

Appeared in the 4th International Conference on Audio- and Video-Based Biometric Person Authentication, pp 10–18, June 9 - 11, 2003, Guildford, UK

An Image Preprocessing Algorithm for Illumination Invariant Face Recognition

Ralph Gross and Vladimir Brajovic

The Robotics Institute, Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA 15213
{rgross,brajovic}@cs.cmu.edu

Abstract

Face recognition algorithms have to deal with significant amounts of illumination variations between gallery and probe images. State-of-the-art commercial face recognition algorithms still struggle with this problem. We propose a new image preprocessing algorithm that compensates for illumination variations in images. From a single brightness image the algorithm first estimates the illumination field and then compensates for it to mostly recover the scene reflectance. Unlike previously proposed approaches for illumination compensation, our algorithm does not require any training steps, knowledge of 3D face models or reflective surface models. We apply the algorithm to face images prior to recognition. We demonstrate large performance improvements with several standard face recognition algorithms across multiple, publicly available face databases.

1 Introduction

Besides pose variation, illumination is the most significant factor affecting the appearance of faces. Ambient lighting changes greatly within and between days and among indoor and outdoor environments. Due to the 3D shape of the face, a direct lighting source can cast strong shadows that accentuate or diminish certain facial features. Evaluations of face recognition algorithms consistently show that state-of-the-art systems can not deal with large differences in illumination conditions between gallery and probe images [1–3]. In recent years many appearance-based algorithms have been proposed to deal with the problem [4–7]. Belhumeur showed [5], that the set of images of an object in fixed pose but under varying illumination forms a convex cone in the space of images. The illumination cones of human faces can be approximated well by low-dimensional linear subspaces [8]. The linear subspaces are typically estimated from training data, requiring multiple images of the object under different illumination conditions. Alternatively, model-based approaches have been proposed to address the problem. Blanz et al. [9] fit a previously constructed morphable 3D model to single images. The algorithm works well across pose and illumination, however, the computational expense is very high.

In general, an image $I(x, y)$ is regarded as product $I(x, y) = R(x, y)L(x, y)$ where $R(x, y)$ is the reflectance and $L(x, y)$ is the illuminance at each point (x, y)

[10]. Computing the reflectance and the illuminance fields from real images is, in general, an ill-posed problem. Therefore, various assumptions and simplifications about L , or R , or both are proposed in order to attempt to solve the problem. A common assumption is that L varies slowly while R can change abruptly. For example, homomorphic filtering [11] uses this assumption to extract R by high-pass filtering the logarithm of the image.

Closely related to homomorphic filtering is Land's "retinex" theory [12]. The retinex algorithm estimates the reflectance R as the ratio of the image $I(x, y)$ and its low pass version that serves as estimate for $L(x, y)$. At large discontinuities in $I(x, y)$ "halo" effects are often visible. Jobson [13] extended the algorithm by combining several low-pass copies of the logarithm of $I(x, y)$ using different cut-off frequencies for each low-pass filter. This helps to reduce halos, but does not eliminate them entirely.

In order to eliminate the notorious halo effect, Tumblin and Turk introduced the low curvature image simplifier (LCIS) hierarchical decomposition of an image [14]. Each component in this hierarchy is computed by solving a partial differential equation inspired by anisotropic diffusion [15]. At each hierarchical level the method segments the image into smooth (low-curvature) regions while stopping at sharp discontinuities. The algorithm is computationally intensive and requires manual selection of no less than 8 different parameters.

2 The Reflectance Perception Model

Our algorithm is motivated by two widely accepted assumptions about human vision: 1) human vision is *mostly* sensitive to scene reflectance and *mostly* insensitive to the illumination conditions, and 2) human vision responds to local changes in contrast rather than to global brightness levels. These two assumptions are closely related since local contrast is a function of reflectance.

Having these assumptions in mind our goal is to find an estimate of $L(x, y)$ such that when it divides $I(x, y)$ it produces $R(x, y)$ in which the local contrast is appropriately enhanced. In this view $R(x, y)$ takes the place of *perceived sensation*, while $I(x, y)$ takes the place of the *input stimulus*. $L(x, y)$ is then called *perception gain* which maps the input sensation into the perceived stimulus, that is:

$$I(x, y) \frac{1}{L(x, y)} = R(x, y) \quad (1)$$

With this biological analogy, R is *mostly* the reflectance of the scene, and L is *mostly* the illumination field, but they may not be "correctly" separated in a strict physical sense. After all, humans perceive reflectance details in shadows as well as in bright regions, but they are also cognizant of the presence of shadows. From this point on, we may refer to R and L as reflectance and illuminance, but they are to be understood as the *perceived sensation* and the *perception gain*, respectively.

To derive our model, we turn to evidence gathered in experimental psychology. According to Weber's Law the sensitivity threshold to a small intensity

change increases proportionally to the signal level [16]. This law follows from experimentation on brightness perception that consists of exposing an observer to a uniform field of intensity I in which a disk is gradually increased in brightness by a quantity ΔI . The value ΔI from which the observer perceives the existence of the disk against the background is called brightness discrimination threshold. Weber noticed that $\frac{\Delta I}{I}$ is constant for a wide range of intensity values. Weber's law gives a theoretical justification for assuming a logarithmic mapping from input stimulus to perceived sensation (see Figure 1(a)).

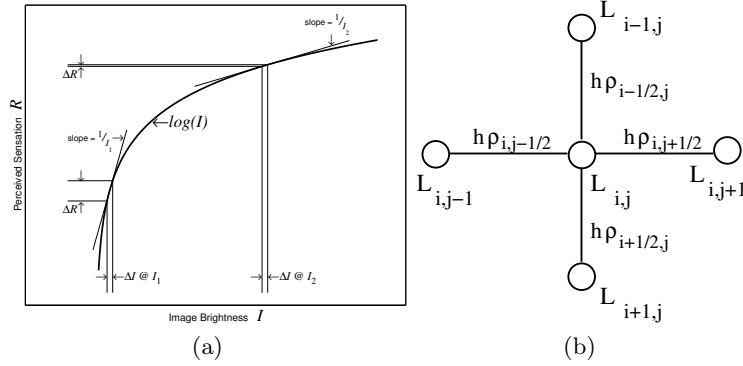


Fig. 1. (a) Compressive logarithmic mapping emphasizes changes at low stimulus levels and attenuates changes at high stimulus levels. (b) Discretization lattice for the PDE in Equation (5).

Due to the logarithmic mapping when the stimulus is weak, for example in deep shadows, small changes in the input stimulus elicit large changes in perceived sensation. When the stimulus is strong, small changes in the input stimulus are mapped to even smaller changes in perceived sensation. In fact local variations in the input stimulus are mapped to the perceived sensation variations with the gain $\frac{1}{I}$, that is:

$$I(x, y) \frac{1}{I_\Psi(x, y)} = R(x, y), \quad (x, y) \in \Psi \quad (2)$$

where $I_\Psi(x, y)$ is the stimulus level in a small neighborhood Ψ in the input image. By comparing Equation (1) and (2) we arrive at the model for the perception gain:

$$L(x, y) = I_\Psi(x, y) \doteq I(x, y) \quad (3)$$

where the neighborhood stimulus level is by definition taken to be the stimulus at point (x, y) . As seen in Equation 4 we regularize the problem by imposing a smoothness constraint on the solution for $L(x, y)$. The smoothness constraint

takes care of producing I_Ψ ; therefore, the replacement by definition of I_Ψ by I in Equation 3 is justified. We do not need to specify any particular region Ψ . The solution for $L(x, y)$ is found by minimizing:

$$J(L) = \iint_{\Omega} \rho(x, y)(L - I)^2 dx dy + \lambda \iint_{\Omega} (L_x^2 + L_y^2) dx dy \quad (4)$$

where the first term drives the solution to follow the perception gain model, while the second term imposes a smoothness constraint. Here Ω refers to the image. The parameter λ controls the relative importance of the two terms. The space varying permeability weight $\rho(x, y)$ controls the anisotropic nature of the smoothing constraint.

The Euler-Lagrange equation for this calculus of variation problem yields:

$$L + \frac{\lambda}{\rho}(L_{xx} + L_{yy}) = I \quad (5)$$

Discretized on a rectangular lattice, this linear partial differential equation becomes:

$$L_{i,j} + \lambda \left[\frac{1}{h\rho_{i,j-\frac{1}{2}}} (L_{i,j} - L_{i,j-1}) + \frac{1}{h\rho_{i,j+\frac{1}{2}}} (L_{i,j} - L_{i,j+1}) + \frac{1}{h\rho_{i-\frac{1}{2},j}} (L_{i,j} - L_{i-1,j}) + \frac{1}{h\rho_{i+\frac{1}{2},j}} (L_{i,j} - L_{i+1,j}) \right] = I \quad (6)$$

where h is the pixel grid size and the value of each ρ is taken in the middle of the edge between the center pixel and each of the corresponding neighbors (see Figure 1(b)). In this formulation, ρ controls the anisotropic nature of the smoothing by modulating permeability between pixel neighbors. Equation 6 can be solved numerically using multigrid methods for boundary value problems [17]. Multigrid algorithms are fairly efficient having complexity $O(N)$, where N is the number of pixels [17]. Running our non-optimized code on a 2.4GHz Pentium 4 produced execution times of 0.17 seconds for a 320x240-pixel image, and 0.76 seconds for a 640x480-pixel image.

The smoothness is penalized at every edge of the lattice by weights ρ (see Figure 1(b)). As stated earlier, the weight should change proportionally with the strength of the discontinuities. We need a *relative* measure of local contrasts that will equally "respect" boundaries in shadows and bright regions. We call again upon Weber's law and modulate the weights ρ by Weber's contrast ¹

$$h\rho_{\frac{a+b}{2}} = \frac{\Delta I}{I} = \frac{|I_a - I_b|}{\min(I_a, I_b)} \quad (7)$$

where $\rho_{\frac{a+b}{2}}$ is the weight between two neighboring pixels whose intensities are I_a and I_b .

¹ In our experiments equally good performance can be obtained by using Michelson's contrast $(I_a + I_b)/(I_a - I_b)$.

3 Face Recognition Across Illumination

3.1 Databases and Algorithms

We use images from two publicly available databases in our evaluation: CMU PIE database and Yale database. The *CMU PIE* database contains a total of 41,368 images taken from 68 individuals [18]. The subjects were imaged in the CMU 3D Room using a set of 13 synchronized high-quality color cameras and 21 flashes. For our experiments we use images from the more challenging illumination set which was captured without room lights (see Figure 2).

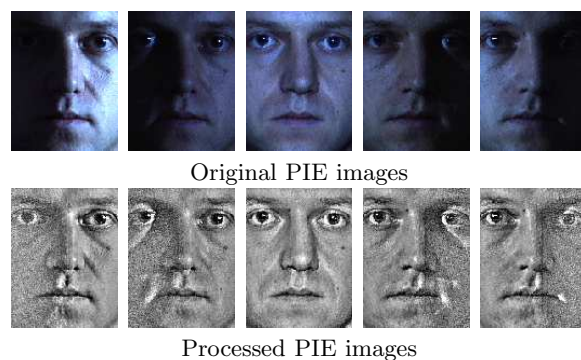


Fig. 2. Result of removing illumination variations with our algorithm for a set of images from the PIE database.

The Yale Face Database B [6] contains 5760 single light source images of 10 subjects each seen under 576 viewing conditions: 9 different poses and 64 illumination conditions. Figure 3 shows examples for original and processed images. The database is divided into different subsets according to the angle the light source direction forms with the camera's axis (12° , 25° , 50° and 77°)

We report recognition accuracies for two algorithms: Eigenfaces (Principal Component Analysis (PCA)) and FaceIt, a commercial face recognition system from Identix. Eigenfaces [19] is a standard benchmark for face recognition algorithms [1]. FaceIt was the top performer in the Facial Recognition Vendor Test 2000 [2]. As comparison we also include results for Eigenfaces on histogram equalized and gamma corrected images.

3.2 Experiments

The application of our algorithm to the images of the CMU PIE and Yale databases results in accuracy improvements across *all conditions and all algorithms*. Figure 4 shows the accuracies of both PCA and FaceIt for all 13 poses of the PIE database. In each pose separately the algorithms use one illumination condition as gallery and all other illumination conditions as probe. The reported

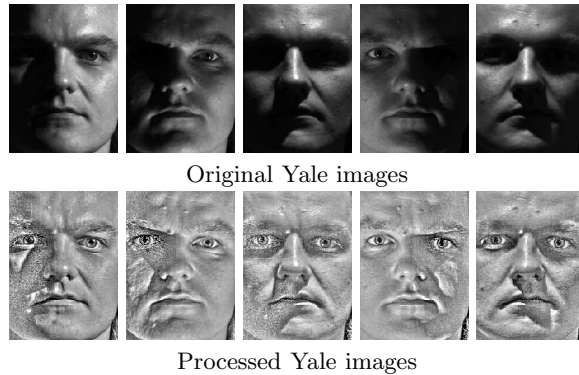


Fig. 3. Example images from the Yale Face Database B before and after processing with our algorithm.

results are averages over the probe illumination conditions in each pose. The performance of PCA improves from 17.9% to 48.6% on average across all poses. The performance of FaceIt improves from 41.2% to 55%. On histogram equalized and gamma corrected images PCA achieves accuracies of 35.7% and 19.3%, respectively.

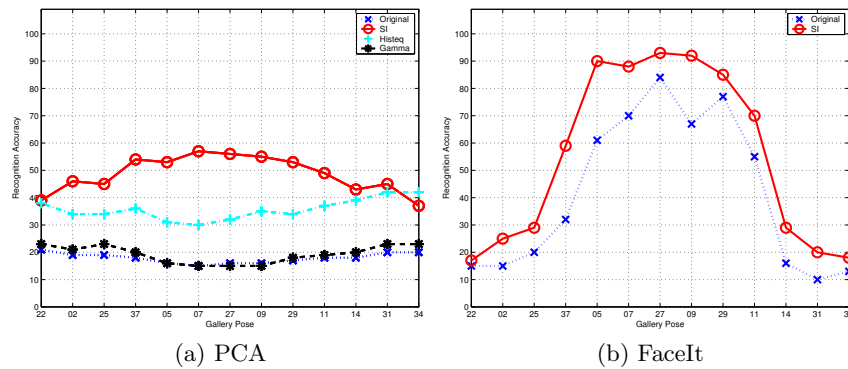


Fig. 4. Recognition accuracies on the PIE database. In each pose separately the algorithms use one illumination condition as gallery and all other illumination conditions as probe. Both PCA and FaceIt achieve better recognition accuracies on the images processed with our algorithm (SI) than on the original. The gallery poses are sorted from right profile (22) to frontal (27) and left profile (34).

Figure 5 visualizes the recognition matrix for PCA on PIE for frontal pose. Each cell of the matrix shows the recognition rate for one specific gallery/probe illumination condition. It is evident that PCA performs better in wide regions of the matrix for images processed with our algorithm. For comparison the recognition matrix for histogram equalized images is shown as well.

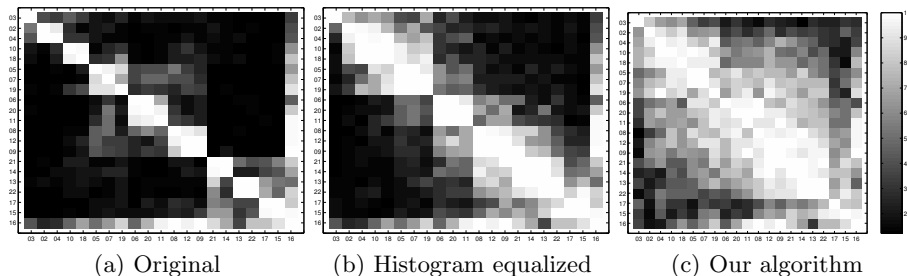


Fig. 5. Visualization of PCA recognition rates on PIE for frontal pose. Gallery illumination conditions are shown on the y-axis, probe illumination conditions on the x-axis, both spanning illumination conditions from the leftmost illumination source to the rightmost illumination source.

We see similar improvements in recognition accuracies on the Yale database. In each case the algorithms used Subset 1 as gallery and Subsets 2, 3 and 4 as probe. Figure 6 shows the accuracies for PCA and FaceIt for Subsets 2, 3 and 4. For PCA the average accuracy improves from 59.3% to 93.7%. The accuracy of FaceIt improves from 75.3% to 85.7%. On histogram equalized and gamma corrected images PCA achieves accuracies of 71.7% and 59.7%, respectively.

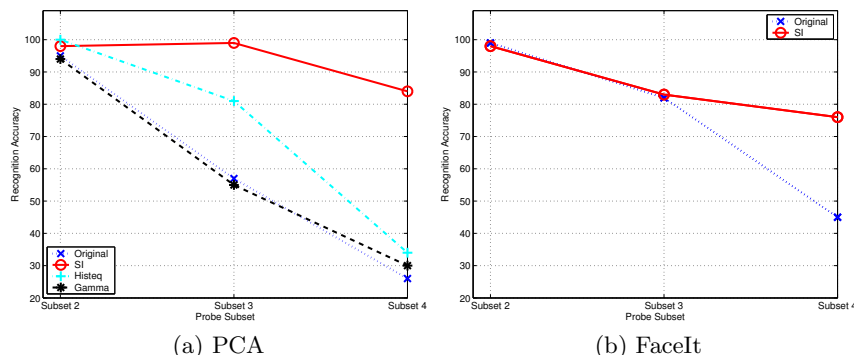


Fig. 6. Recognition accuracies on the Yale database. Both algorithms used images from Subset 1 as gallery and images from Subset 2, 3 and 4 as probe. Using images processed by our algorithm (SI) greatly improves accuracies for both PCA and FaceIt.

4 Conclusion

We introduced a simple and automatic image-processing algorithm for compensation of illumination-induced variations in images. The algorithm computes the estimate of the illumination field and then compensates for it. At the high level, the algorithm mimics some aspects of human visual perception. If desired, the user may adjust a single parameter whose meaning is intuitive and simple to un-

derstand. The algorithm delivers large performance improvements for standard face recognition algorithms across multiple face databases.

Acknowledgements

The research described in this paper was supported in part by National Science Foundation grants IIS-0082364 and IIS-0102272 and by U.S. Office of Naval Research contract N00014-00-1-0915.

References

1. Phillips, P., Moon, H., Rizvi, S., Rauss, P.: The FERET evaluation methodology for face-recognition algorithms. *IEEE PAMI* **22** (2000) 1090–1104
2. Blackburn, D., Bone, M., Philips, P.: Facial recognition vendor test 2000: evaluation report (2000)
3. Gross, R., Shi, J., Cohn, J.: Quo vadis face recognition? In: Third Workshop on Empirical Evaluation Methods in Computer Vision. (2001)
4. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE PAMI* **19** (1997) 711–720
5. Belhumeur, P., Kriegman, D.: What is the set of images of an object under all possible lighting conditions. *Int. J. of Computer Vision* **28** (1998) 245–260
6. Georgiades, A., Kriegman, D., Belhumeur, P.: From few to many: Generative models for recognition under variable pose and illumination. *IEEE PAMI* (2001)
7. Riklin-Raviv, T., Shashua, A.: The Quotient image: class-based re-rendering and recognition with varying illumination conditions. In: *IEEE PAMI*. (2001)
8. Georgiades, A., Kriegman, D., Belhumeur, P.: Illumination cones for recognition under variable lighting: Faces. In: *Proc. IEEE Conf. on CVPR*. (1998)
9. Blanz, V., Romdhani, S., Vetter, T.: Face identification across different poses and illumination with a 3D morphable model. In: *IEEE Conf. on Automatic Face and Gesture Recognition*. (2002)
10. Horn, B.: *Robot Vision*. MIT Press (1986)
11. Stockam, T.: Image processing in the context of a visual model. *Proceedings of the IEEE* **60** (1972) 828–842
12. Land, E., McCann, J.: Lightness and retinex theory. *Journal of the Optical Society of America* **61** (1971)
13. Jobson, D., Rahman, Z., Woodell, G.: A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Trans. on Image Processing* **6** (1997)
14. Tumblin, J., Turk, G.: LCIS: A boundary hierarchy for detail-preserving contrast reduction. In: *ACM SIGGRAPH*. (1999)
15. Perona, P., Malik, J.: Scale-space and edge detection using anisotropic diffusion. *IEEE PAMI* **12** (1990) 629–639
16. Wandel, B.: *Foundations of Vision*. Sunderland MA: Sinauer (1995)
17. Press, W., Teukolsky, S., Vetterling, W., Flannery, B.: *Numerical Recipes in C*. Cambridge University Press (1992)
18. Sim, T., Baker, S., Bsat, M.: The CMU Pose, Illumination, and Expression (PIE) database. In: *IEEE Int. Conf. on Automatic Face and Gesture Recognition*. (2002)
19. Turk, M., Pentland, A.: Eigenfaces for recognition. *Journal of Cognitive Neuroscience* **3** (1991) 71–86