Pokerface: Partial Order Keeping and Energy Repressing Method for Extreme Face Illumination Normalization

Felix Juefei-Xu and Marios Savvides
CyLab Biometrics Center, Electrical and Computer Engineering
Carnegie Mellon University, Pittsburgh, PA 15213, USA
felixu@cmu.edu, msavvid@ri.cmu.edu

Abstract

*We propose a new method called the **Pokerface** for extreme face illumination normalization. The Pokerface is a two-phase approach. It first aims at maximizing the minimum gap between adjacently-valued pixels while keeping the partial ordering of the pixels in the face image under extreme illumination condition, an intuitive effort based on order theory to unveil the underlying structure of a dark image. This optimization can be formulated as a feasibility search problem and can be efficiently solved by linear programming. It then smooths the intermediate representation by repressing the energy of the gradient map. The smoothing step is carried out by total variation minimization and sparse approximation. The illumination normalized faces using our proposed Pokerface not only exhibit very high fidelity against neutrally illuminated face, but also allow for a significant improvement in face verification experiments using even the simplest classifier. Simultaneously achieving high level of faithfulness and expressiveness is very rare among other methods. These conclusions are drawn after benchmarking our algorithm against 22 prevailing illumination normalization techniques on both the CMU Multi-PIE database and Extended YaleB database that are widely adopted for face illumination problems.*

1. Introduction

Techniques for normalizing or correcting illumination variations has gained a lot of prominence both in traditional face recognition/surveillance systems and photo/imaging applications in hand-held mobile devices.

From our perspective, the two pillars of quality illumination normalization are **faithfulness** and **expressiveness**. Faithfulness means that the algorithm can recover an illumination normalized image that is of high visual fidelity compared to the neutrally illuminated image of the same person taken under the same setting (pose, expression, *etc*.). Expressiveness means that the illumination normalized image can fully express the identity information of the subject and thus can improve face recognition performance, and most importantly, even when using the simplest possible classifier, *e.g*. the nearest neighbor classifier based on normalized cosine distance. However, most existing approaches address one of these aspects, but not both. In past work, the importance of expressiveness has been addressed, but faithfulness has often been overlooked. Researchers strive to extract illumination-invariant features that can greatly improve the face recognition performance, but those features may not be faithful to the neutrally illuminated face at all. Take local binary patterns (LBP) [29] and many of its variants for example. They have been shown to exhibit some illumination tolerance and can dramatically improve the performance of face recognition [29]. However, LBP images show poor visual fidelity and it is very hard to tell the identity of the subject just by looking at it.

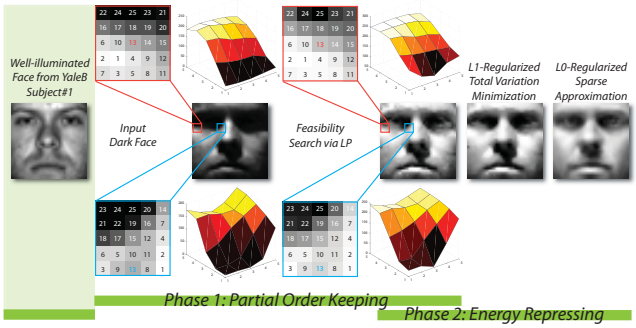

Figure 1. The flowchart of the Pokerface. Phase 1 Partial Order Keeping is accomplished by feasibility search via LP, and Phase 2 Energy Repressing is achieved by ℓ_1-regularized total variation minimization and ℓ_0-regularized sparse approximation.

But why is this such a bad thing to do if we are eventually relying on the computers to unveil the identity of the subject? The answer is as follows. Currently, commercial face recognition systems, widely used among law enforcement agencies, simply cannot deal with images such as the LBP

images. The commercial face matchers are trained using natural face images and it is next to impossible for them to match and LBP image of the subject to the original natural face. Also, faithfulness is important for any applications that has a human-in-the-loop component because human brains interpret natural face images the best. We also notice that, for many algorithms, the level of faithfulness degrades terribly especially for the cases under harsh or even extreme illumination conditions. In this paper, we present a novel approach for extreme face illumination normalization that is outstanding in both faithfulness and expressiveness.

2. Related Work

Let us discuss some of the recent works on illumination normalization techniques applied to face images. A very interesting work by Han *et al*. [12] explicitly exploits maximizing the separability of different subjects' faces as the objective for illumination preprocessing. The authors first decompose the input face via scale-space decomposition and then the normalized face is computed as a linear combination of the decomposed coarse-to-fine bands. Most importantly, the combining coefficients are learned from a training set by maximizing the Fisher criterion. The entire optimization is solved by simulated annealing. Chen *et al*. [5] extract illumination invariant features based on natural images statistics, with which they derive a Wiener filter approach to separate the illumination-invariant features from an image. Wang *et al*. [34] decompose the face image into two channels corresponding to the large- and small-scale features [35], and then a threshold filter is applied to smooth the small-scale features while robust PCA applied to remove shadow on the large-scale features. Finally, the two channels are combined to obtain the normalized image. Matsukawa *et al*. [27] also base their method on small- and large-scale features [35]. They focus on the large-scale feature (LF) in the presence of cast shadows where large-scale features are decomposed using PCA basis, and then the normalized LF is the quotient image of the original LF and the approximated LF using top-K PCA basis. In the end, small- and large-scale features are combined to produce a final output. Arandjelović [2] uses a cascade of processing steps for extracting robust edges and build discriminative and invariant facial features using image gradient directions near those robust edges. However, the author computes illumination-invariant features for face recognition rather than normalizing the illumination. Han *et al*. [11] has provided a comparative study on 12 representative illumination preprocessing methods and grouped them into 3 categories: (1) grey-level transformation *e.g*. histogram equalization, logarithmic transform, (2) gradient or edge extraction *e.g*. Laplacian of Gaussian, and (3) reflectance field estimation *e.g*. work of Tan and Tiggs [32]. The authors have the following conclusion that we also share: "for face recognition purpose, better visualization effect after illumination preprocessing does not imply higher recognition accuracy." This reiterates why satisfying both goals is important for a quality illumination normalization method.

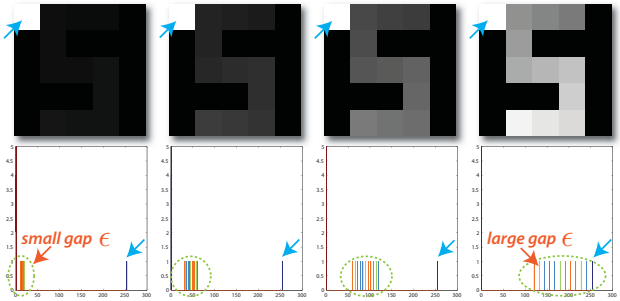

Figure 2. A toy example showing the main idea behind the Pokerface which is to maximize the minimum gap ϵ between adjacently-valued pixels while keeping the partial ordering. The bins inside the green ellipse correspond to the "S" pattern.

3. Main Idea

The Pokerface is very different from any previous work, it is hard to group it into any category summarized in [11]. The main idea behind the Pokerface is very intuitive and straight-forward. It aims at transforming a dark image patch to a bright (illumination normalized) one with distinguishable details by keeping the partial orders of the pixels. As depicted in Figure 2, for each of the four images shown in the first row, a letter "S" pattern is embedded with the rest of the pixels being pure black (intensity=0) except the top-left pixel being pure white (intensity=255). However, we can hardly discern the "S" pattern from the black background in the left-most image. The histograms for these four images are shown in the second row. As can be observed, when the "S" pattern pixels have small intensity gap (ϵ) between adjacently-valued pixels, the pattern is not as salient as the pattern showing larger gaps between adjacently-valued pixels such as the right-most image. The goal of the Pokerface is exactly to maximize the minimum gap (ϵ) between adjacently-valued pixels while preserving the partial ordering of the pixels. The latter is to guarantee that the local edge information is well preserved when normalizing a dark image to a bright one. We will formulate this optimization problem in the context of order theory in the next section.

4. Proposed Method

4.1. Overview and Problem Formulation

The big picture of the Pokerface is a two-phase pipeline as shown in Figure 1. Phase 1: *Partial Order Keeping*. Phase 2: *Energy Repressing*. Thus the name **Pokerface**. In Phase 1, the partial order keeping is accomplished by solving an optimization problem that maximizes the gap

(ϵ) between adjacently-valued pixels with the partial order preserving constraints. This optimization problem will be converted to a feasibility search problem using linear programming as will be detailed in the following subsection. In Phase 2, the "energy" we refer to here is the energy of the gradient map of the intermediate image found by the feasibility search step in Phase 1. Repressing the energy would lead to a smoother representation. The energy repressing phase is achieved by two steps: an ℓ_1-regularized total variation minimization step and an ℓ_0-regularized sparse approximation step. We will explain each component of the Pokerface in the subsections to follow. In order to formulate this problem, it is imperative that we first go over some definitions from order theory [3, 6, 31].

Definition 1. *A binary relation $\mathfrak{R}$ on a nonempty set $\mathbb{X}$ is* ***reflexive*** *if $x \mathfrak{R} x$ for every $x \in \mathbb{X}$. It is* ***antisymmetric*** *if $x \mathfrak{R} y$ implies $y \mathfrak{R} x$, for every $x, y \in \mathbb{X}$. It is* ***transitive*** *if $x \mathfrak{R} y \mathfrak{R} z$ implies $x \mathfrak{R} z$, for every $x, y, z \in \mathbb{X}$.*

Definition 2. *A binary relation $\succsim$ on a nonempty set $\mathbb{X}$ is a* ***preorder*** *on $\mathbb{X}$ if it is transitive and reflexive. It is a* ***partial order*** *on $\mathbb{X}$ if it is an antisymmetric preorder on $\mathbb{X}$.*

Definition 3. *A* ***preordered set*** *is an ordered pair $(\mathbb{X}, \succsim)$, where $\mathbb{X}$ is a nonempty set and $\succsim$ is a preorder on $\mathbb{X}$.*

Definition 4. *A preordered set $(\mathbb{X}, \succcurlyeq)$ is a* ***partially ordered set****, or* ***poset****, if $\succcurlyeq$ is a partial order on $\mathbb{X}$.*

Definition 5. *Let $(\mathbb{X}, \succcurlyeq)$ and $(\mathbb{Y}, \succcurlyeq)$ be two partially ordered sets. The* ***order-preserving map*** *from $(\mathbb{X}, \succcurlyeq)$ into $(\mathbb{Y}, \succcurlyeq)$ is a function $f : \mathbb{X} \mapsto \mathbb{Y}$ such that $a \succcurlyeq b$ implies $f(a) \succcurlyeq f(b)$ for every $a, b \in \mathbb{X}$.*

For every pixel $X_{i,j}$ in the (zero-padded) face image under extreme illumination condition (or "dark face"), we consider an odd-sized $N \times N$ ($N = 3, 5, 7, ...$) square region around it such that $X_{i,j}$ is the center pixel of the patch. These N^2 pixels in the patch form a partially ordered set $(\mathbb{X}_{i,j}, \succcurlyeq)$. Let $Y_{i,j}$ be the counterpart of $X_{i,j}$ in the bright (illumination normalized) face where $(\mathbb{Y}_{i,j}, \succcurlyeq)$ is also a partially ordered set containing the center pixel $Y_{i,j}$ and its neighboring pixels within a patch of the bright face.

Moreover, let function $f_{i,j} : \mathbb{X}_{i,j} \mapsto \mathbb{Y}_{i,j}$ be an order-preserving mapping from $(\mathbb{X}_{i,j}, \succcurlyeq)$ to $(\mathbb{Y}_{i,j}, \succcurlyeq)$ such that $a \succcurlyeq b$ implies $f_{i,j}(a) \succcurlyeq f_{i,j}(b)$ for every $a, b \in \mathbb{X}_{i,j}$.

We adopt a shifting-window approach where we only establish the binary relations between the center pixel and its $(N^2 - 1)$ neighbors, rather than a fully-paired case with $\frac{1}{2}\binom{N^2}{2}$ binary relations. Readers can easily verify that establishing the relations between only the center pixel and its neighbors for an $N \times N$ patch under shifting-window, is equivalent to establishing fully-paired relations within each window of width $(\frac{N-1}{2} + 1)$, as depicted in orange color in Figure 3.

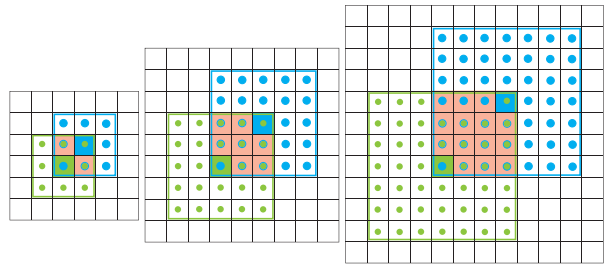

Figure 3. For each $N \times N$ patch, the center pixel is plotted as a colored square, and its neighbors are plotted as dots with the same color. With the window shifting, the orange region is the largest region where pixels inside can have fully-paired binary relations.

Now, we can formulate the main idea of the Pokerface using the following optimization where we aim at maximizing the minimum gap (ϵ) between partially ordered pixels while keeping the same partial ordering ($\succcurlyeq$).

$$\begin{aligned} &\underset{\epsilon}{\text{maximize}} \quad \epsilon \\ \text{subject to} \quad &\forall a, b \in \mathbb{X}_{i,j}, \ \ \forall i, j \\ &f_{i,j}(a) \succcurlyeq f_{i,j}(b) + \epsilon \\ 0 \leq \min(f_{i,j}(a), &f_{i,j}(b)) \leq \max(f_{i,j}(a), f_{i,j}(b)) \leq 255 \end{aligned} \tag{1}$$

However, this optimization requires explicit knowledge of every mapping function $f_{i,j}$ which is neither efficient nor feasible to learn. Also, solving such an optimization is NP-hard. Thanks to the reformulation to be discussed in the next subsection, we are able to obtain the bright face without explicitly knowing the mapping functions $f_{i,j}$, while, most importantly, satisfying the partial order constraints.

4.2. Feasibility Search via LP

The optimization problem (1) can be re-formulated as a feasibility search problem using linear programming.

Suppose we have a dark face of size $m \times n$ as shown in Figure 4. Each $X_{i,j}$ corresponds to the pixel intensity at location (i, j) and the arrow pointing from $X_{i,j}$ to $X_{k,l}$ means that the pixel intensity at (i, j) is greater than that at (k, l). If (i, j) and (k, l) have the same intensity, the arrow points to both directions and is shown in blue.

We want our illumination normalization algorithm to preserve the partial ordering in every local image patch. In other words, the bright face image should have exactly the same partial ordering as the input dark face. Take $X_{3,3}$ for instance, in its 3×3 neighborhood (the green box), these 9 pixels form a partially ordered set $(\mathbb{X}_{3,3}, \succcurlyeq)$, and the partial order characteristic ($\succcurlyeq$) can be written as the following by comparing the center pixel $X_{3,3}$ to its 8 neighbors:

$$\begin{aligned} &X_{3,3} < X_{2,2} \qquad X_{3,3} > X_{3,2} \qquad X_{3,3} < X_{4,2} \\ &X_{3,3} > X_{4,3} \qquad X_{3,3} > X_{4,4} \qquad X_{3,3} > X_{3,4} \\ &X_{3,3} = X_{2,4} \qquad X_{3,3} < X_{2,3} \end{aligned} \tag{2}$$

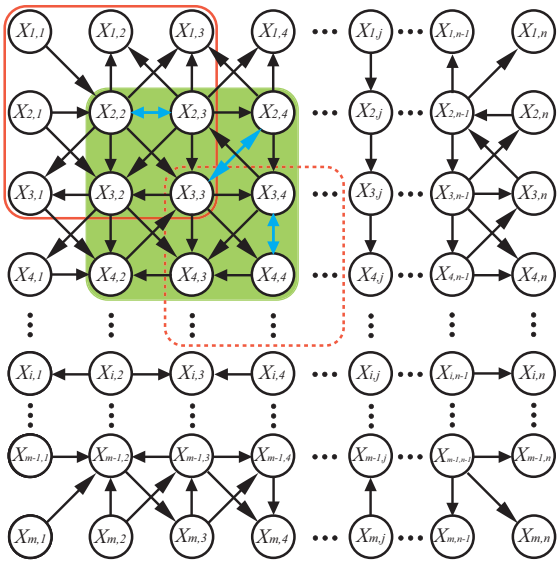

Figure 4. In Phase 1 Partial Order Keeping, the illumination normalized image $\mathbf{y} = \{Y_{i,j}\}, \forall i, j$ should have the same partial ordering as the dark face $\mathbf{x} = \{X_{i,j}\}, \forall i, j$. The arrows indicate binary relations between pixels and the orange box indicates the shifting window.

If $X_{i,j}$ is greater than $X_{k,l}$, then it is equivalent to say that $X_{i,j}$ is greater or equal to $X_{k,l} + 1$ since pixel intensities can only be integers: $X_{i,j} > X_{k,l} \Leftrightarrow X_{i,j} \geq X_{k,l} + 1$. If $X_{i,j}$ is equal to $X_{k,l}$, then it is equivalent to say that $X_{i,j}$ is greater or equal to $X_{k,l} + 0$ and $X_{k,l}$ is greater or equal to $X_{i,j} + 0$, where a single equality relation is represented by *two* inequality relations. $X_{i,j} = X_{k,l} \Leftrightarrow X_{i,j} \geq X_{k,l} + 0 \; and \; X_{k,l} \geq X_{i,j} + 0$. So, the relations in (2) become:

$$
\begin{aligned}
&(+1)X_{3,3} + (-1)X_{2,2} \leq -1, (-1)X_{3,3} + (+1)X_{3,2} \leq -1 \\
&(+1)X_{3,3} + (-1)X_{4,2} \leq -1, (-1)X_{3,3} + (+1)X_{4,3} \leq -1 \\
&(-1)X_{3,3} + (+1)X_{4,4} \leq -1, (-1)X_{3,3} + (+1)X_{3,4} \leq -1 \\
&(+1)X_{3,3} + (-1)X_{2,3} \leq -1, (-1)X_{3,3} + (+1)X_{2,4} \leq 0 \\
&(+1)X_{3,3} + (-1)X_{2,4} \leq 0
\end{aligned} \tag{3}
$$

Relation (3) is a set of linear constraints that pixel $X_{3,3}$ has to satisfy in order to keep the partial ordering. By scanning through all the pixels in the input dark image, we can generate the complete list of constraints. A careful scanning will visit every pair of adjacent pixels exactly once, eliminating duplicate constraints from considering the flipped version of a previously visited pair.

According to this setup, let $\mathbf{x}$ be a vector containing all the $X_{i,j}$'s, we can write all of the linear constraints in matrix form: $\mathbf{Ax} \leq \mathbf{b}$, where $\mathbf{A}$ should be a sparse matrix whose non-zero elements are either $+1$ or -1 and the locations of $+1$ and -1 indicate which $X_{i,j}$ and $X_{k,l}$ are being compared. $\mathbf{b}$ should be a vector whose elements are either -1 or 0, indicating the corresponding RHS of the inequality constraints, which is the minimum gap between pairs of pixels. Since we are dealing with 8-bit grayscale image, all the $X_{i,j}$ should be within the range 0 to 255.

Now, let us take a step back and consider how to use these conditions from the dark face to obtain the bright face without explicitly knowing the mapping functions $f_{i,j} : \mathbb{X} \mapsto \mathbb{Y}$ that maps each dark face patch $(\mathbb{X}_{i,j}, \succcurlyeq)$ to bright face patch $(\mathbb{Y}_{i,j}, \succcurlyeq)$ while preserving the partial order. Recall that $Y_{i,j}$ is the counterpart of $X_{i,j}$, and each patch around $Y_{i,j}$ forms a partially ordered set $\mathbb{Y}_{i,j}$ having the same ordering $(\succcurlyeq)$ as $\mathbb{X}_{i,j}$. Let $l^l_{i,j}$ be the lth lower-valued neighboring pixel of $Y_{i,j}$, and similarly, $h^h_{i,j}$ be the hth higher-valued neighboring pixel of $Y_{i,j}$. The optimization (1) can be reformulated as:

$$
\begin{aligned}
&\underset{\epsilon}{\text{maximize}} \quad \epsilon \\
\text{subject to} \quad &(-1)Y_{i,j} + (+1)l^l_{i,j} \leq -\epsilon, \;\; \forall l, \;\; \forall i, j \\
&(+1)Y_{i,j} + (-1)h^h_{i,j} \leq -\epsilon, \;\; \forall h, \;\; \forall i, j \\
&0 \leq Y_{i,j} \leq 255, \;\; \forall i, j
\end{aligned} \tag{4}
$$

Since we know that each $l^l_{i,j}$ and $h^h_{i,j}$ is actually some pixel $Y_{k,l}$ in the bright face, whose partial order should be precisely captured by the linear constraints $\mathbf{Ay} \leq \epsilon\mathbf{b}$, $(\epsilon = 1, 2, 3, ...)$, where $\mathbf{y}$ is a vector containing all the $Y_{i,j}$'s. It is worth noting that here matrix $\mathbf{A}$ and vector $\mathbf{b}$ are directly obtained from the partial order characteristics of the dark face. For a known gap value ϵ, finding the bright face $\mathbf{y}$ under the aforementioned linear constraints $\mathbf{Ay} \leq \epsilon\mathbf{b}$ is a feasibility search problem [4], which can be efficiently solved using linear programming by setting the objective function to be 0. The feasibility search problem, also called the satisfiability problem, can be regarded as the special case of mathematical optimization where the objective value is the same for every solution, and thus any solution is optimal. The optimization is often written as:

$$
\begin{aligned}
&\text{find} \quad \mathbf{y} \\
&\text{subject to} \quad \mathbf{Ay} \leq \epsilon\mathbf{b} \\
&0 \leq \mathbf{y}_s \leq 255, \quad s = 1, 2, \ldots, m \times n
\end{aligned} \tag{5}
$$

The Pokerface aims at finding the bright face with the largest gap value ϵ. Thanks to the limited value ϵ can take, this part can be done greedily. In other words, we increase ϵ until the solution is no longer feasible. In the Pokerface, we choose $\epsilon = 3$, and patch width $N = 7$. When considering a larger image patch, there are more constraints that the center pixel has to satisfy, and consequently, the gap ϵ usually cannot be very large. Otherwise, the search for $\mathbf{y}$ becomes infeasible and no solution can be found.

4.3. ℓ_1-Regularized Total Variation Minimization

After Phase 1 of the Pokerface, the intermediate face may show some non-smooth artifacts. This is because dur-

Figure 5. Different levels of smoothing by varying τ in the total variation minimization stage. This example is from YaleB+.

ing Phase 1, there are no smoothness constraints to be satisfied in the optimization. This is done purposefully, we want a subsequent step down the line to perform the smoothing task, rather than solving a single convoluted multi-purpose optimization.

The concept of total variation (TV) was introduced in computer vision first by Rudin, Osher and Fatemi [30]. However, it is very difficult to be minimized by conventional methods. Therefore, we resort to an iterative *split Bregman* method [9] for solving the total variation minimization problem. The split Bregman method can solve a series of ℓ_1-regularized problem in the form of:

$$\underset{u}{\text{minimize}}\, |\Phi(u)| + H(u) \tag{6}$$

where $|\cdot|$ denotes the ℓ_1 norm, and both $|\Phi(u)|$ and $H(u)$ are convex functions. Following this, the isotropic total variation minimization problem can be formulated as:

$$\underset{u}{\text{minimize}} \sum_i \sqrt{(\nabla_x u)_i^2 + (\nabla_y u)_i^2} + \frac{\mu}{2}\|u - f\|_2^2 \tag{7}$$

where f represents the original noisy image and u is the smooth image after TV minimization. The key to the split Bregman method is that the ℓ_1 and ℓ_2 portions of the energy in Equation (6) are decoupled.

For better visualizing the effect of the penalty coefficient, we rewrite Equation (7) as follows:

$$\underset{u}{\text{minimize}}\, \tau \sum_i \sqrt{(\nabla_x u)_i^2 + (\nabla_y u)_i^2} + \frac{1}{2}\|u - f\|_2^2 \tag{8}$$

We can vary τ to get different levels of smoothing. For large τ, we will end up with very washed out images, and for small τ it will still be noisy as shown in Figure 5. An example of illumination normalized image after ℓ_1-regularized total variation minimization is shown in Figure 1, with much smoother appearance. In the Pokerface, $\tau = 8$.

4.4. ℓ_0-Regularized Sparse Approximation

In some cases, even after total variation minimization, there still remains some splotchy artifacts, such as some black/white spots on the subject's left cheek and chin area, as shown in Figure 5. This is because in Phase 1, if some pixels are already the local extrema in the dark face, when recovering the bright face, they will be pushed to 0 or 255. To account for that, we resort to an ℓ_0-regularized sparse approximation method for removing the splotchy artifacts. First, we need to obtain an overcomplete dictionary.

Dictionary learning methods, such as the K-SVD [1] algorithm, are very well known for image denoising [7] and inpainting [25], which serves our purpose here exactly. K-SVD aims at extending the K-means algorithm with the analogy that the cluster centers are the elements of the learned dictionary and the memberships are determined by the sparse approximations of the signals in that dictionary. It provides a solution to the following optimization:

$$\underset{\mathbf{D},\boldsymbol{\alpha}}{\text{minimize}}\, \|\mathbf{X} - \mathbf{D}\boldsymbol{\alpha}\|_F^2 \text{ subject to } \|\boldsymbol{\alpha}_i\|_0 < K_1,\ \forall i \tag{9}$$

where $\mathbf{X}$, $\mathbf{D}$ and $\boldsymbol{\alpha}$ are the data, the learned overcomplete dictionary and the sparse approximation matrix respectively. Here $\|.\|_0$ is the ℓ_0 pseudo-norm measuring sparsity. The sparse approximations of the data elements are allowed to have some maximum sparsity $\|\mathbf{x}\|_0 \leq K_1$.

With the learned dictionary $\mathbf{D}$, any intermediate face $\mathbf{y}$ after TV minimization step can be sparsely approximated by the elements of $\mathbf{D}$ following:

$$\boldsymbol{\alpha}_{\mathbf{y}} = \arg\min_{\alpha} \|\mathbf{y} - \mathbf{D}\boldsymbol{\alpha}\| \text{ subject to } \|\boldsymbol{\alpha}\|_0 < K_2 \tag{10}$$

which can be efficiently solved using any sparse coding algorithm such as the orthogonal matching pursuit (OMP) [28]. The sparse approximation of $\mathbf{y}$ is therefore $\hat{\mathbf{y}} = \mathbf{D}\boldsymbol{\alpha}_{\mathbf{y}}$.

So far, the Pokerface is agnostic to image size and registration. The patch-based K-SVD method can be applied so that the Pokerface can remain agnostic. However, if we do deal with well-aligned images with a fixed size, we can take advantage by applying full-image-based K-SVD. The patch-based K-SVD is usually trained using images patches of size 8×8 and obtain an overcomplete dictionary $\mathbf{D}_{\text{patch}} \in \mathbb{R}^{64\times M}, M \gg 64$, where M is the number of atoms in the learned dictionary. The query image is decomposed into many 8×8 blocks and each block is then sparsely approximated using $\mathbf{D}_{\text{patch}}$. The resulting reconstruction usually exhibits block artifacts which can be alleviated by applying a shifting-window approach. We use full-image-based K-SVD where we train the dictionary $\mathbf{D}_{\text{full}} \in \mathbb{R}^{(m\times n)\times M}, M \gg (m \times n)$ using full face images of size $m \times n$. Experimentally, the full-image-based K-SVD leads to better reconstruction. The only caveat is that since the dimension of the dictionary atoms are the same as the image, learning an overcomplete dictionary will require much more training data. An example of illumination normalized image after applying the ℓ_0-regularized sparse approximation is shown in Figure 1.

5. Experiments

In this section, we will evaluate both the faithfulness and expressiveness of the proposed Pokerface. We also

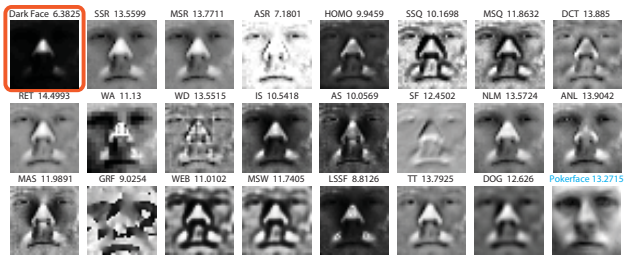

Figure 6. Example from YaleB+ subject #1 showcasing various illumination normalization algorithms. PSNR is also displayed for this particular image.

benchmark it against 22 prevailing illumination normalization techniques with open-source implementations consolidated in [33], with optimally tuned parameters. Unfortunately, implementation of some of the recent work discussed in Section 2 are not publicly available. We also omit the basic algorithms like histogram equalization, logarithmic transform, gamma intensity correction, Gaussian high-pass, directional gray-scale derivative, *etc*. for the sake of brevity. The 22 algorithms for comparison are: (1) single scale retinex (SSR), (2) multi scale retinex (MSR), (3) adaptive single scale retinex (ASR), (4) homomorphic filtering (HOMO), (5) single scale self quotient image (SSQ), (6) multi scale self quotient image (MSQ), (7) discrete cosine transform (DCT), (8) retina modeling (RET), (9) wavelet (WA), (10) wavelet denoising (WD), (11) isotropic diffusion (IS), (12) anisotropic diffusion (AS), (13) steering filter (SF), (14) non-local means (NLM), (15) adaptive non-local means (ANL), (16) modified anisotropic diffusion (MAS), (17) gradientfaces (GRF), (18) single scale Weberfaces (WEB), (19) multi scale Weberfaces (MSW), (20) large and small scale features (LSSF), (21) Tan and Triggs (TT), and (22) difference of Gaussian filtering (DOG). One example for each algorithm is shown in Figure 6, along with the Pokerface.

Face Illumination Databases

There exist multiple publicly available databases targeting face illumination problems. Among these, The CMU Multi-PIE (MPIE) database [10] is the largest publicly available illumination database in terms of subject size. Yale Face Database B (YaleB) [8] and Extended Yale Face Database B (ExtYaleB) [26] contain the most extreme illumination conditions. However, less constrained database such as the FRGC ver 2.0 [21, 13, 36] is not an ideal illumination database because it contains interference from many other factors *e.g*. facial expression, image blur, pose variations, *etc*. Only by looking at purely the illumination factor and its relationship to face recognition performance, can we draw more meaningful and pinpointed conclusions.

The MPIE comprises of images with less harsh illumination conditions. There are 249 subjects in the database with 20 images per subject showing frontal face under various illumination variations. Therefore the total number of images we conduct the experiments on is $249 \times 20 = 4980$.

Figure 7. Visual results of the Pokerface on 4 subjects of MPIE under all illumination variations.

The YaleB contains single light source images of 10 subjects each seen under 576 viewing conditions (9 poses $\times$ 64 illumination conditions). For every subject in a particular pose, an image with ambient (background) illumination was also captured. The ExtYaleB contains images of 28 human subjects under 9 poses and 64 illumination conditions. We combine the unique subjects in both database and call it YaleB+. So we have $10+28 = 38$ subjects in total. We only choose the frontal image with all the illumination variations. The entire database on which we conduct our experiments contains $38 \times 64 = 2432$ images.

Faithfulness: Fidelity Experiments

For the faithfulness experiments, we investigate the fidelity of the illumination normalized image against the neutrally-illuminated face of the same subjects for both databases. We adopt the broadly used peak signal-to-noise ratio (PSNR) [24, 22, 17, 16, 20] as the fidelity measurement. All the dark images are resized to be 32×32 and then various algorithms are applied, including the Pokerface, for illumination normalization. We report the average PSNR for each of the algorithms on both databases in the PSNR columns in Table 1 and 2. In addition, some visual results of the Pokerface are shown in Figure 7 and 8.

Expressiveness: Face Verification Experiments

For the expressiveness, we conduct face verification experiments by matching all the images to themselves. We utilize the simplest nearest neighbor classifier based on the normalized cosine distance (NCD) [23, 18, 14, 15, 19] between each query and galley image. The features are the raw representation out of each algorithm. Thus, the resulting similarity matrix will be of size 4980×4980 for MPIE, and 2432×2432 for YaleB+, whose entry $\mathrm{SimM}(i, j)$ is the NCD between query i and gallery j. The experimental results are reported using receiver operating characteristic (ROC) curves. Both the verification rate (VR) at 0.1%

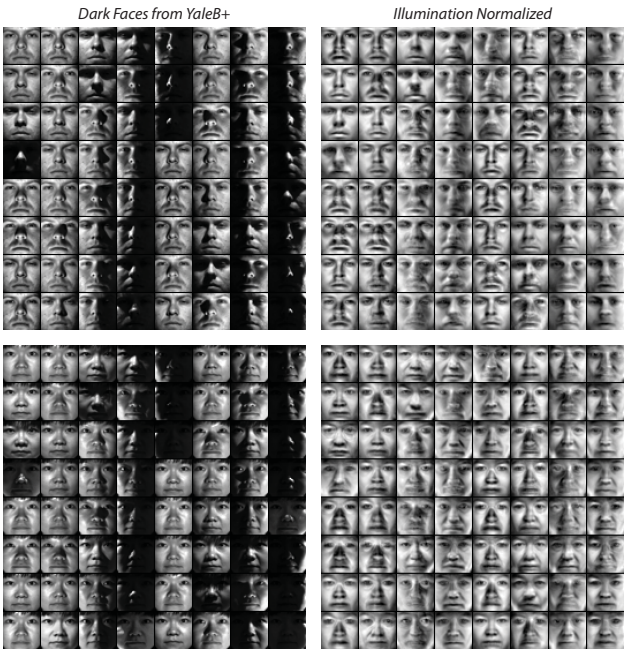

Figure 8. Visual results of the Pokerface on 2 subjects of YaleB+ under all illumination variations .

(0.001) false accept rate (FAR) and equal error rate (EER) are computed. Figure 9 and 10 contain ROC curves for the Pokerface and 22 other competing algorithms, for MPIE and YaleB+ respectively. VR at 0.1% FAR and EER are also consolidated in Table 1 and 2.

Discussion

In Table 1 and 2, we also indicate the rankings of VR and PSNR for the 22 competing algorithms with the proposed Pokerface. In addition, the sum of the 2 rankings are shown in the $\sum$ column with the next column showing the total ranking for both VR and PSNR. As mentioned in the beginning of this paper, many algorithms that have high VR scores, may not have high PSNR scores and vice versa. It is very rare for an illumination normalization algorithm to be outstanding in both faithfulness and expressiveness, which is indeed the case for our proposed Pokerface algorithm.

6. Conclusions

The faithfulness and expressiveness are the two important criteria to determine whether an illumination normalization algorithm is of high quality or not. We have presented a practical and effective Pokerface method for extreme face illumination normalization. Such a method exhibits very high level of faithfulness and expressiveness at the same time, which is outstanding among many competing algorithms. The formulation of the Pokerface is also very intuitive because it aims at maximizing the minimum gap between adjacently-valued pixels while keeping the partial ordering of the pixels in the dark face. We reformulate this optimization as a feasibility search problem which is efficiently solved by LP. Next, a smoothing step involving total variation minimization and sparse approximation is exercised for improved enhancement quality. The effectiveness of the Pokerface in terms of both faithfulness and expressiveness is confirmed after benchmarking our algorithm against 22 prevailing illumination normalization techniques on both MPIE and YaleB+ database which are widely adopted for face illumination problems.

References

[1] M. Aharon, M. Elad, and A. Bruckstein. K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation. *Signal Processing, IEEE Transactions on*, 54(11):4311–4322, Nov 2006. 5

[2] O. Arandjelović. Gradient Edge Map Features for Frontal Face Recognition under Extreme Illumination Changes. In *British Machine Vision Conference (BMVC)*, 2012. 2

[3] G. Birkhoff. *Lattice Theory (3rd Revised ed.)*. American Mathematical Society, 1940. 3

[4] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004. 4

[5] L.-H. Chen, Y.-H. Yang, C.-S. Chen, and M.-Y. Cheng. Illumination Invariant Feature Extraction Based on Natural Images Statistics - Taking Face Images as an Example. In *IEEE CVPR*, 2011. 2

[6] B. A. Davey and H. A. Priestley. *Intro. to Lattices and Order (2nd ed.)*. Cambridge University Press, 2002. 3

[7] M. Elad and M. Aharon. Image Denoising via Sparse and Redundant Representations over Learned Dictionaries. *Image Processing, IEEE Transactions on*, 15(12):3736–3745, Dec 2006. 5

[8] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman. From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose. *IEEE TPAMI*, 23(6):643–660, jun 2001. 6

[9] T. Goldstein and S. Osher. The Split Bregman Method for L1-regularized Problems. *SIAM Journal on Imaging Sciences*, 2(2):323–343, 2009. 5

[10] R. Gross, I. Matthews, J. F. Cohn, T. Kanade, and S. Baker. Multi-PIE. In *IEEE FG*, sep 2008. 6

[11] H. Han, S. Shan, X. Chen, and W. Gao. A Comparative Study on Illumination Preprocessing in Face Recognition. *Pattern Recognition*, 46:1691–1699, 2013. 2

[12] H. Han, S. Shan, X. Chen, S. Lao, and W. Gao. Separability Oriented Preprocessing for Illumination-insensitive Face Recognition. In *European Conference on Computer Vision (ECCV)*, 2012. 2

[13] F. Juefei-Xu, M. Cha, J. L. Heyman, S. Venugopalan, R. Abiantun, and M. Savvides. Robust Local Binary Pattern Feature Sets for Periocular Biometric Identification. In *Biometrics: Theory Applications and Systems (BTAS), 4th IEEE Int'l Conf. on*, pages 1–8, sep 2010. 6

[14] F. Juefei-Xu, M. Cha, M. Savvides, S. Bedros, and J. Trojanova. Robust Periocular Biometric Recognition Using Multi-level Fusion of Various Local Feature Extraction Techniques. In *IEEE 17th International Conference on Digital Signal Processing (DSP)*, 2011. 6

[15] F. Juefei-Xu, K. Luu, M. Savvides, T. Bui, and C. Suen. Investigating Age Invariant Face Recognition Based on Periocular Biometrics. In *Biometrics (IJCB), 2011 International Joint Conference on*, pages 1–7, Oct 2011. 6

[16] F. Juefei-Xu, D. K. Pal, and M. Savvides. Hallucinating the Full Face from the Periocular Region via Dimensionally Weighted K-SVD. In *IEEE CVPR Workshops*, June 2014. 6

[17] F. Juefei-Xu, D. K. Pal, and M. Savvides. NIR-VIS Heterogeneous Face Recognition via Cross-Spectral Joint Dictionary Learning and Reconstruction. In *IEEE CVPR Workshops*, June 2015. 6

[18] F. Juefei-Xu and M. Savvides. Unconstrained Periocular Biometric Acquisition and Recognition Using COTS PTZ Camera for Uncooperative and Non-cooperative Subjects. In *Applications of Computer Vision (WACV), 2012 IEEE Workshop on*, pages 201–208, Jan 2012. 6

[19] F. Juefei-Xu and M. Savvides. An Augmented Linear Discriminant Analysis Approach for Identifying Identical Twins with the Aid of Facial Asymmetry Features. In *IEEE CVPR Workshops*, pages 56–63, June 2013. 6

[20] F. Juefei-Xu and M. Savvides. An Image Statistics Approach towards Efficient and Robust Refinement for Landmarks on Facial Boundary. In *Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on*, pages 1–8, Sept 2013. 6

[21] F. Juefei-Xu and M. Savvides. Subspace Based Discrete Transform Encoded Local Binary Patterns Representations for Robust Periocular Matching on NIST's Face Recognition Grand Challenge. *IEEE Trans. on Image Processing*, 23(8):3490–3505, aug 2014. 6

[22] F. Juefei-Xu and M. Savvides. Encoding and Decoding Local Binary Patterns for Harsh Face Illumination Normalization. In *IEEE International Conference on Image Processing (ICIP)*, Sept 2015. 6

[23] F. Juefei-Xu and M. Savvides. Pareto-optimal Discriminant Analysis. In *IEEE ICIP*, Sept 2015. 6

[24] F. Juefei-Xu and M. Savvides. Single Face Image Super-Resolution via Solo Dictionary Learning. In *IEEE International Conference on Image Processing (ICIP)*, Sept 2015. 6

[25] M.-S. Koh and E. Rodriguez-Marek. Turbo Inpainting: Iterative K-SVD with a New Dictionary. In *IEEE International Workshop on Multimedia Signal Processing*, pages 1–6, Oct 2009. 5

[26] K.-C. Lee, J. Ho, and D. J. Kriegman. Acquiring Linear Subspaces for Face Recognition under Variable Lighting. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(5):684–698, may 2005. 6

[27] T. Matsukawa, T. Okabe, and Y. Sato. Illumination Normalization of Face Images with Cast Shadows. In *IEEE Int'l Conf. on Pattern Recognition (ICPR)*, Nov 2012. 2

[28] Y. Pati, R. Rezaiifar, and P. Krishnaprasad. Orthogonal Matching Pursuit: Recursive Function Approximation with Application to Wavelet Decomposition. In *Asilomar Conf. on Signals, Systems and Comput.*, 1993. 5

[29] M. Pietikainen, A. Hadid, and T. Ahonen. *Computer Vision Using Local Binary Patterns*. Springer, 2011. 1

[30] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear Total Variation Based Noise Removal Algorithms. *Physica D*, 60:259–268, 1992. 5

[31] B. S. W. Schröder. *Ordered Sets: An Introduction*. Birkhäuser Boston, 2003. 3

[32] X. Tan and B. Triggs. Enhanced Local Texture Feature Sets for Face Recognition under Difficult Lighting Conditions. *Image Processing, IEEE Transactions on*, 19(6):1635–1650, june 2010. 2

[33] V. Štruc and N. Pavešić. Performance Evaluation of Photometric Normalization Techniques for Illumination Invariant Face Recognition. *Advances in Face Image Analysis: Techniques and Technologies*, 2010. 6

[34] H. Wang, M. Ye, and S. Yang. Shadow Compensation and Illumination Normalization of Face Image. *Machine Vision and Applications*, 24:1121–1131, 2013. 2

[35] X. Xie, W. S. Zheng, J. Lai, P. C. Yuen, and C. Y. Suen. Normalization of Face Illumination Based on Large-and Small-Scale Features. *Image Processing, IEEE Transactions on*, 20(7):1807–1821, july 2011. 2

[36] N. Zehngut, F. Juefei-Xu, R. Bardia, D. K. Pal, C. Bhagavatula, and M. Savvides. Investigating the Feasibility of Image-Based Nose Biometrics. In *IEEE International Conference on Image Processing (ICIP)*, Sept 2015. 6

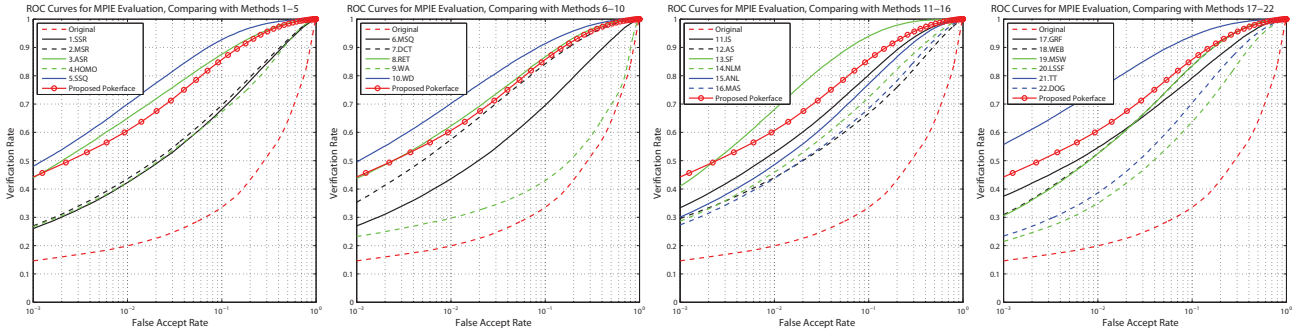

Figure 9. ROC curves for our proposed Pokerface and other 22 competing algorithms on MPIE. These 22 algorithms are split into four sub-figures, and in each sub-figure, ROC curves for the original image performance and the Pokerface performance are plotted for comparison.

Table 1. Results including verification rate (VR) at 0.1% false accept rate (FAR), equal error rate (EER), and peak-signal-to-noise ratio (PSNR) are tabulated for experiments on MPIE. VR/PSNR/overall rankings are also shown by Rk_{VR}, Rk_{PSNR}, and $Rk_{\sum}$ respectively. $\sum$ column is the sum of Rk_{VR} and Rk_{PSNR}, which leads to the overall ranking $Rk_{\sum}$.

Method	**VR**	**(EER)**	Rk_{VR}	**PSNR**	Rk_{PSNR}	$\sum$	$Rk_{\sum}$	**Method**	**VR**	**(EER)**	Rk_{VR}	**PSNR**	Rk_{PSNR}	$\sum$	$Rk_{\sum}$
Original	0.1457	(0.2017)	—	—	—	—	—	AS	0.2958	(0.2224)	15	12.1041	17	32	19
SSR	0.2633	(0.2070)	20	11.5426	20	40	22	SF	0.4118	(0.0749)	8	11.9702	18	26	15
MSR	0.2672	(0.2044)	19	11.4762	21	40	23	NLM	0.2914	(0.1885)	16	11.6009	19	35	21
ASR	0.4406	(0.1095)	5	8.6825	23	28	16	ANL	0.2999	(0.1585)	14	13.5821	9	23	12
HOMO	0.2674	(0.2185)	18	12.8675	12	30	17	MAS	0.2755	(0.2043)	17	14.0922	5	22	10
SSQ	0.4868	(0.0851)	3	12.2546	16	19	7	GRF	0.3763	(0.1551)	9	8.9773	22	31	18
MSQ	0.4300	(0.1157)	7	12.7599	14	21	9	WEB	0.3131	(0.1299)	12	13.3675	11	23	13
DCT	0.3547	(0.1270)	10	13.5237	10	20	8	MSW	0.3091	(0.1255)	13	14.3141	4	17	5
RET	0.4395	(0.1120)	6	14.0250	7	13	3	LSSF	0.2146	(0.2211)	23	14.8537	3	26	14
WA	0.2307	(0.3710)	22	12.7786	13	35	20	TT	0.5602	(0.0719)	1	12.6407	15	16	4
WD	0.4986	(0.0879)	2	13.9621	8	10	2	DOG	0.2364	(0.1852)	21	14.7781	2	23	11
IS	0.3363	(0.1443)	11	14.0754	6	17	6	Pokerface	0.4445	(0.1217)	4	14.9831	1	5	1

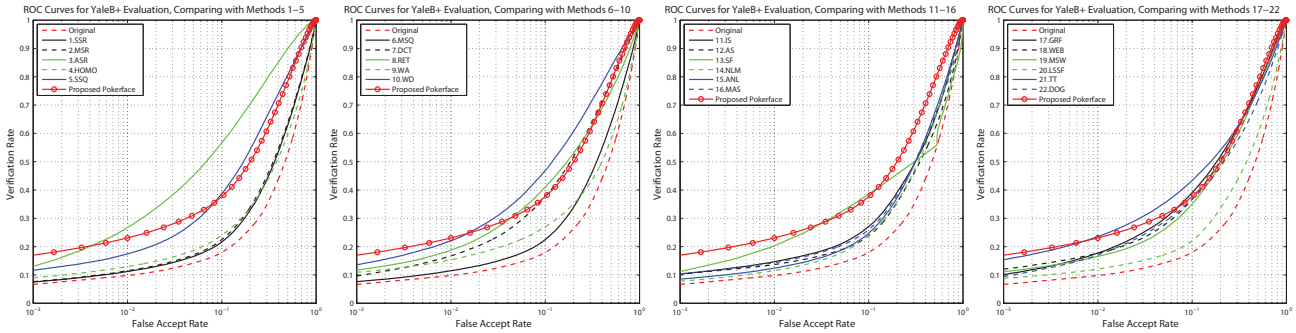

Figure 10. ROC curves for our proposed Pokerface and other 22 competing algorithms on YaleB+. These 22 algorithms are split into four sub-figures, and in each sub-figure, ROC curves for the original image performance and the Pokerface performance are plotted for comparison.

Table 2. Results including verification rate (VR) at 0.1% false accept rate (FAR), equal error rate (EER), and peak-signal-to-noise ratio (PSNR) are tabulated for experiments on YaleB+. VR/PSNR/overall rankings are also shown by Rk_{VR}, Rk_{PSNR}, and $Rk_{\sum}$ respectively. $\sum$ column is the sum of Rk_{VR} and Rk_{PSNR}, which leads to the overall ranking $Rk_{\sum}$.

Method	**VR**	**(EER)**	Rk_{VR}	**PSNR**	Rk_{PSNR}	$\sum$	$Rk_{\sum}$	**Method**	**VR**	**(EER)**	Rk_{VR}	**PSNR**	Rk_{PSNR}	$\sum$	$Rk_{\sum}$
Original	0.0750	(0.4808)	—	—	—	—	—	AS	0.1040	(0.4342)	14	11.6312	15	29	17
SSR	0.0755	(0.4319)	22	10.3763	20	42	23	SF	0.1136	(0.4519)	9	12.9689	5	14	4
MSR	0.0749	(0.4252)	23	10.3884	19	42	22	NLM	0.0766	(0.4207)	21	10.7308	18	39	21
ASR	0.1332	(0.2482)	4	6.4917	23	27	16	ANL	0.0862	(0.4069)	20	12.1193	13	33	18
HOMO	0.0906	(0.4434)	18	11.5867	16	34	19	MAS	0.1043	(0.4047)	13	12.9829	4	17	5
SSQ	0.1181	(0.3160)	6	10.1418	21	27	15	GRF	0.1024	(0.3423)	15	9.0207	22	37	20
MSQ	0.1173	(0.3497)	8	10.9452	17	25	13	WEB	0.1213	(0.3434)	5	12.1227	12	17	7
DCT	0.0973	(0.3374)	16	12.3461	9	25	11	MSW	0.1129	(0.3568)	10	12.2681	11	21	10
RET	0.1174	(0.3441)	7	12.3458	10	17	6	LSSF	0.0879	(0.4397)	19	12.7387	7	26	14
WA	0.1096	(0.4546)	11	12.1002	14	25	12	TT	0.1559	(0.3458)	2	13.2149	3	5	2
WD	0.1382	(0.2963)	3	12.7586	6	9	3	DOG	0.0961	(0.3718)	17	13.2827	2	19	8
IS	0.1047	(0.4117)	12	12.6833	8	20	9	Pokerface	0.1709	(0.3403)	1	13.9678	1	2	1