Expert Systems With Applications 62 (2016) 347–357

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Multiscale morphology based illumination normalization with enhanced local textures for face recognition

Sanghun Lee, Chulhee Lee*

School of Electrical and Electronic Engineering, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul, Republic of Korea

ARTICLE INFO

Article history:
Received 22 March 2016
Revised 9 May 2016
Accepted 17 June 2016
Available online 18 June 2016

Keywords:
Face recognition
Illumination normalization
Texture information
Quotient image

ABSTRACT

A key challenge of face recognition is to obtain illumination invariant face images while preserving the discriminative features. The locations and shapes of small-scale features (e.g. eyebrows, eyes, nostrils, a mouth, etc.) are usually treated as key features for face recognition. However, it has also been observed that the local texture information of facial regions contains intrinsic facial features and needs to be enhanced to improve performance. To compensate for the illumination effects that appeared while extracting both the small-scale features and the texture information, we used multiscale morphological techniques. We used a generalized dynamic morphological quotient image (GDMQI) method based on Retinex theory and multiscale morphological closing to solve the artifact problem discussed in previous works. The proposed method consisted of two main steps: (i) illumination estimation and (ii) texture enhancement. The proposed method showed improved performance when using the CMU PIE, AR and Extended Yale-B databases.

1. Introduction

Face recognition is increasingly being used in intelligent surveillance systems, identity authentication for security systems, face database matching and other identification systems. However, uncontrolled and various lighting conditions can cause unpredictable illumination effects on faces (Zou, Kittler, & Messer, 2007) such as atypical shadow regions, which can make robust face recognition quite difficult.

1.1. Existing models and methods

Researchers have tried to model the human visual system by using lightness algorithms (Blake, 1985; Horn, 1973; Land, 1983, 1986; Land & McCann, 1971), such as Land's Retinex theory. A popular model, which was formulated as the intensity equation in (Hurlbert, 1986), describes the idea that the intensity signal of an object captured by light sensors is the product of illumination and reflectance. Such models have led to the development of several algorithms for illumination-invariant representations of images.

Illumination compensation is an ill-posed problem since the levels of both illumination and reflectance are unknown in most cases. Thus, some assumptions are necessary to simplify the problem. The intensity equation is often simplified as a Lambertian reflectance model (Basri & Jacobs, 2003) under the assumption that a surface reflects lights uniformly in all directions. Belhumeur and Kriegman showed that if a convex object is a Lambertian reflector, then an illumination cone (i.e. the number of distinct surface normal values) can be obtained from three properly chosen images (Belhumeur & Kriegman, 1998). Based on the illumination cone model, Georghiades et al. used a set of training face images to generate a complete model of the Lambertian reflectance face map (Georghiades, Belhumeur, & Kriegman, 2001; Georghiades, Kriegman, & Belhumeur, 1998), but this required several well-posed face images under various illumination conditions. Similarly, a quotient image (QI) (Shashua & Riklin-Raviv, 2001) was computed from a probe image and three bootstrap images under different lighting conditions.

Another common assumption is that illumination varies smoothly and is locally constant (Jobson, Rahman, & Woodell, 1997a; Land, 1983, 1986; Land & McCann, 1971; Wang, Li, & Wang, 2004; Zhang, Tian, He, & Yang, 2007). Jobson et al. introduced a single-scale retinex (SSR) model (Jobson, Rahman, & Woodell, 1997b), within which the illumination component was estimated as a smoothed original image with a normalized Gaussian function. This approach was extended to a multi-scale retinex (MSR) model (Jobson et al., 1997a), which used a weighted sum of the SSRs. However, the assumption that the amount of illumination varied smoothly was invalid for the shadow boundaries. Generally, in self-quotient images (SQI) (Wang et al., 2004), an anisotropic filter (a weighted Gaussian filter) is used to preserve shadow edges

* Corresponding author. Fax: +82 2 312 4584.
E-mail addresses: nica@yonsei.ac.kr (S. Lee), chulhee@yonsei.ac.kr (C. Lee).

http://dx.doi.org/10.1016/j.eswa.2016.06.039
0957-4174/

while estimating the illumination component. Unlike the QI, the SQI method requires no additional alignment and bootstrap images, but its over-compensated regions may lower recognition performance. Chen et al. used the total variation model (Rudin, Osher, & Fatemi, 1992; Strong & Chan, 2003) which is well known for its edge-preserving properties, and proposed a total variation based quotient image method (TVQI) (Chen, Yin, Zhou, Comaniciu, & Huang, 2005, 2006). In spite of its good performance, its iterative optimization proved to be very time-consuming. Recently, by using a double-density dual-tree complex wavelet transform method (DD-DTCWT), Baradarani et al. decomposed an image into frequency subbands to estimate the illumination component (Baradarani et al., 2013). Unlike when using other methods, the amount of reflectance is directly estimated in the Weber-face (WF) method (Wang, Li, Yang, & Liao, 2011) by computing a ratio image between the center pixel and the difference of the center pixel from its neighboring pixels. Wu et al. extended the Weber-face method by using multi-scale patches, which is known as the weighted generalized Weber-face (wGWF) method (Wu et al., 2014).

On the other hand, finding local descriptors for facial features is one of the most popular approaches. The local binary pattern (LBP), introduced by Ojala et al., is a nonparametric local texture descriptor that represents the relationship between a pixel and its neighboring pixels (Ojala, Pietikäinen, & Mäenpää, 2002). Ahonen et al. applied the LBP to face recognition by dividing facial images into 7×7 windows and constructing a histogram for each window (Ahonen, Hadid, & Pietikäinen, 2004; Ahonen, Hadid, & Pietikainen, 2006). In the local ternary pattern (LTP) (Tan & Triggs, 2007, 2010), the LBP noise problem was resolved by using 3-valued codes and a user-specified threshold. Recently, Liu et al. used a local histogram specification (LHS-L) method (Liu, Yang, Gao, & Cui, 2014) to eliminate high frequency illumination features that were not considered in (Tan & Triggs, 2007, 2010).

1.2. Multiscale morphological methods

Because of their edge-preserving nature, multiscale morphological operations have been used in several applications: noise removal (Mukhopadhyay & Chanda, 2002), segmentation (Mukhopadhyay & Chanda, 2003), and medical imaging (Jackway, 1996). For face recognition, Zhang et al. proposed a dynamic morphological quotient image (DMQI) (Zhang et al., 2007), based on multiscale morphological closing operations. When estimating the amount of illumination in the DMQI, the edge-preserving property of morphological closings is helpful for avoiding halo artifacts on the shadow boundaries. Basically, the DMQI is based on the Retinex theory (Land & McCann, 1971), in which an observed image $I(x, y)$ is decomposed into an illumination image $L(x, y)$ and a reflectance image $R(x, y)$, such that $I(x,y) = L(x,y)R(x,y)$. For a given face image, the authors in (Zhang et al., 2007) estimated the illumination $L(x, y)$ by combining morphological closings with various sizes of structuring elements. The conditional function $DClose(x, y)$ was defined as:

$$DClose(x,y) = \begin{cases} C_l(x,y) & C_l(x,y) > \alpha C_s(x,y) \\ C_s(x,y) & C_l(x,y) < \beta C_s(x,y) \\ C_m(x,y) & \text{otherwise} \end{cases} \quad (1)$$

where $C_u = I \bullet T_{u\times u}$ and $\bullet$ denotes a morphological closing with a structuring element $T_{u\times u}$ on a gray-scale image. l, m and s represent the sizes of the large, medium, and small structuring elements, respectively ($l > m > s$). α and β are heuristically selected coefficients with $\alpha > \beta > 1$. The DMQI can be defined as follows: $DMQI(x,y) = I(x,y)/DClose(x,y)$.

The idea is to use different sizes of structuring elements for estimating the amount of illumination and extracting the small-scale features. It is well known that morphological closing suppresses dark details less than structuring elements (Gonzalez, 2009; Mukhopadhyay & Chanda, 2003). Therefore, we chose a structuring element larger than the small-scale features to suppress the dark details. On the other hand, a morphological closing operation with a small structuring element is good at estimating local illumination (Zhang et al., 2007). Based on this observation, the DMQI uses structuring elements of different sizes to handle objects of various sizes. However, grainy pixels appeared in some regions (such as in the shadows) since the pixel-wise division operation amplified small value changes in the dark areas. A grainy noise reduction algorithm (Zhang et al., 2007) was used as a post-processing procedure to reduce the number of grainy pixels.

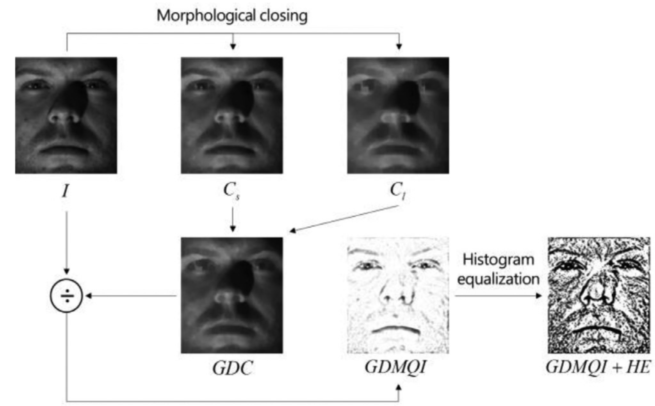

Fig. 1. Proposed illumination normalization.

However, we observed potential problems in the DMQI. First, the conditional function, used to combine the three images, produced some hard edges around the boundaries of the small-scale features (e.g. eyebrows, eyes, nostrils, mouth, etc.). Such edges are artifacts, which need to be suppressed to achieve improved performance. To address this problem, we used a generalized DMQI (GDMQI) with a logistic function for smooth transitions. Second, the texture information of the facial regions in both the DMQI and the GDMQI showed less influence than the small-scale features when we calculated the similarity between the two images. We applied histogram equalization and observed that the GDMQI combined with histogram equalization produced good performance when using the CMU PIE, AR and Extended Yale-B databases. As shown in Fig. 1, the proposed illumination normalization method consisted of two stages: (1) illumination estimation and (2) texture enhancement. To effectively deal with misaligned face images, a modified correlation coefficient was used as a similarity measure, which will be presented in Section 3.

2. Methodology

In this section, the proposed method will be discussed in detail. We begin with the introduction of a generalized form of DMQI. Also, we examine some characteristics of the GDMQI method, which can utilize the local texture information of facial regions.

2.1. Generalized DMQI (GDMQI)

As explained in Section 1, the common assumption used in the illumination-reflectance model is that the illumination $L(x, y)$ slowly varies and can be represented by low-frequency components. However, this assumption is not suitable for abrupt shadow boundaries. Accordingly, edge-preserving smoothing methods were used in (Chen et al., 2005, 2006; Wang et al., 2004). In our case, the DMQI preserved edges since the morphological closing