Attila Szilágyi: A comparison of different face classifiers based on Local Binary Patterns for emotional expression, gender, age and race
Argumentum 7 (2011), 357-365
Debreceni Egyetemi Kiadó

Working paper

Attila Szilágyi

A comparison of different face classifiers based on Local Binary Patterns for emotional expression, gender, age and race

Abstract

Computer image processing is a diversified area of information technology. Automatic age estimation, emotion, gender and race recognition based on the human face is an interesting and challenging task. In addition, this knowledge can be useful in such fields as the security industry, psychology and entertainment.

The purpose is to extract certain attributes of a person from a video stream or image. This process is built up from several steps: loading the input image from a sequence, face detection, cropping the desired face, preprocessing and classification. Our goal is to find the appropriate techniques both for preprocessing and classification.

In the experiments below we used face cropping and Local Binary Patterns for preprocessing and Support Vector Machines for classification. We performed the observations for each partition of each aspect using different parameter configurations. The results are collected in tables.
Keywords: face analysis, gesture recognition, local binary patterns

1 Introduction

Image processing is an exciting tool of informatics and automation. In general, we always want to extract implicit information held by a digital image. In certain situations it’s often easy for a human but not for a computer. Furthermore, the process can be done in different abstract levels of recognition. Face detection and further analysis primary belongs to the classification problems of image processing.

For recognition and analysis by computer, the human face is a challenging object. There are series of fields where face detection and analysis can be invoked: expression recognition, sex detection, age estimation, examination of different ethnic groups and so on. In addition, the above tasks can be closely related to each other and thus resolved all at once in compound applications. Furthermore the solutions and results of any of these problems are often helpful in other recognition tasks as well. Hence it is an intensively researched topic of image processing with an extensive literature.

The purpose is to extract as accurately as possible certain attributes of a person from a video stream or image. This classification process is built up from several important steps: first we load the input image from a sequence, then we attempt to detect the faces, crop the

Attila Szilágyi: A comparison of different face classifiers based on Local Binary Patterns for emotional expression, gender, age and race
Argumentum 7 (2011), 357-365
Debreceni Egyetemi Kiadó

desired face, do some preprocessing and finally perform the classification. The procedure is shown in Figure 1.

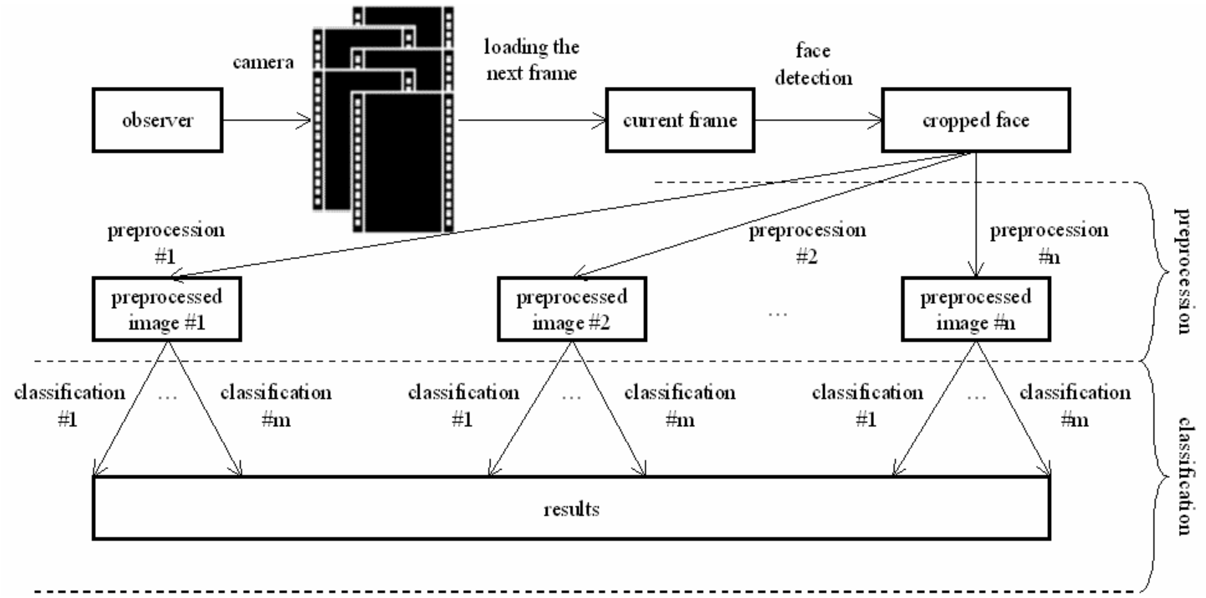

Figure 1: The steps of classification process

This paper covers a new method based on the so called LBP operator. The novelty lies in the application and analysis of concurrent LBP parameters together.

Facial gestures can represent a channel of non verbal communication. Image processing methods combined parallel with linguist processing can be an effective way to build a multimodal system.

2 Related research

2.1 Face detection

Our purpose is the extraction of features of a face. So, the first step is face detection. Finding the desired face is a prerequisite of any other meaningful operation; therefore we have to use an effective and reliable object detection method with the following reqirements: good detection rate, precision, high speed and robustness.

There are many different approaches which try to solve this problem: geometric feature-based, knowledge-based, appearance-based and so on. These are discussed in comprehensive surveys (Hjelmås 2001; Yang 2002). We chose the appearance-based approach and in our case the face is the object to be detected. The algorithm used is an implementation of the extended face detector of Lienhart and Maydt (Lienhart 2002), originally published by Viola and Jones (Viola 2001).

Attila Szilágyi: A comparison of different face classifiers based on Local Binary Patterns for emotional expression, gender, age and race
Argumentum 7 (2011), 357-365
Debreceni Egyetemi Kiadó

2.2 Analyzing faces

For humans, gender recognition is a very efficient and quickly learned process acquired in childhood (Wild 2000). Gender recognition is mainly done based on the eye region and face outline, while texture of the skin is essential as well (Bruce 1994; Yamaguchi 1995). Age estimation is similar, with skull shape and the texture of the face being very important for age recognition (Todd 1980).

The first attempts to recognize sex using computer vision techniques were based on neural networks. Golomb et al. applied a fully connected two-layer neural network that was trained and used to identify gender from 30 x 30 face images (Golomb 1991). Brunelli and Poggio (Brunelli 1992) used HyperBF networks to distinguish male and female. A Support Vector Machine (SVM) was tested for low resolution images to classify gender by Yang et al. (Yang 2000). Shakhnarovich et al. (Shakhnarovich 2002) utilized a face detector for demography analysis (gender and ethnicity). Wu et al. (Wu 2003) introduced an automatic real-time gender classification system based on Adaboost. Sun et al. (Sun 2006) also used Adaboost with Local Binary Patterns (LBP).

3 Technical description

In general a face classification procedure is built up from three parts: face localization, preprocessing and classification. In this section we briefly describe the techniques that we used during the experiments.

3.1 Face detection

The original method was introduced by Viola and Jones (Viola 2001). It is an extremely rapid detector with high detection rates. This is mainly due to three considerations. The first is a new representation of the image called “Integral Image” which allows very rapid computation of the used features. The second is the learning algorithm, based on Adaboost, which selects only the most significant features from a large set resulting efficient classifiers. The third consideration is a special organization of increasingly complex classifiers called “cascade structure”. It allows for the quick discard of unlikely background regions while spending more time on more promising object-like regions. In addition, Lienhart and Maydt (Lienhart 2002) extended the method with a new feature pool and a post-optimization process improving efficiency.

3.2 Preprocessing

Preprocessing is a mapping operation which serves to highlight the most discriminative features of the image for the corresponding classification. The input vector contains the pixels of the input image. The aim is to find a mapping for the input data which:

- groups the vectors closely in one place if they belong to the same class and
- keeps as much distance as possible between the groups of different classes.

Attila Szilágyi: A comparison of different face classifiers based on Local Binary Patterns for emotional expression, gender, age and race
Argumentum 7 (2011), 357-365
Debreceni Egyetemi Kiadó

Producing output images like that and using them during learning improves the performance of the classification. Two different methods were used:

- Basic preprocessing
- Local Binary Patterns (LBP)

First, we perform basic preprocessing of the input image, which involves face cropping, grayscaling, resizing to a predefined size, and histogram equalization. Then we apply the LBP operator to the resulting image.

The original LBP operator was introduced by Ojala et al. (Ojala 1996) for texture description. It labels the pixels of an image by thresholding the 3x3-neighborhood of each pixel with the center value. This value is considered as a binary number. Then a set of histograms of these labels made at different regions can be used as a texture descriptor.

The main limitation of the original method comes from its predetermined 3x3 size; however, by using circular neighborhood and bilinear interpolation, it was extended to any size (denoted by R) and number of pixels in the neighborhood (denoted by N) (Ojala 2002).

In addition, we made a further extension using a *series* of subdivisions all at once. Using multiple subdivisions allows more control in the extraction of environmental information. A subdivision determines the regions for the histograms and a subdivision is given by the dimensions of the Split parameter (see Figure 2).

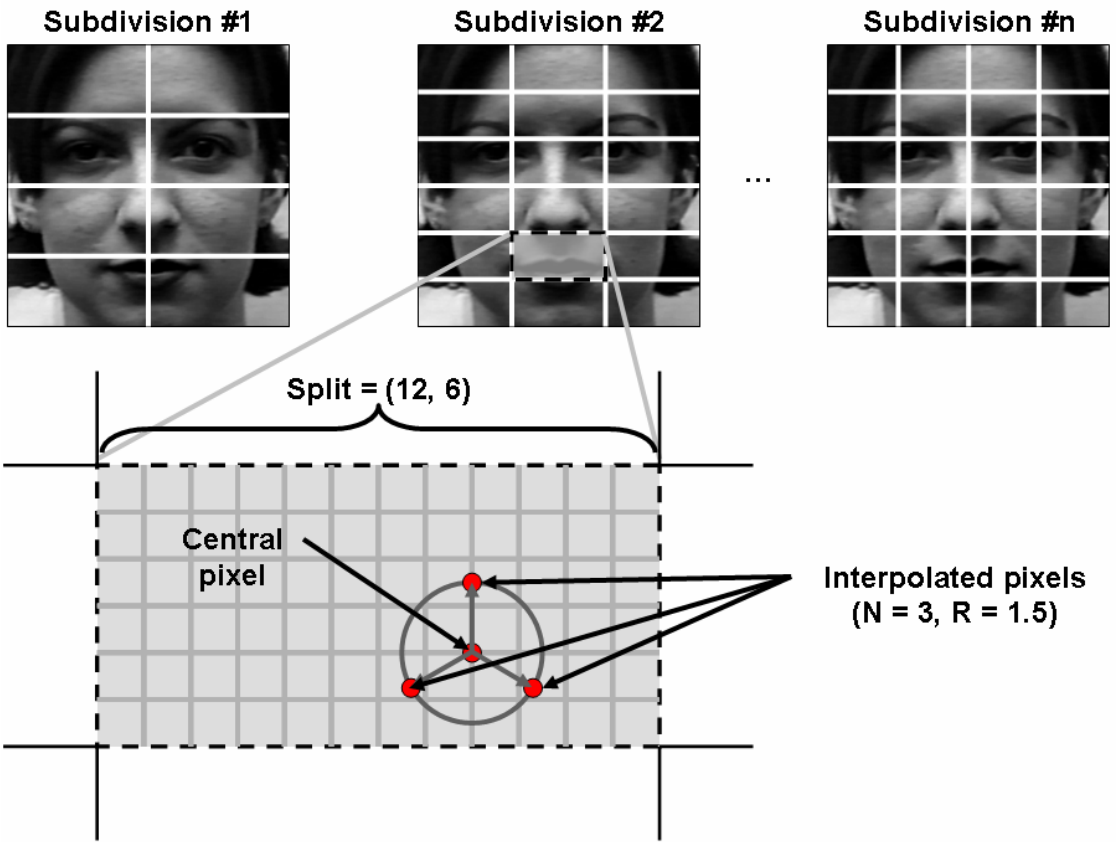

Figure 2: LBP parameters

Attila Szilágyi: A comparison of different face classifiers based on Local Binary Patterns for emotional expression, gender, age and race
Argumentum 7 (2011), 357-365
Debreceni Egyetemi Kiadó

3.3 Classification

SVM is a powerful supervised machine learning technique for classification and regression. Suppose a given set of training samples, each annotated by the corresponding classlabel. SVM maps this dataset into a higher (perhaps infinite) dimensional feature space, and then finds a linear separating hyperplane with the maximal margin between the mapped sample points of the different classes. Since SVM makes binary decisions, the multi-class case is originated from the two-class problem with the one-against-rest technique. It means that if we have more than two classes the task is to determine whether the incoming sample belongs to one given class or to another.

4 Evaluation

We used the Cohn-Kanade and the FERET face databases. The Cohn-Kanade contains persons in different emotional states without any annotations, so after manual labelling of images it’s appropriate to train gender and emotional classifiers. On the other hand, FERET has the helpful advantage of the accompanying annotations about sex, estimated age and skin color.

4.1 Experimental setups

Applying the mentioned databases, we prepared five databases (divided to train and test sets) in order to categorize facial expression, gender, age and race (Table 1).

	Partitions	Base
Gender	{male, female}	Cohn-Kanade
Gender	{male, female}	FERET
Expression	{angry, disgusted, scared, happy, natural, sad, surprised}	Cohn-Kanade
Age	{10 - 29, 30 - 49, 50+}	FERET
Race	{asian, black, hispanic, white}	FERET

Table 1: Training sets

The applied LBP configurations are also shown in Table 2.

	N	R	Splits	nOfFeatures
LBP 1	4	1	(5, 5)	1296
LBP 2	8	1	(5, 5)	20736
LBP 3	8	2	(5, 5)	20736
LBP 4	4	1	(10, 10)	256
LBP 5	8	1	(10, 10)	4096
LBP 6	8	2	(10, 10)	4096
LBP 7	4	1	(10, 10), (10, 15), (15, 15)	592
LBP 8	8	1	(5, 6), (8, 10), (12, 16)	27648

Table 2: LBP configurations

Attila Szilágyi: A comparison of different face classifiers based on Local Binary Patterns for emotional expression, gender, age and race
Argumentum 7 (2011), 357-365
Debreceni Egyetemi Kiadó

4.2 Results

In this section we collect the results of the tests in the following tables. The measured values should be interpreted as the averaged percentage of the successfully classified samples of the positive and the negative test set.

4.2.1 Emotion

We performed two tests for emotion on different test sets. Test 1 refers to a very small subset of the Cohn-Kanade database containing only 4 or 7 samples for each emotion (see Table 3).

	Basic	LBP 1	LBP 2	LBP 3	LBP 4	LBP 5	LBP 6	LBP 7	LBP 8
Angry	37.27	**69.69**	46.66	48.48	51.81	57.27	56.96	39.39	47.57
Disgusted	77.20	77.20	55.88	**81.61**	**81.61**	75.00	71.32	72.05	76.47
Scared	47.05	50.73	58.08	60.29	58.08	50.00	**69.11**	56.61	46.32
Happy	91.24	98.38	91.24	92.85	**100.00**	98.38	96.77	98.38	96.77
Natural	59.44	**85.71**	64.28	66.58	66.58	68.20	69.81	64.97	69.81
Sad	**68.48**	56.06	33.33	42.42	50.00	54.54	51.51	43.93	50.00
Surprised	90.10	91.66	91.66	83.33	83.33	**100.00**	91.66	83.33	91.66
Average	67.25	75.63	63.02	67.94	70.20	71.91	72.45	65.52	68.37

Table 3: Emotion test 1

Test 2 was done on the whole Cohn-Kanade database (see Table 4).

	Basic	LBP 1	LBP 2	LBP 3	LBP 4	LBP 5	LBP 6	LBP 7	LBP 8
Angry	60.40	59.23	58.74	52.82	55.44	**67.48**	60.06	58.28	52.66
Disgusted	75.70	73.36	51.60	**76.60**	67.25	61.45	73.56	72.94	60.69
Scared	54.42	50.80	56.42	54.92	52.57	54.90	**71.79**	51.50	48.87
Happy	91.18	92.21	87.63	89.90	92.28	91.91	**92.66**	91.52	92.16
Natural	**77.41**	76.30	66.84	70.12	75.11	73.10	76.98	75.74	75.06
Sad	**87.06**	52.06	60.01	71.80	50.00	52.35	51.40	61.31	54.71
Surprised	84.10	83.55	78.32	79.57	82.63	81.18	**84.55**	83.48	82.31
Average	75.75	69.64	65.65	70.82	67.90	68.91	73.00	70.68	66.64

Table 4: Emotion test 2

According to the observations, the operators with larger radius (LBP 3, 6) or larger grid (LBP 4, 5, 6) are more effective. In some cases face cropping may be sufficient.

4.2.2 Gender

For gender recognition we used both a Cohn-Kanade and a FERET based classifier. On the one hand they were tested on a test set based on the original database, on the other hand a cross testing was performed between the databases. This was executed in the following way:

• Test 1: Cohn-Kanade based classifier test on the base database.
• Test 2: FERET based classifier test on the base database.

Attila Szilágyi: A comparison of different face classifiers based on Local Binary Patterns for emotional expression, gender, age and race
Argumentum 7 (2011), 357-365
Debreceni Egyetemi Kiadó

• Test 3: Cohn-Kanade based classifier test on the FERET database.
• Test 4: FERET based classifier test on the Cohn-Kanade database.

In the case of the Cohn-Kanade based training set for gender recognition, it was possible to use faces in different emotional states to improve performance.

	Basic	LBP 1	LBP 2	LBP 3	LBP 4	LBP 5	LBP 6	LBP 7	LBP 8
Test 1	**84.77**	82.70	77.59	81.99	77.77	76.86	83.64	78.23	81.83
Test 2	**87.11**	79.50	71.82	77.57	77.62	78.99	81.47	78.96	80.66
Test 3	74.11	73.81	77.85	76.40	79.36	75.52	71.02	76.80	**80.16**
Test 4	52.98	59.80	65.71	65.00	**67.70**	56.07	61.77	62.25	62.69
Average	74.74	73.95	73.24	75.24	75.61	71.86	74.48	74.06	76.34

Table 5: Gender tests

Both classifiers work notably better on the training set based on the original database with only simple face cropping. However, the Cohn-Kanade based classifier shows much better generalization performance.

4.2.3 Age

Age estimation is shown on Table 6.

	Basic	LBP 1	LBP 2	LBP 3	LBP 4	LBP 5	LBP 6	LBP 7	LBP 8
10-29	62.42	63.79	63.72	63.52	66.32	65.03	63.06	64.04	**66.41**
30-49	**59.88**	54.31	53.81	56.51	58.72	59.51	56.49	59.71	56.56
50-	58.40	**76.62**	50.00	50.00	75.44	75.42	71.43	72.79	75.02
Average	60.23	64.91	55.84	56.68	66.83	66.65	63.66	65.51	66.00

Table 6: Age test

The age groups of 10-29 and 50+ are classified relatively well, but the classification of people in the 30-49 age group is not reliable.

4.2.4 Race

Observations on race recognition are summarized in Table 7.

	Basic	LBP 1	LBP 2	LBP 3	LBP 4	LBP 5	LBP 6	LBP 7	LBP 8
Asian	67.07	67.68	56.70	54.57	70.88	68.59	69.20	**71.34**	65.54
Black	**72.57**	54.05	50.00	50.00	59.05	51.35	51.35	65.94	50.00
Hispanic	50.00	50.00	50.00	50.00	**52.00**	50.00	50.00	49.87	50.00
White	76.43	75.49	64.87	63.13	74.84	76.04	76.92	**77.13**	74.20
Average	66.52	61.81	55.39	54.43	64.19	61.50	61.87	66.07	59.94

Table 7: Race test

Argumentum 7 (2011), 357-365
Debreceni Egyetemi Kiadó

The detection of the different ethnic groups on grayscaled images is very poor. The lack of color data does not allow correct accuracy.

5 Conclusions

In this paper we have studied face classifiers for emotion, gender, age and race recognition. For training and testing we applied the FERET and the Cohn-Kanade face databases with different parameter configurations. The experiments show the efficiency of LBP and SVM for these kind of classification problems. LBP worked better for emotion and basic preprocessing was sufficient for gender. The achieved results correspond to previous research. For race recognition the lack of the color data caused uncertain results and produced unreliable classifiers. Furthermore age estimation is also a difficult task when using only grayscaled images. Finally we also saw the good generalization capability of the Cohn-Kanade database on sex detection.

References

Bruce, V. & Langton, S. (1994): The use of pigmentation and shading information in re-cognising the sex and identities of faces. *PERCEPTION-LONDON*, 23: 803-803.

Brunelli, R. & Poggio, T. (1992): Hyperbf networks for gender classification. *Proceedings of the DARPA Image Understanding Workshop*, 311-314.

Golomb, B.A. Lawrence, D.T. & Sejnowski, T.J. (1991): Sexnet: A neural network identifies sex from human faces. *Advances in neural information processing systems*, 3: 572-577.

Hjelmås, E. & Low, B.K. (2001): Face detection: A survey. *Computer vision and image understanding*, 83 (3): 236-274.

Lienhart, R. & Maydt, J. (2002): An extended set of haar-like features for rapid object detection. *IEEE ICIP*, volume 1, 900-903. Citeseer.

Ojala, T., Pietikäinen, M. & Harwood, D. (1996): A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29 (1): 51-59.

Ojala, T., Pietikäinen, M. & Mäenpää, T. (2002): Multiresolution gray scale and rotation invariant texture analysis with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24 (7): 971-987.

Shakhnarovich, G., Viola, P.A. & Moghaddam, B. (2002): A unified learning framework for real time face detection and classification. *International Conference on Automatic Face and Gesture Recognition*, 14-21. Citeseer.

Sun, N., Zheng, W., Sun, C., Zou, C. & Zhao, L. (2006): Gender classification based on boosting local binary pattern. *Advances in Neural Networks-ISNN 2006*, 194-201.

Todd, J.T., Mark, L.S., Shaw, R.E. & Pittenger, J.B. (1980): The perception of human growth. *Scientific American*, 242 (2): 132.

Attila Szilágyi: A comparison of different face classifiers based on Local Binary Patterns for emotional expression, gender, age and race
Argumentum 7 (2011), 357-365
Debreceni Egyetemi Kiadó

Viola, P. & Jones, M. (2001): Rapid Object Detection using a Boosted Cascade of Simple Features. *Proc. IEEE CVPR 2001*. Citeseer.

Wild, H.A., Barrett, S.E., Spence, M.J., O'Toole, A.J., Cheng, Y.D. & Brooke, J. (2000): Recognition and sex categorization of adults' and children's faces: Examining performance in the absence of sex-stereotyped cues. *Journal of Experimental Child Psychology*, 77 (4): 269-291.

Wu, B., Ai, H. & Huang, C. (2003): Real-time gender classification. *Proceedings of SPIE*, volume 5286, 498.

Yamaguchi, M.K., Hirukawa, T. & Kanazawa, S. (1995): Judgment of gender through facial parts. *PERCEPTION-LONDON*, 24: 563-563.

Yang, M., Moghaddam, B. & Moghaddam, B. (2000): Gender Classification with Support Vector Machine. *4th IEEE Intl. Conf. on Automatic Face and Gesture Recognition*, 306-311. Citeseer.

Yang, M.H., Kriegman, D.J. & Ahuja, N. (2002): Detecting faces in images: A survey. *IEEE Transactions on Pattern analysis and Machine intelligence*, 34-58.

Attila Szilágyi
University of Debrecen
Image Processing Group of Debrecen
szilagyi.attila.hu@gmail.com