A Technique for Compensating Light Source Direction in Face Recognition Applications

PAULO SERGIO DE SOUZA COELHO[1], ANTONIO A. F. DE OLIVEIRA[1], CLAUDIO ESPERANÇA[1]

[1]Laboratório de Computação Gráfica
Programa de Engenharia de Sistemas - COPPE / UFRJ
psergio,oliveira,esperanc@lcg.ufrj.br

Abstract. A process to obtain the directional component of the main light source illuminating a human face in an image is described. In the first step of this process, two measures using low resolution data are obtained from the image. The value of these measures is then compared with those obtained from a synthetic model illuminated from different directions. This process, however, is not sufficient to indicate a single direction as is proved using a simplified model for face parts. An additional step is required to break the tie by using high frequency data. Once the light direction is determined, correction values are added to the image pixels to obtain an approximation of the face imaged under standard illumination conditions thereby making recognition much easier.

Keywords: Face recognition, light compensation, image databases.

1 Introduction

Automated Face Recognition is a topic of increasing interest in Computer Vision. The context of its application ranges from identifying criminals – filmed during a robbery, for instance – to recognizing authorized personal before granting access to restricted areas.

However, the comparison of a face image with those of stored in a database can be considerably more difficult if conditions like ambient and directional illumination and the pose of the face vary too much[1]. To reduce this problem, in a pre-processing phase one may try to transform the original image of a face into another depicting that same face imaged under normalized conditions. Obtaining such a corrected version requires that the conditions under which the original image was taken are detected. This work addresses specially the detection of the directional light source illuminating the face when the image was taken. That source is assumed to be unique and once its direction is estimated both the ambient light and correction values for each pixel intensity can also be calculated.

The image information collected is constrained to the regions of the forehead and the cheeks. Those regions admit reasonable planar approximations and the texture determined by men beard is mostly avoided. Bangs can be a real problem but, if the hair hanging over the forehead does not occlude a considerable part of it, one can try to exclude the pixels imaging both the bang and the shadow determined by it on the forehead. However, for the sake of simplicity, in this text, one will assume that bangs do not invade the forehead regions where measures are made.

The measures obtained from the image are compared with corresponding ones taken from frontal images of a synthetic face model illuminated from different directions. One must observe that according to [BRU 97] the face model does not need to be a very detailed one. The set of directions considered, $\Gamma = \{d_{kj} = (\theta_k, \varphi_j), k = 1, \ldots, K;\ j = 1, \ldots, J\}$, was determined so that the area of the spherical quadrilateral defined by $(\theta_i, \varphi_j), (\theta_i, \varphi_{j+1}), (\theta_{i+1}, \varphi_j)$ and $(\theta_{i+1}, \varphi_{j+1})$ is equal to $\pi^2/\ 648$ what in average corresponds to an interval of 5^o both in θ and in φ. The range of φ_j can be constrained to $[-60^o, 60^o]$ and θ_i must also have an upper bound $< 90^o$, as explained below. Let μ_{kj} be the vector whose coordinates are the measures obtained when the light illuminating the synthetic model comes from the direction d_{kj}. The pairs (d_{kj}, μ_{kj}) can be used to train a neural net for the task of reconstructing the function $\mu = f(d)$. f, which can be assumed to have the form $\sum_{m=1}^{M} c_m.e^{-(||W(\mu - t_m)||^2)}$. The purpose of the training phase is to obtain parameters c_m and $t_m, m = 1, \ldots, M$; and a correlation matrix W which minimizes a weighted sum of the quadratic errors $||d_{kj} - f(\mu_{kj})||^2$. In a more classical approach, an n^{th} - order Voronoi diagram of the set $\{\ \mu_{kj}, k = 1, \ldots, K;\ j = 1, \ldots, J\}$ is constructed. The direction illuminating an image is chosen among the d_{kj} corresponding to the n μ_{kj} closest to the image measures vector in a second phase thorough a selection process like that described in section 3. The reason for considering more than one solution is the possibility of different directions generating the same values

[1] Notice that this problem is similar, but not identical, to the problem of tracking under changing illumination conditions (see, for instance, [hag 96]).

for all measures used. In this case, given a vector of measures μ, although one might have a perfect reconstruction of the function f, it is impossible to decide among the directions in $f^{-1}(\mu)$. Nevertheless, if the cardinality(n) of that set is approximately known and if one has an estimate about how the directions in $f^{-1}(\mu)$ are distributed on the unit sphere, then it is possible to choose up to n candidates sufficiently separated from each other and subject them to another contest in which different criteria are employed. This is the approach used in this work. We prefer this approach rather than increasing the set of measures, mainly because we have not found other measures as robust as the two already used.

In the first phase of the process described here, two measures obtained from the image of the cheeks and the forehead are used to reduce the number of possible alternatives for the light direction to at most three. To decide among these directions the border of the shadow determined by each one of them on the model is examined. In the first step, low resolution information is processed, while the second step treats high frequency information in a much more constrained context. The measures used in the first step must be independent of the ambient light, the directional light intensity and the skin albedo since all these factors were kept equal to constant standard values when the model images were taken. The limit of three solutions used in the first part is justified in section 2 using a reasonable approximation of the face geometry. A possible implementation of the second phase is to place a snake [COH 91] along the border of the shadow determined on a cheek by any of the directions output by phase 1 and evaluating the energy spent until it converges to an contour in the original image. The lowest energy solution is, finally, the one indicated. Some difficulties of implementing that process are overviewed in [COH 91] and will not be discussed here.

Once the light direction is determined, a compensation value is calculated for each pixel on the face image. These compensation values are added to the pixel intensities in order to obtain an image of the face under standard illumination conditions. That correction makes it much easier to compare a face image with those of stored in a database. Those compensation values are first calculated for the synthetic model and then transferred to the face image pixels by means of the warping process as described in section 3.

2 The Approach Used in the First Phase

The four conditions below will be assumed henceforth:

$i)$ Besides the ambient light, there is a single source of directional light. When there are more than one of these sources, it may not be possible to determine their directions completely, even analyzing the shadow border on a region(a cheek, half the forehead) rather than considering only the total amount of light received by that region, as is done here.

$ii)$ There are no obstacles between that source and the face being imaged.

$iii)$ The interocular axis is horizontal or equivalently there is no asymmetry between the parts of the face on the two sides of the nose line. This is necessary since no pose estimation will be performed.

$iv)$ The angle between the light direction and the z-axis has an upper bound $U = \pi/2 - \varepsilon$. ε must be at least, sufficiently large to preclude the shadow of the chin from reaching the cheeks.

The two measures (M_1 and M_2) are obtained in the following way. The forehead is divided into three parts: left(F_l), central(F_c) and right (F_r). The range of the central part is from the middle of an eyebrow to the middle of the other. Let τ_l, τ_c and τ_r be the average pixel intensity in each one of these parts, considering the original image or a filtered version of it. The first measure (M_1) is given by $\dfrac{(\tau_l - \tau_r)}{(\tau_l + \tau_r - 2.\tau_c)}$. Regions C_l and C_r on each cheek are also considered. Let ϕ_l and ϕ_r be the average intensity of the pixels on those regions. The second measure (M_2) is obtained by $\dfrac{\phi_l - \tau_c}{\phi_r - \tau_c}$.

To identify the cheeks and forehead regions in the image of a face in a first step, eyes, eyebrows, nose and lips are identified by a process using Yow and Cippola's approach [YOW 97]. Cheeks are then determined from the position of the eyes, nose and the extremities of the lips. The forehead is identified from the eyebrows and the limits of the hair line.

Under the assumptions i-iv given above, the shadow on a cheek Q is determined only by the nose and the forehead. The shadow of the nose is limited by a line which can be reasonably approximated by two linear segments on the central y-z plane. In the simplified model (SM) used below, however, these two segments are replaced by a single one (see Figure 1).

The border of the forehead shadow is that of the eyebrow arch which starts at a junction with the nose line and can be considered to be part of an ellipsis parallel to the x-z plane. That border will be approximated by a line segment in SM.

A very simple illumination model will also be assumed. By this model the intensity of a pixel p is proportional to the following expression:

$$\rho(p).(I_a + E.\,\Delta(p,d). < \overrightarrow{n(p)},\, \overrightarrow{d} >) \quad \text{where:}$$

- $\rho(p)$ is the albedo of the area patch imaged in p,
- I_a is the intensity of ambient light,
- E is the intensity of the directional light source,

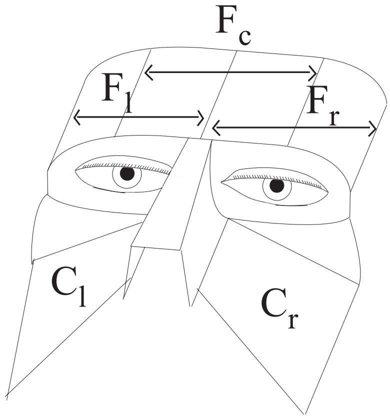

Figure 1: Simplified 3D model of a human face.

- $\Delta(p, d)$ is 1 if the face patch imaged in p is in the shade when the illuminant direction is d. Otherwise its value is 0.

- $\overrightarrow{n(p)}$ is the normal to the surface patch represented in p,

- $\overrightarrow{d} = (d_x, d_y, d_z)$ is the light source direction.

In the model SM, the central part of the forehead(F_c) is assumed to be a rectangle on the frontal plane. Let $[-a, a] \times [b, b+w] \times \{0\}$ be such a rectangle. The lateral parts of the forehead(F_r and F_l) are regions, symmetrical in relation to the plane $y - z$ consisting of bounded cylindrical hulls parallel to e_z. $F_r = \{(x, y, z) \mid (x, y) \in [a, a+l] \times [b, b+w]$ and $z = \sqrt{\frac{\rho.x^2}{(x^2 - \psi^2)}}\}$. To obtain F_l, simply replace $[a, a+l]$ by $[-a, -a-l]$.

The cheek regions C_r and C_l are considered to be planar in this model so as to simplify the estimation of the area of their parts which are in the shade. C_r is given by $\{(x, y, z) \mid (x, y) \in [c, c+w] \times [b, b+l]$ and $z = \alpha x + \beta y + \gamma\}$, $\alpha, \beta, \gamma \in R$. As mentioned above, for the sake of simplicity we replace the elliptical arch delimiting the shadow of F_s on C_s, $s = l, r$; by a straight line segment. The line delimiting the nose shadow on a cheek will be also supposed to be a linear segment.

Let $d = (d_x, d_y, d_z)$ be the illuminant direction being searched. The role of M_1 and M_2 in the determination of d can now be explained using the simplified model SM.

The value of τ_r can be written as follows:

$$\tau_r = \rho[E.(\lambda_x.d_x + \lambda_z.d_z) + A] \qquad (1)$$

where λ_x and λ_z are parameters of the model. Due to the symmetry between F_r and F_l in relation to the plane $y - z$, τ_l can be expressed by a similar equation obtained replacing λ_x by $-\lambda_x$ in equation (1). Thus, $\tau_l - \tau_r = 2.\rho.E.\lambda_x.d_x$ and, since there are no shadows on F_c, then $\tau_c = \rho.(E.d_z + A)$. Hence, $\tau_l + \tau_r - 2.\tau_c = 2.\rho.E.(\lambda_z - 1).d_z$ and therefore $\frac{d_x}{d_z}$ can be obtained from $M_1 = \frac{(\tau_l - \tau_r)}{(\tau_l + \tau_r - 2.\tau_c)}$ if the parameters λ_x and λ_z of the model are known.

Basically, M_2 is employed to determine $\frac{d_y}{d_z}$. Let us analyze the case where $d_y \leq 0$. The analysis of the other case is similar. If $d_y \leq 0$, considering all simplifications made above, then the part of C_r in the shade is the union of two quadrilaterals, one in the shadow of the nose and the other in the shadow of the forehead. Summing the area of those quadrilaterals we have an expression with the following form:

$$s_r = \frac{[r_1 + r_2(\frac{d_y}{d_z})]^2}{r_3 + r_4(\frac{d_y}{d_z})} + r_5(\frac{d_y}{d_z}) + r_6$$

where the $r_i, i = 1, \ldots, 6$ are expressions involving parameters of the model and $\frac{d_x}{d_z}$ which has already been determined from M_1. The non-linear term is a simple consequence of the fact that the nose line is not in general parallel to the "plane of the cheek".

The shade on the left cheek is determined only by the forehead and covers an area $s_l = q_1(\frac{d_y}{d_z}) + q_2$, where $q_i, i = 1, 2$ are also functions of the model parameters and $\frac{d_x}{d_z}$.

A vector normal to plane C_r is parallel to $(\alpha, 0, -1)$ and, by symmetry, also normal to plane C_l, is parallel to $(-\alpha, 0, -1)$. In view of that and considering the illumination model given above, we may infer that

$\phi_r = \rho[A + E/\sqrt{(1+\alpha^2)}(\alpha.d_x - d_z)(1 - s_r)]$ and $\phi_l = \rho[A + E/\sqrt{(1+\alpha^2)}(-\alpha.d_x - d_z)(1 - s_l)]$.

In the above equations we consider that the values of s_d and s_e have already been normalized in relation to the total area of each cheek.

Since $\tau_c = \rho.(E.d_z + A)$, then

$$M_2 = \frac{\phi_l - \tau_c}{\phi_r - \tau_c} = \frac{(\alpha\frac{d_x}{d_z} - 1)(1 - s_r) + \sqrt{(1+\alpha^2)}}{(-\alpha\frac{d_x}{d_z} - 1)(1 - s_l) + \sqrt{(1+\alpha^2)}} \qquad (2).$$

Replacing s_r and s_l by its expressions given above, one can transform 2 into a cubic equation in $\frac{d_y}{d_z}$. Analyzing the expression of the roots of that equation given by the Cardan's formula, one concludes that it can have more than one real solution even for not so extravagant values of the model parameters. For that reason having obtained a direction $(D = (D_x, D_y, D_z))$ through for instance, a trained neural net whose input is M_1 and M_2, we still search for other(two) good solutions reasonably distant from D around the plane $(\frac{x}{y} = \frac{D_x}{D_z})$.

3 Correcting the Image.

To calculate the values that have to be added to the intensity of each pixel in order to obtain an image of the face under standard illumination conditions, we propose the following two-step process:

Step 1: Obtain an image of the model illuminated from the direction indicated by the process described above. Then, perform a pixel-by-pixel subtraction between that image and an image of the same model under standard illumination conditions. Every pixel of the resulting image contains now a correction value.

Step 2: The correction values calculated in *step 1* are transferred to the original image by a warping transformation. Both the model image and the original image parts where the correction must be performed are partitioned by triangulations with the same topology. An example of such a triangulation is depicted in Figure 2. At first, a correspondence between the triangles in the two triangulations is established. Once that correspondence is obtained, the 1-to-1 mapping required by the warping process is defined by matching points in each pair of corresponding triangles based on their respective barycentric coordinates. Of course, anti-aliasing techniques have to be applied in some cases.

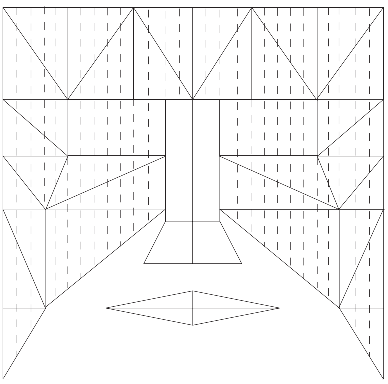

Figure 2: Example of a 2D triangulation of a face model

4 Conclusion

In this article, two measures are used to estimate the directional light component illuminating the image of a face. Those measures, although reasonably robust, are not sufficient to indicate a single solution in all cases. This determines the execution of a second phase where a different criterion is applied. In that phase, high frequency information is used, but that is not mandatory. Finding other measures which together with M_1 and M_2 can point out only one solution is also a possible alternative for perfecting the process.

5 Bibliography

[BRU 97] BRUNELLI, R. *Estimation of pose and illuminant direction for face processing*. Image and Vision Computing 15 (1997) 741-748.

[COH 91] COHEN, Laurent. *Note on active contour models and balloons.* CVGIP : Image Understanding 53 (1991) 211-218.

[HAG 96] HAGER, Gregory D. and BELHUMEUR, Peter N. *Real Time Tracking of Image Regions with Changes in Geometry and Illumination.* Proc. of the IEEE conf. on Computer Vision and Pattern Recognition (1996).

[YOW 97] YOW, Kin Chong and CIPOLLA, Roberto. *Feature-based human face detection*. Image and Vision Computing 15 (1997) 713-735.