Automated Estimation of Human Age, Gender and Expression

EE368 Digital Image Processing Final Project Report

Yaoyu Tao
Department of Electrical Engineering, Stanford University
Stanford, CA 94305, USA
taoyaoyu@stanford.edu

***Abstract*—Recognition of facial variations has been a hot research area for the past decade. Age, gender and expression are three important facial variations attaining increasing attentions. This report presents a software system for automatic estimation of age, gender and expression. To make the system robust to any input image containing a face, a pre-processing stage is implemented to calibrate the face. Pre-processing stage also involves detection of eyes and nose. The age estimation subsystem is composed of aging feature extraction and feature classification. We use local binary patterns (LBP) and Gabor filter to describe the aging features. Feature classification is implemented based on histograms from LBP analysis. A minimum distance classifier is built upon training using images from FG-NET aging database. Furthermore, linear discriminant analysis (LDA) with fisher faces is used to identify gender and expression information. Experimental results based on training images demonstrate accuracies of 39.8%, 95.1% and 54.2% for age, gender and expression, respectively. Real-world testing also shows the estimation capability of this system.**

Keywords—Facial variations, age, gender, expression, automatic estimation, local binary pattern, linear discriminant analysis

I. Introduction

Human face, as a window to the soul, conveys a significant amount of nonverbal information that facilitates real-world human-to-human communications. People can easily extract many kinds of useful information from a face image, such as gender, expression, identity, approximate age, etc. Automatic extraction of facial information on a machine is attractive as well as challenging. Such machine can play an important role in many applications such as human-computer interaction, surveillance, content-based indexing and searching, biometrics and targeted advertising.

The problem has inspired researchers leading to a diverse set of solutions. Gender and expression identification have been studied extensively and can be identified accurately by simple linear discriminant analysis (LDA) with fisher faces [1]-[4]. However, directly applying LDA for age estimation usually results in poor performance, mainly due to the fact that LDA is insufficient to separate aging patterns.

Previous age estimation methods aiming for higher accuracy can be divided into two categories: active appearance model- (AAM-) based and non-AAM-based methods [5]. AAMs are statistical face models and have been used for age estimation; they involve modeling of face's shape and appearance, which are generated based on multiple feature points. Age is then considered as a function of feature vectors learned from AAMs.

Another approach to tackle this problem is non-AAM-based methods. The work in [6] used a single-level LBP operator to extract age features from skin regions of human face and the method has been shown effective on age classification. The estimation is fast but the accuracy is strongly dependent on the method of choosing skin regions (where the frequency components are extracted). An enhancement version of this method is proposed in [7] powered by sorted vector machine (SVM) combined with multi-level LBP and Gabor filter, offering more than 20% accuracy improvement. Compared to AAM methods, non-AAM methods have much lower complexity while still maintaining high accuracy, which makes them good candidates for practical adoptions.

In this report, we implemented an age estimator by extracting aging patterns with single level LBP and Gabor filter. This design choice yields good estimation accuracy with reasonable run time. We extracted aging patterns including skin features as well as wrinkles and lines. Single level LBP is applied to selected skin regions on a human's face that best represent age information. Gabor filter, on the other hand, extracts wrinkle and line information specifically. Commonly used FG-NET aging database is used to obtain these features under different ages upon which a minimum distance classifier is built. Gender and expression estimators are built separately based on LDA and databases are obtained from EE368 course website and mid-term exam. Experiments are carried out based on selected database images as well as real-world images taken during the poster session. This design has shown its good capability to estimate age, gender and expression.

II. Algorithms and Their Implementations

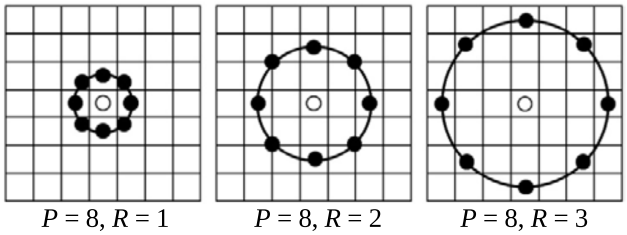


Fig. 1 Positions of neighboring according to P and R

For completeness, we briefly introduce LBP, Gabor filter and LDA algorithms implemented in this estimator and the rationales of using them.

A. Local Binary Pattern

LBP operator is one of the best performing texture descriptors and used widely in texture classification, segmentation, and face detection/recognition. It has the advantage of computational simplicity while still offering good performance.

As a person is getting older, facial blemishes such as freckles, age spots and fine wrinkles increase on the face skin. These micro-structures can be detected efficiently using LBP method. The basic concept of LBP is to assign a code to each pixel comparing it to its neighbors. The creation of the LBP code is expressed by the equation [] below

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p \text{, where } s(x) = \begin{cases} 1, x \geq 0 \\ 0, x < 0 \end{cases}$$

where P is the number of neighboring pixels, R is the distance from the center to the neighboring pixels, g_c is the gray value of the center pixel, $g_p(p = 1, \ldots, P\text{-}1)$ are the gray values of the p equally spaced pixels on the circle of radius R that forms a circularly symmetric neighbor set, and s is the threshold function of x.

The LBP feature vector can be created in the following manner [8]:

1. Divide the specified examined region into $m \times m$ cells (eg. 8×8 pixels for each cell).
2. For each pixel in a cell, compare the pixel to each of its n (eg. 8) neighbors. Follow the pixels along a circle. Examples are shown in Fig.1 for different combinations of P and R.

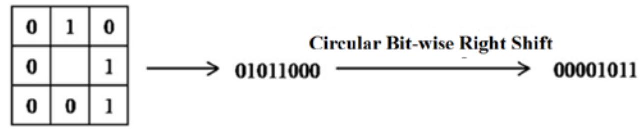


Fig.2 An example of generating n-digit binary number (n = 8)

3. If the center pixel's value is greater than the neighbor's value, assign a "1". Otherwise, assign a "0". This gives an n digits binary number. An example of n = 8 is shown below in Fig. 2.
4. Compute the histogram, over the cell, of the frequency of each "number" occurring.
5. Concatenate histograms of all cells. This gives the feature vector for the examined region.

Note that the feature vector can be processed using SVM or other machine-learning (ML) algorithms to classify images. Such classifiers can them be used for facial texture analysis for age estimation (In this design we implemented SVM but codes still have bugs).

B. Gabor Filter

Gabor filter is a linear filter commonly used for edge detection. Frequency and orientation representations of such filter are similar to those of human visual system [9], and they have been found to be particularly appropriate for texture representation and discrimination. [7] uses a set of Gabor filters with different frequencies and orientations that help extracting useful features such as wrinkles or lines from facial images.

The two dimensional Gabor filter in the spatial domain is defined by

$$g(x,y) = \frac{1}{2\pi\sigma_x\sigma_y}\exp[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right) + 2\pi jWx]$$

where σ_x and σ_y are the standard deviations of the x- and y-axes, respectively, and W is the radial frequency of the sinusoids. Similar to LBP analysis, examined regions have to be set properly for Gabor filters to extract valid wrinkle/line features. The region selection will be discussed in Section III.

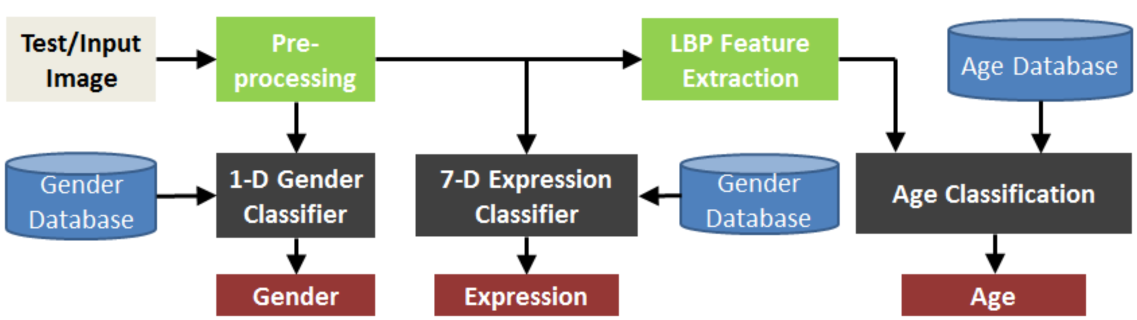
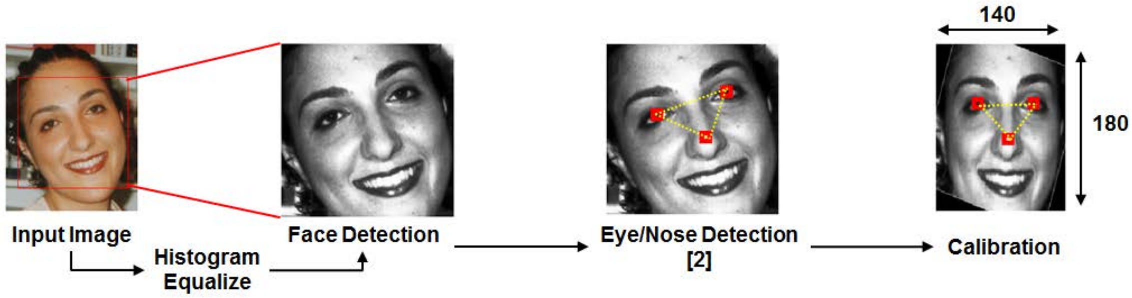


Fig. 3 Block Diagram of Automated Estimation of Age, Gender and Expression



(a) (b) (c) (d)

Fig. 4 Pre-processing Stages: Face Detection, Eye/Nose Detection and Calibration

C. *Linear Discriminant Analysis*

LDA is a generalization of Fisher's linear discriminant to find a linear combination of features that separates two or more classes of objects. It has been shown to perform well for facial gender and expression identification. Details of this method are covered during lectures and we will skip it here.

III. System Implementation

The implementation of proposed software system follows the block diagram as shown in Fig.3. The first step is to build the age, gender and expression classifiers based on database images. Gender and expression classifiers are built by applying 1-dimensional or multi-dimensional fisher LDA analysis on the database images. Age classifier, on the other hand, is built from LBP-based analysis plus Gabor filter (not shown in the block diagram as the system still have bugs when adding Gabor filter). Once we have the classifiers, pre-processing is a required stage to make the input image compatible with the classifier specifications. Particularly for age estimation, a LBP feature extraction stage is applied before age classifier. The rest of the section will be organized in three parts: 1) pre-processing stage, 2) database training stage and 3) estimation stage.

A. *Pre-processing Stage*

The pre-processing stage essentially transforms input face images of varying size to calibrated images ready for classifier and feature extraction. There are three steps in the pre-processing stage: face detection and image equalization, eye/nose detection, and calibration.

Step 1: Face detection and histogram equalization

Given an image containing a face to be estimated, the system first needs to detect face location and alleviate the effects of non-face regions. Face detection in our design is implemented with MATLAB vision system toolbox (CascadeObjectDetector) using Viola-Jones algorithm. It returns the bounding box values on the face region as shown in Fig. 4 (a) and (b). Note that histogram equalization is also applied to enhance the contrast.

Step 2: Eye/nose detection

After face detection and histogram equalization, we need to rotate the image so that face can be aligned in vertical direction. This can be done by first detect eyes' positions. The method we used here for eye detection is given in paper [10], [11][1], where an enhanced pictorial structure model is used for precise eye localization. Details for this method can be found in the two papers. The resulting image after eye/nose detection is shown in Fig. 4 (c), where two eyes as well as nose are highlighted by red square boxes.

Step 3: Calibration

Once we know the positions of two eyes, we are able to rotate the image. Assume the two eyes' positions are (x_1, y_1) and (x_2, y_2), then the rotation angle θ can be determined by

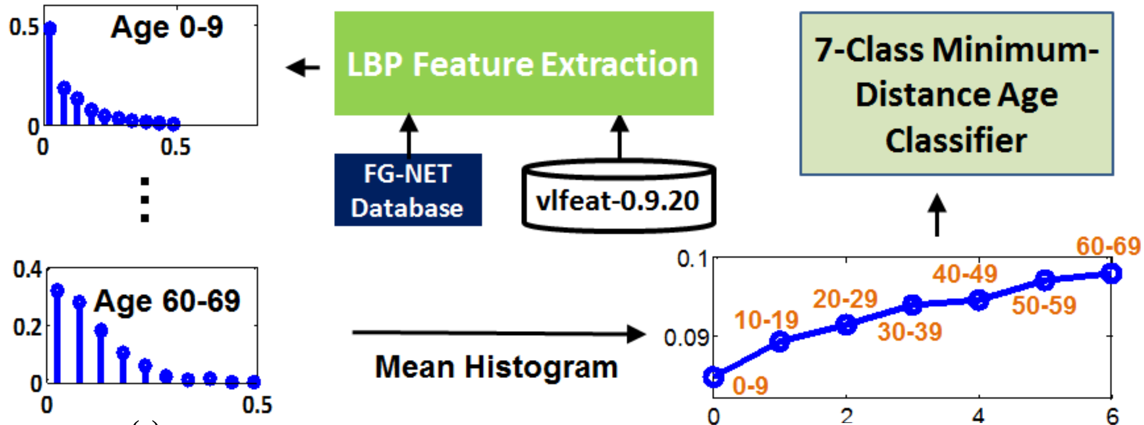
$$\theta = tan^{-1}\left(\frac{y_1 - y_2}{x_1 - x_2}\right)$$

We also set a standard image size of 140x180 for all images sent to the classifiers to make it easy for database training. Upon completion of image rotating, image resizing

Database		Class #	Image #	Image Size		Notes	
FG-NET Aging		**7**	**1002**	**444x489**		**68 points/image**	
Age Range	0-9	10-19	20-29	30-39	40-49	50-59	60-69
%	37.03	33.83	14.37	7.88	4.59	1.50	0.8
Gender		**2**	**400**	**201x246**		**Generate 1 fisher face**	
Expression		**7**	**147**	**120x150**		**Generate 8 fisher face**	
Range	Smile	Anger	Contempt	Digest	Fear	Sad	Surprise
%	23.8	16.3	3.4	21.8	7.5	6.8	32.0

Table I Database Summary

1 With the permission of the author Dr. Xiaoyang Tan, I used some of his original codes for the eye/nose detection.



(a) (b)

Fig. 5 Database Training for FG-NET

based on bilinear interpolation is applied to images that are not compatible with this standard. Image after the calibration stage is shown in Fig. 4(d). Note that calibrated eye/nose positions can be calculated from non-rotated image with equation below.

$$\begin{bmatrix} x_{rot} \\ y_{rot} \end{bmatrix} = \begin{bmatrix} cos\theta & sin\theta \\ -sin\theta & cos\theta \end{bmatrix} \times \left(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} x_{cent} \\ y_{cent} \end{bmatrix} \right) + \begin{bmatrix} x_{rot.cent} \\ y_{rot.cent} \end{bmatrix}$$

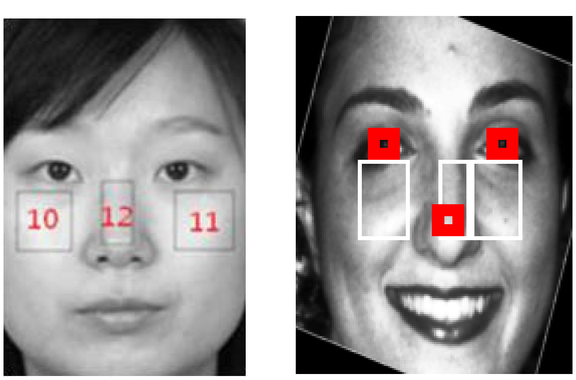
After calibration stage, image is ready to be sent to any of the classifiers for estimation.

B. Databased Training

Database training stage processes all databased images and forms classifiers for future test images. There are three separate classifiers in this system: LBP-based age classifier, LDA-based gender classifier and expression classifier. As mentioned in Section I, we don't use LDA classifier for age since its performance is much worse than LBP-based approach.

Particularly, we use FG-NET database for age estimation and databases from EE368 course materials for gender and expression recognition. Table I summarizes the database information as a reference.

1. FG-NET database training and age classifier



(a) Desired skin regions to extract features (Fig. 10 from [])

(b) Illustration of implemented region selection

Fig. 6 Region selection to extract skin features

The database training process for FG-NET database is shown in Fig. 5. FG-NET contains 1002 image with ages ranging from 0 to 69 as shown in Table I. We group this images into 7 groups with ages 0-9, 10-19, 20-29, …, 60-69. The first key problem to solve is to select proper regions to extract LBP vector and apply Gabor filter. Similar to [7], [12], [13], we decided to pick the regions as shown in Fig. 6 (a). It has been proved that aging patterns are especially significant in the labeled regions.

We first divide the entire face image into 8x8 cells. Since the eyes' and nose's positions are (x_1, y_1), (x_2, y_2) and (x_n, y_n), respectively, we can quickly determine the cell positions of these key components. Assume left eye corresponds to cell [x_{c1}, y_{c1}], right eye corresponds to cell [x_{c2}, y_{c2}], and nose corresponds to cell [x_{cn}, y_{cn}], we pick the following regions for skin feature extraction:

Region 1: [(x_{c1} – 2):(x_{c1} + 1), (y_{cn}: y_{c1})]

Region 2: [(x_{c2} – 1):(x_{c2} + 2), (y_{cn}: y_{c2})]

Region 3: [x_{cn} , (y_{cn}: y_{c1})]

Fig. 6(b) shows the region selection results of our implementation. We can see that the regions are accurately selected as desired.

Once region selection is completed, LBP and Gabor filter are applied to these regions. We use vlfeat-0.9.20 to compute the LBP vector. The resulting LBP vectors of each age range are then classified by SVM (results not shown since codes have bugs). An alternative approach is to directly associate the histogram means with ages. This is working because older people will have more wrinkles and lines and their skins tend to be leathery, leading to a larger portion of big LBP.values. Histograms with age 0-9 images and age 60-69 images are shown in Fig. 5 (a). We can clearly see that bins are spread towards bigger LBP values with older ages.

Based on this approach, we further run days of simulations on all the training images and plot the relationship between ages and histogram means, as shown in Fig. 5 (b). This shows a nicely proportional relationship between age and LBP

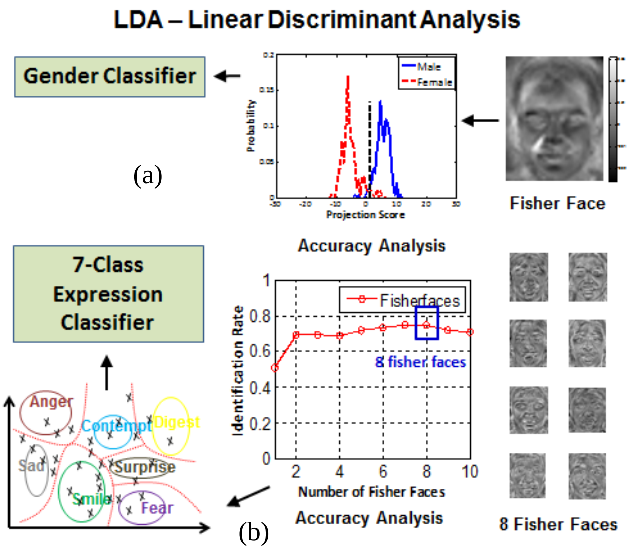


Fig. 7 Gender and expression classifier using LDA

histogram mean. A minimum distance is then built based on the histogram mean

On the other hand, Gabor filter is implemented as a helper in the age estimation to specifically emphasize on wrinkle features. We use a Gabor filter with 4 scales and 6 orientations. Due to time limited, we are not able to integrate the Gabor filter with correct functionality to the system.

2. Gender and expression database training

The training process for gender and expression follows standard LDA process taught in class. Fig. 7 (a) shows the two-class gender identification based on a single fisher face and Fig. (b) shows the 7-class expression identification based on 8 fisher faces. We pick 8 fisher faces for expression because it gives the highest accuracy over database images. Classifiers are then built from these LDA, as shown in Fig. 7.

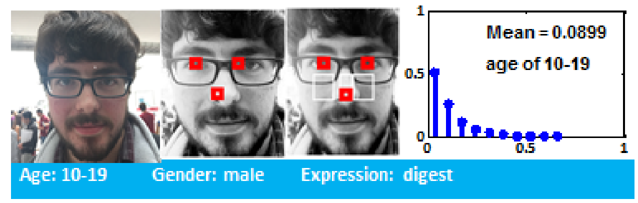
IV. Experiments and Analysis

A. *Experiment 1*

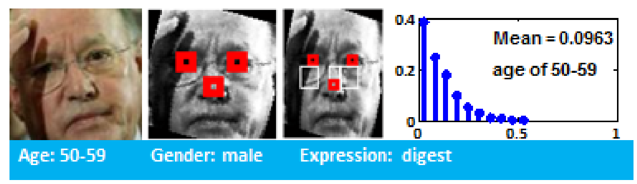
We first test system accuracy based on training database. The results are shown in Table II. Note that we manually split age database and expression database into male and female groups to get the gender identification accuracy. After fixing some mistakes and computational bugs, the final accuracies listed in Table II have some mismatches compared with those on the poster obtained from the first few runs.

From the experiments we can see that age estimation based

(a) Prof. Girod photo from personal website (I mean no offence)



(b) EE368 fellow classmate 1 (permitted orally to use photo during poster session)



(c) Image from YGA image database

Fig. 7 Testing with images outside the training database

on FG-NET database yields an accuracy of 46.4%, improving roughly 30% compared to the nature estimation rate of 15%. This accuracy is lower than the best estimation rates reported in reference papers [5]-[7], [12]-[13], due to the fact that we don't get SVM codes working correctly; however, the estimation accuracies are still considerably good. Gender estimation using single fisher face provides high accuracies over all the database images as expected. Expression estimation accuracy based on expression database is around 74.2%, and it can be further enhanced if we have more images in the database.

B. *Experiment 2*

We also test our system with some random input images not in the database. Fig. 8 shows 3 real-world estimation examples. Estimation results of age, gender and expression are considerably good. During the poster session, we also tried the system with pictures of EE368 fellow classmates. Fig. 8 (b) shows that our system can work well even with glasses on face as long as the glass area is not too big entering the LBP

Class/Accuracy	Age Database Images	Gender Database Images	Expression Database Images
Age	46.4%	N/A	N/A
Gender	81.3%	95.1%	83.4%
Expression	N/A	N/A	74.2%
All (exclude if N/A)	39.8%	95.1%	54.2%

Table II Experimental results based on 500 randomly-select training images examined regions. We can conclude that our system is able to accurately estimate age, gender and expression in real world.

ACKNOWLEDGMENT

Thanks for Prof. Bernd Girod, Prof. Gordon Wetzstein, Dr. Jean-Baptiste Boin and Huizhong Chen for teaching this wonderful class. Special thanks to mentor Huizhong Chen for giving me project advice. Thanks to Dr. Xiaoyang Tan for sharing ideas and implementation details on accurate eye/nose detection.

REFERENCES

[1] EE368 Course Website, http://web.stanford.edu/class/ee368/

[2] W. Zhao, R. Chellappa, P. J. Phillips, A. Rosenfeld, "Face Recognition: A literature Survey", *ACM Journal of Computing Surveys (CSUR)*. vol. 35, issue. 4, Dec 2003

[3] Jain, A.; Huang, J., "Integrating independent components and linear discriminant analysis for gender classification," *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on* , vol., no., pp.159,163, 17-19 May 2004

[4] Bekios-Calfa, J.; Buenaposada, J.M.; Baumela, L., "Revisiting Linear Discriminant Techniques in Gender Recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* , vol.33, no.4, pp.858,864, April 2011

[5] D. T. Nguyen, S. R. Cho, and K. R. Park, "Human age estimation based on multi-level local binary pattern and regression method," in Future Information Technology, vol. 309 of Lecture Notes in Electrical Engineering, pp. 433–438, Springer, 2014.

[6] S. E. Choi, Y. J. Lee, S. J. Lee, K. R. Park, and J. Kim, "A comparative study of local feature extraction for age estimation," in Proceedings of the 11th International Conference on Control, Automation, Robotics and Vision (ICARCV '10), pp. 1280–1284, Singapore, December 2010.

[7] S. E. Choi, Y. J. Lee, S. J. Lee, K. R. Park, J. Kim, "Age estimation using a hierarchitecal classifier based on global and local face features", Journal of Pattern Recognition, vol .44, no.6, p1262-1281, Feb 2011.

[8] Local Binary Pattern http://en.wikipedia.org/wiki/Local_binary_patterns

[9] B.S. Manjunathi, W.Y. Ma, Texture features for browsing, retrieval of image data, IEEE Transactions on Pattern Analysis and Machine Intelligence 18 (8) (1996) 837–842.

[10] X.Tan, S.Song, Z-H.Zhou and S.Chen. Enhanced Pictorial Structures for Precise Eye Localization under Uncontrolled Conditions, In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'09),pp. 1621–1628, Miami, Florida, USA, June 2009.

[11] X.Tan and B.Triggs. Enhanced Local Texture Feature Sets for Face Recognition under Difficult Lighting Conditions, In: Proceedings of the 2007 IEEE International Workshop on Analysis and Modeling of Faces and Gestures (AMFG'07), LNCS 4778, pp.168-182, 2007.

[12] Gunay, A.; Nabiyev, V.V., "Automatic age classification with LBP," *Computer and Information Sciences, 2008. ISCIS '08. 23rd International Symposium on* , vol., no., pp.1,4, 27-29 Oct. 2008

[13] Xin Geng; Zhi-Hua Zhou; Smith-Miles, K., "Automatic Age Estimation Based on Facial Aging Patterns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* , vol.29, no.12, pp.2234,2240, Dec. 2007