Face Position and Angle Estimation Using Color Features and Normalized Moments

C. Emiliano Solórzano Espíndola, carlosemiliano04@gmail.com

ETSEIB- Universitat Politècnica de Catalunya

***Abstract*— The present works evaluates an approach for the estimation of relative position and angle estimation. The face is first located using color features in the RGB domain to locate the zones that represents skin. Once these zones are located, the first and second order statistical normalized moments are calculated for the estimation of position and angle. A new sub-image is then created centered on the face and oriented correctly.**

I. Introduction

Face detection is one of the most recurring subjects in computer vision, that's why there are many works published every year since the last few decades treating this subject, developing new algorithms depending on the goal of each one. This is because of the many applications that can be given to this, like face recognition, face tracking or create a 3D-model of a subject.

Also, face detection is still an area of active research since a completely successful approach or model has not be proposed to solve this problem. A human face is a dynamic object having high degree of variability in its appearance, which makes face detection a difficult problem in computer vision. In this field, accuracy and speed of identification is a main issue. Some of the most used face identification algorithms used nowadays, like Viola-Jones [1], depend on the face being in a range of relative angles with respect to the camera, thus not being rotational invariant.

In this work face detection is used to estimate the relative position and angle of the subject on a given color image. The first part is treated as a segmentation problem in which the areas corresponding to the human skin are identified. After that the face has to be located in this new fragments of skin by detecting some features in it.

Some of the present approaches used to detect the face are the following:

A. PCA (color)

The color of human skin is distinctive from the color of many other natural objects, hence color is a very important feature that can be used for face detection. Analyzing the skin-tone color statistics, one observes that skin colors are distributed over a small area in the chrominance plane and the major difference between skin tones is intensity [2]. Thus, the image is first converted into a color space capable of separating into a luminance channel and two chrominance components.

With the remaining components, a conditional Gaussian probability function is created where $p(w_{ij}|S)$ denotes the probability of belonging to the skin class *S*.

$$p(w_{ij}|S) = \frac{\exp\left[-\frac{1}{2}\left(w_{ij}-\mu_S\right)^T \sum_S \left(w_{ij}-\mu_S\right)\right]}{2\pi\left|\sum_S\right|^{\frac{1}{2}}}$$

Figure 1: Original image and probability image

Now a new image can be created with the probability of each pixel belonging to the skin. The major benefit of using the skin probability image is that a facial region is enhanced compared to the background and the influence of different lightning conditions is reduced since the skin probability doesn't depends on it.

After that PCA is applied to the resulting image. The approximation of a face using M "eigenfaces" is given by $x = P_M \, y + \bar{x}$ where P_M is the orthogonal projection matrix into the M-dimensional principal subspace. The residual reconstruction error $\in^2$ indicates how well the test pattern can be approximated in the face space. Thus, the “distance from face space" (DFFS) defined by equation can be used to determine if the image pattern represents a face.

B. Fleck and Forsyth algorithm (color)

In this case the process can be divided in two phases, the first one is skin detection and labeling and the second is find holes in the image that can represent the eyes.

In the first step we take the RGB image and from there extract the *log-opponent* chromacity representation, as in [3].

After that, it uses the texture amplitude values to find regions of low texture information, since the skin has a very smooth texture. In those regions, we further select skin based on measures of hue and saturation, so as their color matches that of the skin

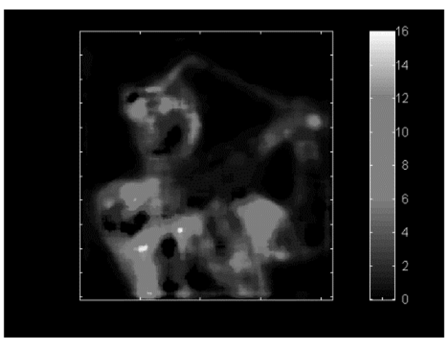

Figure 2. Texture map using the illumination values

Once that it has computed the texture map a threshold discriminates the parts of the image needed. The range that can describe the skin will depend also on the hue and saturation of the given image. If a pixel falls into the range it is marked as being skin in a binary map.

On the binary map the regions can be label, after some binary operations. After that the original image without dilatation is subtracted and find where the holes are and create a new image with the absolute value, this holes can represent the eyes or other facial features, Once we have done this we can find which of the labeled groups is the nearest to this holes, and thus where the face is.

Figure 3. Labeled groups of skin. It can be seen that in the face group there are the holes of the nostrils that will give us where the face is.

II. Selected Method

Finally, the selected method was based on the works on normalized color as in [4]. In this case it is also based on the detection of skin using the conditional probability of representing skin

For the detection of the region that represent the face, the normalized RGB space (*rgb*) is first obtained, this allows to minimize the variation due to the illumination since the face is a highly curved surface. The skin was found that can be tracked with high efficiency using only the *r* and *g* components [5].

$$r = \frac{R}{R+G+B}, \quad g = \frac{G}{R+G+B}$$

Then a conditional probability densities for a pixel being part of the skin using the *r* and *g* channels histogram can be estimated. Then we can estimate the probability of being skin $p(Pix \in Skin|r,g)$ given the values in r and g with some examples and training. A Gaussian distribution can be used for the estimation of the pixel’s probability. The database used for the examples is the *RGB-Color database* provided by Bonn-Rhine-Sieg University of Applied Sciences [6].

$$p(Pix \in Skin|r,g) = \frac{p(r,g|Pix \in Skin)p(Pix \in Skin)}{p(r,g)}$$

After that, a secondary image can be created using a threshold in which can be seen the pixels that surpasses this threshold to be enough representative of humans skin. This technique allows to complete the first part that was segmentation.

Figure 4. Pixels that surpass the threshold representing the zones with skin

The face position tracking con be achieved using moments. The first moment represents the center of gravity and provides a robust estimation of the face position, the second moment can give an estimation of height, width and slant of the face. The first and second moment are used to normalize the position and orientation. The moments can be estimated given the formulas:

$$\mu_x = \frac{1}{S}\sum p(x,y) * x$$
$$\mu_y = \frac{1}{S}\sum p(x,y) * y$$
$${\sigma_x}^2 = \frac{1}{S}\sum p(x,y) * (x-\mu_x)^2$$
$${\sigma_y}^2 = \frac{1}{S}\sum p(x,y) * (y-\mu_y)^2$$
$${\sigma_{xy}}^2 = \frac{1}{S}\sum p(x,y) * (x-\mu_x) * (y-\mu_y)$$

From here the position can be normalized using those moments for finding the relative position of the face with respect to the camera and the most representative area that can contain the face [7]. The important area is then created using an ellipse with described by:

$$\frac{(x-\mu_x)^2}{{\sigma_x}^2} + \frac{(y-\mu_y)^2}{{\sigma_y}^2} = 1$$

It also is oriented using the second order moments to calculate the estimation of the relative inclination angle. This can be calculated as:

$$\Theta = \frac{1}{2}\tan^{-1}\left(\frac{2{\sigma_{xy}}^2}{{\sigma_y}^2 - {\sigma_x}^2}\right)$$

This estimation works in angles between $-\pi/4$ and $\pi/4$

Using all this information a new sub-image is reconstructed in which the face is projected in grayscale. In this one the second moments are also reoriented to give the new axis for cutting non important parts of the image and centered in the face. The criteria for this is using two times the reoriented second moments, since they give a representation of 95% of the data. This last part was implemented for a live tracking using an RGB *webcam* with a 640*480 pixels resolution.

Figure 5. Reorientation and cropping of the zone of interest

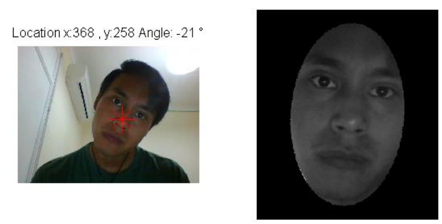

Figure 6. Ellipse with the region of interest representing 95% of the data

III. Evaluation

The first part of the method was proven to robust to illumination change even when a different camera and resolution of the image was used, but has different levels of confidence for different skin tones when using the same threshold and sometimes the hair is taken as part of the skin.

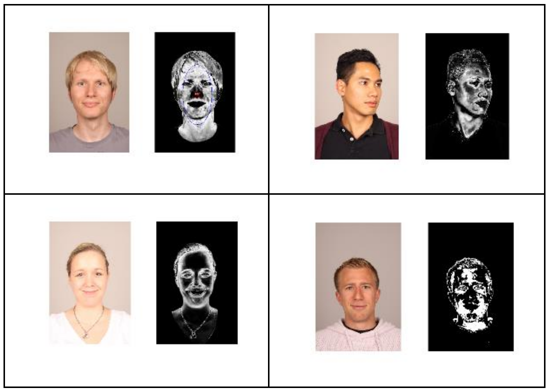

Figure 7.Response with the same threshold to images with different skin tones

This was enhanced by using a second probability density with the probability of belonging to the hair and subtracting it to the skin one multiplied by a parameter *b (the parameters were found experimentally).* Mathematically this can be represented as:

$$p(r,g|Pix \in Skin) = a * p(r,g|Pix \in OriginalSkin) - b * p(r,g|Pix \in Hair)$$

This can be visually represented as a complex Gaussian curve with the mixing of both distributions.

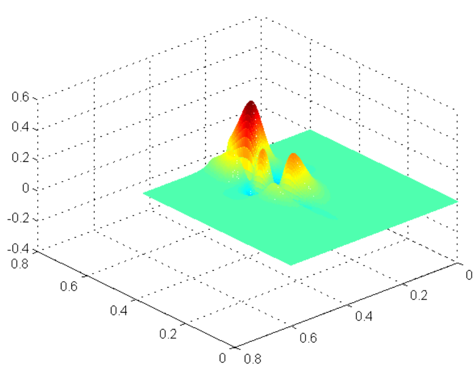

Figure 8. Gaussian mixture curve

Once the classified part is obtained, the next part was the position and angle estimation and the subsequent creation of a sub-image containing the face only. This part was already running on a webcam stream. The estimation was done right several times but with the limitations previously mentioned by the method. This is when the face is tilted more than 45°.

Some examples are here illustrated using different angles and illuminations:

Figure 9. Examples of the algorithm running

Some problems were presented here when the face was on the limits of the image, so the cropping couldn't be done properly, so it was decided that only the upper part of the head was needed, this allowed to eliminate part of the skin represented by the neck.

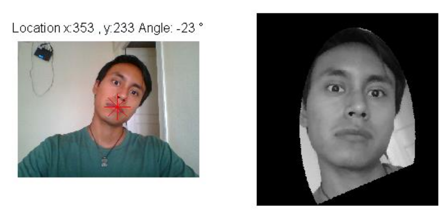

Figure 10. New crop using the upper part of the detected zone

IV. CONCLUSIONS

The algorithm was proven to be robust enough when working with different skin tonalities and in different illumination conditions, and faster like similar discriminant methods like the Mahalanobis classifier. The new sub-image represents accurately the face of the person and with the reorientation serves for posterior algorithms that are not rotational invariant like the face classification ones or *eigenfaces*. Even though it fails when some object of a skin-like color is present on the background or a second person is present on the same image.

There can be some methods to solve this like using a clustering algorithm before the calculations of moments and evaluate them for each clustered object, in this case, different faces.

Other limitation that the algorithm can face is the limitation of the angle estimation. But with a range of 90° of confidence serves good enough for many cases, especially for ones when the person is in front of the computer.

V. REFERENCES

[1] Paul Viola, Michael J Jones. (2004). Robust Real-Time Face Detection. In International Journal of Computer Vision 57, pp. 137-154, Netherlands.

[2] B. Menser and F. Mueller. (1999). Face detection in color images using principal components analysis. In Image Processing and Its Applications, 1999. Seventh International Conference on (Conf. Publ. No. 465), pp 620 - 624. Manchester: IET.

[3] Storring, M. (2004). Computer Vision and Human Skin Colour (Ph.D.). Faculty of Engineering and Science, Aalborg University.

[4] Gourier, N., Hall, D., & Crowley, J. (2004). Estimating Face Orientation from Robust Detection of Salient Facial Structures. FG NET WORKSHOP ON VISUAL OBSERVATION OF DEICTIC GESTURES.

[5] K. Schwerdt, J.L. Crowley. Robust face Tracking using Color. Automatic face and Gesture Recognition, pp 90-95, 2000.

[6] Steiner, Holger; Sporrer, Sebastian; Kolb, Andreas & Jung, Norbert: Design of an Active Multispectral SWIR Camera System for Skin Detection and Face Verification. Journal of Sensors, vol. 2015, Article ID 456368.

[7] Simon X. (1993). Image Analysis by Moments (PhD). Department of Electrical and Computer Engineering, University of Manitoba Winnipeg.