21st International Conference on Pattern Recognition (ICPR 2012)
November 11-15, 2012. Tsukuba, Japan

Age Classification in Unconstrained Conditions Using LBP Variants

Juha Ylioinas, Abdenour Hadid, and Matti Pietikäinen
Center for Machine Vision Research, University of Oulu, Finland
{juyl,hadid,mkp}@ee.oulu.fi

Abstract

Automatic age classification from human faces is a challenging task which has recently attained an increasing attention. Most of the proposed approaches have however been mainly concerning controlled settings. In this paper, we propose a novel method for age classification in unconstrained conditions and provide extensive performance evaluation on benchmark datasets with standard protocols, thus allowing a fair comparison and an easy reproduction of the results. Our proposed method is based on a combination of local binary pattern (LBP) variants encoding the structure of elongated facial micro-patterns and their strength. The experimental analysis points out the complexity of the age classification problem under uncontrolled settings. The proposed method provides state-of-the-art performance that can be used as a reference for future investigations.

1. Introduction

Human face embodies rich amount of information usable in many interesting applications, one of the most fascinating being automated age classification. The problem has inspired researchers leading to a diverse set of solutions, but a significant remark among the proposals is that their feasibility has been mainly evaluated in controlled settings. In this respect, more is needed to prove age classification to be worthwhile in more general and realistic scenarios, i.e., in settings acquired in unconstrained conditions.

Broadly speaking, automatic age classification aims to assign a label to a face regarding the exact age or the age group it belongs. As one of the recent dimensions added to the problem of face recognition, it is useful in many applications such as more affective Human-Computer Interaction (HCI), surveillance monitoring, forensics, audience measurement and reporting, electronic customer relationship management and so on [3].

Age classification from human faces is a very challenging problem because the appearance of a particular face varies due to changes in pose, expressions, illumination, and other factors such as make-up, occlusions, image degradations caused by blur and noise etc. In addition to these difficulties which are shared with the standard problems of face recognition, aging is a very complex process that is extremely difficult to model: a group of people of the same age may look very different depending on, for example, environment, lifestyle, genes etc.

Existing methods for age classification from facial images fundamentally differ in (i) the face representation and (ii) the classification scheme. Many face image representation methods have been studied such as anthropometric models, active appearance models (AAM), aging pattern subspace, and age manifold. An extensive review of age representation methods can be found in [3]. Regarding age classification schemes, the existing methods are based on either pure classification or regression analysis. Perhaps, among the pioneering studies on age classification are those proposed by Kwon and Vitoria Lobo [8], Lanitis et al. [9], and Guo et al. [5]. Although relatively successful in some scenarios (e.g. high-quality and occlusion-free images, clean background and neutral facial expression), most existing methods tend to suffer under uncontrolled settings as noted in [12].

In this work, we propose a novel method for automatic age classification in unconstrained conditions. Our approach encodes the textures of the facial regions using variants of the very successful local binary pattern (LBP) features. Our approach aims to combine complementary measures in terms of spatial structure (patterns) and contrast (the strength of the patterns) for efficient facial representation.

The lack of standard databases and associated protocols made the evaluation of the progress in the field of age classification very difficult. Very recently, efforts have been made within the BeFIT project (`http://fipa.cs.kit.edu/befit/`) by proposing standardized datasets and protocols for evaluating face analysis methods. We strongly approve this valuable initiative and report the results of our present work using such benchmark datasets and protocols thus allowing a fair comparison and an easy reproduction of our results.

2. Face representation using LBP complementary measures

It has been widely accepted in the computer vision community that the best results are most often produced by a combination of complementary descriptors rather than using any single operator alone. From this observation, we propose an efficient facial representation for age classification using LBP complementary measures.

The local binary pattern (LBP) operator is a simple yet powerful gray-scale invariant texture primitive, derived from a general definition of texture in a local neighborhood. The original LBP operator works in a 3×3 neighborhood, using the center value as a threshold to label each pixel and considering the result as a binary number.

A more generic form of the operator, referred as $\mathrm{LBP}_{P,R}$, imposes no limitations to the size of the neighborhood or to the number of sampling points. The formulation of the generic LBP operator is given by:

$$\mathrm{LBP}_{P,R} = \sum_{p=0}^{P-1} t(g_p - g_c)2^p, \tag{1}$$

where g_c corresponds to the gray value of the center pixel (x_c, y_c), g_p refers to gray values of P equally spaced pixels on a cicrle of radius R, and t defines a thresholding function with $t(x)=1$ if $x \geq 0$ and $t(x)=0$ otherwise. An extension called *uniform patterns* means that there is a separate label for each uniform pattern and all the non-uniform patterns are labeled with a single label.

LBP codes can be regarded as texture primitives including different types of curved edges, spots, flat areas, and so on. For an efficient facial representation, a given face image is usually divided into local blocks from which LBP histograms are extracted and concatenated into an enhanced feature histogram [1]. The classification is then performed by computing similarities between these concatenated local LBP occurrence histograms.

Different variants of the LBP operator can be adopted to improve the discriminative power and robustness of the operator. Particularly, one can make use of the most frequently occurring local binary patterns to represent textural information (referred to as Dominant LBPs [11]): First, the occurence histograms of all LBP codes are computed. Based on their occurrences these codes are then sorted in descending order to capture dominating patterns, which are finally selected for the model construction. The minimun number of patterns to cover $n\%$ of all LBP occurences is determined prior to extracting the features. Alternatively, Fisher separation criterion can also be used to learn and encode the local binary patterns (FBL-LBP) [6]. The most dominant LBP codes are first determined for each class. Then, all dominant LBP codes of each class are merged forming the global dominant types for the whole group of classes. The fundamental idea is to learn the most reliable and robust pattern types taking into account their intra-class and inter-class properties.

Instead of circular sampling as in basic LBP, an elliptical sampling can be considered yielding in so called elongated LBP (ELBP). This has been shown to be efficient in capturing anisotropic structures of the facial images [10].

A downside of the LBP formulation is that it uses only the signs of local pixel differences on image texture. That is, it totally ignores contrast information which is considered a very important cue for characterizing texture property. Our proposed approach, however, takes into account both of these two measures, i.e. spatial structure (patterns) and contrast (the strength of the patterns).

Instead of considering the joint distribution of LBP codes and a local contrast measure (LBP/C), we consider the completed modelling of LBP (CLBP) in which a local neighborhood is represented by its center pixel and a local difference sign-magnitude transform (LDSMT) [7]. The LDSMT decomposes the input image into two complementary components: the difference signs and the difference magnitudes. The sign component is coded using the LBP operator defined in (1), whereas the magnitude component is coded as:

$$\mathrm{CLBP\_M}_{P,R} = \sum_{p=0}^{P-1} t(m_p, c)2^p, \tag{2}$$

where m_p is the magnitude of local pixel difference, c is a predetermined threshold, usually set as the mean value of local pixel differences in the whole image, and t as in (1). As the magnitude operator encodes the difference in local pixel intensities, it gives a measure of contrast. Furthermore, the center pixel information is coded by globally thresholding the input image using its mean intensity.

To compute both sign and magnitude components, we divide the facial images into several blocks for spatially enhanced representation. We concatenate the sign and magnitude histograms into a single augmented histogram denoted as CLBP_S_M.

As the coding strategy of the magnitude component relies on a predetermined threshold parameter c defined in (2), we propose to divide the face into separate regions before the transformation for obtaining localized thresholds. Our proposed scheme for extracting the spatially enhanced magnitude histogram is illustrated in Fig. 1.

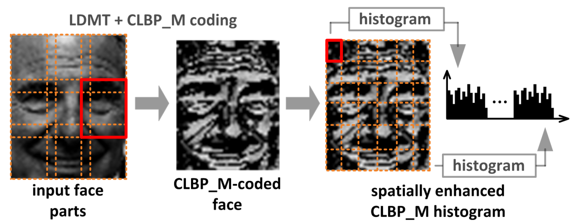

Figure 1. LBP magnitude histogram.

3. Experimental analysis

To gain insight into the effectiveness of our facial representation for unconstrained settings, we conducted extensive experiments following the guidelines of the BeFIT standards for benchmarking age classification methods.

3.1 Setup

Following the BeFIT recommendations, we considered the publicly available *Images of Groups* database consisting of 28,231 facial images collected from 5,080 Flickr images. The face images are taken in uncontrolled conditions. Each face is labeled with an age category. There are seven age categories as follows: 0-2, 3-7, 8-12, 13-19, 20-36, 37-65, and 66+. This grouping roughly corresponds to different life stages. The dataset is very challenging as many faces are of very low image resolutions.

We first cropped and normalized the face samples into images of 60×48 pixels using the eye coordinates. Then, the given LBP variant operator is applied resulting in an LBP face which is divided into 6×6 non-overlapping regions from which pattern occurrences are computed into a spatially enhanced histogram. For the magnitude component, we noticed that operating on nine separate face regions generally yields better results than operating on the whole face. Hence, as illustrated in Fig. 1, the given face was first divided into nine blocks that were separately processed by the local difference magnitude transform (LDMT) and the following magnitude coding. Finally, both sign and magnitude component histograms were concatenated into a single histogram.

For estimating the age of a person in a given test image, we used the extracted facial representations as inputs to a multiclass SVM classifier with a non-linear RBF kernel. The parameters of the SVM classifier were determined using a grid search. For SVM implementation, we used the publically available LIBSVM library [2].

3.2 Results

We compared several LBP variants and their combinations. As suggested in BeFIT, we trained the classifiers using a random selection of 3,500 faces, having equal number of samples per each age category (500 per class). Testing was performed on an independent uniformly distributed set of 1,050 face samples (150 per class).

We started by evaluating four different LBP variants based on local difference signs. We compared raw LBP sign outputs against uniform patterns using two radii values, namely $R=1$ and $R=2$, on a circular neighborhood. Then, we considered dominant and FBL patterns with $n = 90\%$ and for choosing the optimal radius. We also investigated the effect of elliptical sampling, with different orientations $\beta = \{0°, 45°, 90°, 135°\}$, setting the semi-major $A = 3$, the semi-minor $B = 2$, and $P = 10$ to obtain finer sampling. Regarding complementary measures, we considered different variants of magnitude patterns which were then concatenated with sign patterns. We considered both circular and elliptical neighborhoods. The results of the experiments clearly showed that combining magnitude information with uniform sign patterns enhances the age classification performance in all configurations. Furthermore, the results pointed out that combining uniform magnitude with uniform sign patterns using elliptical neighborhood (setting $A=3$, $B=2$, $P=10$, and $\beta=90°$) achieves the best performance.

To gain insight into the performance of our face representation (using the best performing feature combination), we looked at the system performance through the confusion matrix which is shown in Table 1. The matrix points out the difficulty in determining the age groups of 3-7, 8-12, 13-19, and 37-65, whereas the age groups of 0-2 and 66+ were more easily recognizable. Fig. 2 shows some examples of face images processed by our system depicting some correct and incorrect age classification cases.

We also conducted experiments comparing our approach against state-of-the-art methods. The results, shown in Table 2, clearly indicate the supremacy of

Table 1. Confusion matrix (rows denoting actual classes) of using the best feature combination and SVM classification.

	0-2	3-7	8-12	13-19	20-36	37-65	66+
0-2	**126**	21	2	0	0	0	1
3-7	28	**70**	37	6	4	5	0
8-12	7	37	**57**	29	11	7	2
13-19	1	7	29	**63**	34	11	5
20-36	1	4	4	53	**55**	31	2
37-65	2	5	4	20	34	**56**	29
66+	1	2	1	2	4	26	**116**

our approach which outperforms all other methods. It is worth noting that, unlike most of the other methods, our approach is not using any boosting which could further enhance our results (at the cost of time-consuming training).

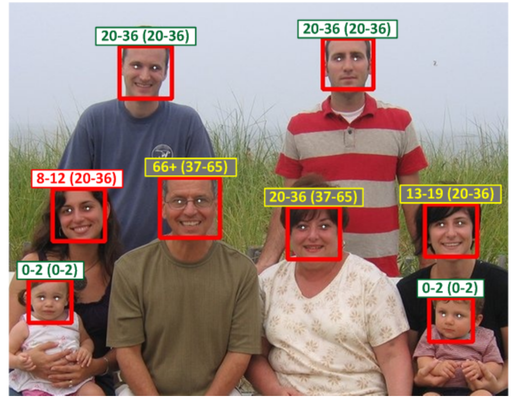

Figure 2. Examples of estimated age categories (ground truth in parantheses).

Table 2. Age classification results and comparison to state-of-the-art.

Approach	rank 1	rank 2
Appearance [4]	38.3 %	71.3 %
Appearance + Context [4]	42.9 %	78.1 %
Gabor + Adaboost [12]	43.7 %	80.7 %
LBP + Adaboost [12]	44.9 %	83.0 %
boosted Gabor + SVM [12]	48.4 %	84.4 %
boosted LBP + SVM [12]	50.3 %	87.1 %
Our approach	**51.7 %**	**88.7 %**

4. Conclusion

In this paper, we described a novel approach to the problem of age group classification by combining variants of the local binary pattern features for representing the face images. We conducted extensive experiments on a demanding real-life face database using the standard BeFIT benchmarking protocol for the evaluation. The obtained results indicated that elongated LBPs capturing the anisotropic facial micro-features, both their spatial structures and their strength, form an efficient feature set for age group classification. Though the experiments revealed the complexity of the age classification problem under uncontrolled settings, our proposed approach provided better results than the state-of-the-art. Our reported results can be used as a reference for future investigations.

Acknowledgments: The financial support of the Academy of Finland is gratefully acknowledged. This work has been partially performed within the EU FP7 project TABULA RASA (257289).

References

[1] T. Ahonen, A. Hadid, and M. Pietikäinen. Face description with local binary patterns: Application to face recognition. *TPAMI*, 28(12):2037–2041, 2006.

[2] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM TIST*, 2:27:1–27:27, 2011.

[3] Y. Fu, G. Guo, and T. Huang. Age synthesis and estimation via faces: A survey. *TPAMI*, 32(11):1955–1976, 2010.

[4] A. Gallagher and T. Chen. Understanding images of groups of people. In *CVPR'09*, pages 256–263, 2009.

[5] G. Guo, G. Mu, Y. Fu, and T. Huang. Human age estimation using bio-inspired features. In *CVPR'09*, pages 112–119, 2009.

[6] Y. Guo, G. Zhao, and M. Pietikäinen. Discriminative features for texture description. *Pattern Recognition*, 45(10):3834–3843, 2012.

[7] Z. Guo, L. Zhang, and D. Zhang. A completed modeling of local binary pattern operator for texture classification. *TIP*, 19(6):1657–1663, 2010.

[8] Y. H. Kwon and N. da Vitoria Lobo. Age classification from facial images. In *CVPR'94*, pages 762–767, 1994.

[9] A. Lanitis, C. Taylor, and T. Cootes. Toward automatic simulation of aging effects on face images. *TPAMI*, 24(4):442–455, 2002.

[10] S. Liao and A. C. S. Chung. Face recognition by using elongated local binary patterns with average maximum distance gradient magnitude. In *ACCV'07*, 2007.

[11] S. Liao, M. Law, and A. Chung. Dominant local binary patterns for texture classification. *TIP*, 18(5):1107–1118, 2009.

[12] C. Shan. Learning local features for age estimation on real-life faces. In *MPVA'10*, pages 23–28, 2010.