An In-depth Examination of Local Binary Descriptors in Unconstrained Face Recognition

Juha Ylioinas, Abdenour Hadid, Juho Kannala and Matti Pietikäinen
Center for Machine Vision Research
University of Oulu
firstname.lastname@ee.oulu.fi

***Abstract*—Automatic face recognition in unconstrained conditions is a difficult task which has recently attained increasing attention. In this domain, face verification methods have significantly improved since the release of the Labeled Faces in the Wild database, but the related problem of face identification, is still lacking considerations, which is partly because of the shortage of representative databases. Only recently, two new datasets called Remote Face and Point-and-Shoot Challenge were published providing appropriate benchmarks for the research community to investigate the problem of face recognition in challenging imaging conditions, in both, verification and identification modes. In this paper we provide an in-depth examination of three local binary description methods in unconstrained face recognition evaluating them on these two recently published datasets. In detail, we investigate three well established methods separately and fusing them at rank- and score-levels. We are using a well-defined evaluation protocol allowing a fair comparison of our results for future examinations.**

I. Introduction

Face recognition has been among the major topics in computer vision research for decades. In biometrics, processing with faces has to offer qualities such as acceptability and effortless collection, which is not the case, for example, with iris or fingerprints. Besides, among the general population, expectations from common consumer electronic devices (like Google glasses) and web services (like automatic face tagging in Facebook) have been rising all the time. In general, face recognition has applications in content-based image retrieval [1], security and surveillance [2], and human computer interaction [3], to name but a few.

The earliest work on automatic face recognition can be traced back to the 1970's, and since then, numerous algorithms have been developed including the classical *Eigenfaces* and *Fisherfaces* methods. An extensive review of face recognition from its early phase until the beginning of 2000's can be found in [4]. Currently, a hot subtopic is unconstrained face recognition, where the verification (one-to-one matching) or identification (one-to-many matching) process is often severely affected by resolution, blurring, noise, not forgetting the variations in pose, illumination, and expression. The major source of these issues is the fact that users can be situated pretty far away from the camera sensor during the image capture, the problem also known as remote face recognition. Another reason for these complications can be the photographer him/herself or the quality of the camera [5]. The key insight is that in unconstrained face recognition one moves further away from cooperative subjects. Moreover, in [6], a face identification scenario, presented by the investigation of the Boston Marathon bombings, was simulated and two state-of-the-art commercial face recognition systems were gauged in matching low quality face images of uncooperative subjects. Although, one rank-one hit for the suspect was presented, it was concluded that the common obstacles continue to confound matchers, and thus, it can be argued that unconstrained face recognition remains highly topical problem. For clarity, from this point onwards, by unconstrained face recognition it is meant any of the following four scenarios commonly discussed in the literature, namely remote face recognition, face recognition at a distance, face recognition from low-resolution images, or any other face recognition scenario where controlled imaging conditions or cooperative subjects (either both or only one) are absent.

One of the milestones in unconstrained face recognition is the release of the Labeled Faces in the Wild (LFW) [7] database, which was designed to promote development of face verification methods in unconstrained scenarios. The utilization of the database has been a big success, and currently, it is regarded as the *de-facto* evaluation benchmark for face verification. In the latest discussion, however, it has been argued whether LFW truly reflects the conditions often confronted in real-life-like scenarios [8]. For unconstrained face identification, things are not better as most of the investigations consider only constrained scenarios, where studio lights and cooperative subjects are used to control variations in pose, illumination and expression, representative evaluation benchmarks being, for example, FERET [9] and FRGC [10]. Only recently, two new benchmarks designed for studying unconstrained face recognition were published, namely the Remote Face [20] and the Point-and-Shoot Challenge [5] datasets. The release of these two datasets finally enables the research community to benchmark existing and novel methods for unconstrained face identification in more plausible conditions.

In this paper, we provide an in-depth examination of three off-the-shelf local binary description methods to tackle the problem of unconstrained face recognition. Our main focus is on closed-set face identification where a one-to-many comparison is made to establish an identity, and where the yet unidentified individual is known to lie as a template in the biometric database. We use two recently published databases particularly designed for studying this problem. Besides this, we investigate how the given description methods perform in face verification contributing the recently announced challenge related to the Point-and-Shoot dataset. The considered description methods are Local Binary Patterns (LBP) [12], Local Phase Quantization (LPQ) [13], and Binarized Statistical Image Features (BSIF) [14]. The main contribution is to benchmark a subset of recent powerful local binary description methods on the problem of unconstrained face recognition using new and challenging databases. To the best of our knowledge, this is the first examination of these two new datasets using well established local binary description methods that usually work as a baseline for many face recognition studies. The description methods chosen to this study are state-of-the-art especially in texture recognition and are among the top local image description methods in many other problems. The second contribution is to provide a comparison on the given three methods, namely LBP, LPQ, and BSIF, and to examine the potential of fusing them in face identification. Our work provides a valuable reference point for future work.

The outline of the paper is as follows: We first describe the local binary descriptors and fusion methods used throughout the study. Then, in the experimental section we provide evaluations on the two new benchmarks gauging the capability of the methods. Finally, we provide some discussion about the results in overall, compare them to the previous works in the literature, and make the concluding remarks.

II. Local binary descriptions for facial representation

One of the focus areas in face recognition is how to represent faces to the automatic recognizer. An ongoing success story has been the use of methods based on local image description, the notable ones including Gabor features, gradient based SIFT and HOG, and Local Binary Patterns (LBP). Especially binarized local image descriptors, like LBP, have gained a great favour in a wide spectrum of face analysis studies. An updated reference of the LBP methods for face representation can be found from [11]. Besides LBP, we next review LPQ and BSIF which were not evaluated in [11].

One of the pioneering works in local image description is Local Binary Patterns (LBP) [12]. The LBP method is based on a texture operator which works in a rectangular pixel neighborhood. In one LBP operation, the center value of the neighborhood is used as a threshold to label a set of surrounding pixels by zeros and ones to form a binary string which characterizes texture properties within this region. Usually, the sampling pattern of the operator follows a certain topology, for example a circle, using interpolation to sample at sub-pixel coordinates. The efficiency of the LBP method can be largely explained by the thresholding operation which makes the descriptor invariant to monotonic gray-scale changes. Fig. 1 depicts three neighborhood examples used to define a texture and calculate a local binary pattern.

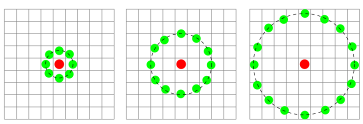


Fig. 1. An example of an LBP calculation using three different neighborhoods.

A blur invariant Local Phase Quantization (LPQ) was proposed for texture description in [13]. The LPQ method is based on an operator that examines the phase component of the Fourier transform in local neighborhoods. The fundamental idea is based on the assumption that the phase component at low frequencies is a blur invariant property under certain commonly fulfilled conditions saying the spatial blurring in an image should be close to a convolution between the image and a centrally symmetric point spread function like a Gaussian or a sinc-function. In LPQ, a short-term Fourier transform is first applied following the calculation of four frequency points. The phase information in the Fourier coefficients is recorded by taking the signs of the real and imaginary parts by using a simple scalar quantizer. Furthermore, in [13], it is said that decorrelating the samples before quantization can further improve the descriptor's performance. The overall LPQ calculation in a given neighborhood yields an eight-bit binary string. The steps of LPQ calculation are described in Fig. 2.

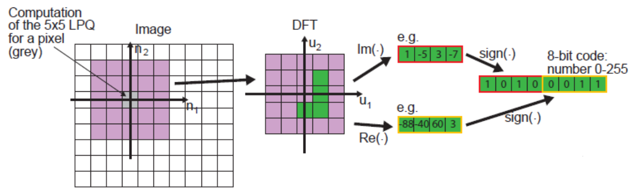


Fig. 2. An example of an LPQ pattern computation for a gray pixel using 5-by-5 neighborhood.

A rather new local binary description method is called Binarized Statistical Image Features (BSIF) [14]. BSIF is a data-driven method, where a filter bank is learnt from a set of natural image patches by maximizing the statistical independence of the filters' responses to those. One of the underlying ideas is to form a justified base for the independent binary quantization of the output coordinates of the filter bank. Indeed, by maximising statistical independence one is able to learn the most optimal set of filters with respect to the following independent quantization of response vector coordinates, which is the fundamental part of all local binary descriptor methods. Also, maximising the statistical independence results in an entropy growth between those individual coordinates leading to an efficient description process in overall. An example filter bank learnt using a set of natural images using a particular set of parameters is shown in Fig. 3.

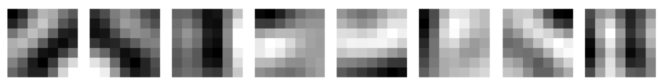


Fig. 3. A bank of 7-by-7 BSIF filters learnt from natural images.

As seen, the common factor of the presented descriptor methods is that they finally output a binary string for each pixel in the given image. What follows then is to decide how these binary strings are used in the recognition. The original procedure with the LBP description, as well as with LPQ and BSIF, is to collect the resulting binary descriptors in the image or its regions into a histogram which acts as an image description as a whole. In general, using histograms over a discrete vocabulary of local texture features is an effective way of image description. However, it has also been shown that the encoded image can be used as such [15], [16].

III. Fusion in face identification

Fusion is among the key methods when it comes to boosting the performance of a recognition system. In general, it can be applied in any level of the system covering sensor, feature, score, rank, and decision levels [17]. Among the most widely used fusion methods are the ones operating on rank and score levels. On the rank level, the consensus decision is made by using the rankings produced by different recognizers. On the score level, similarity scores or measures are combined in order to arrive at a final recognition decision. Score-level fusion is essentially based on score normalization, which is needed as score distributions vary as a function of the recognition algorithm. The idea behind fusing or combining pattern classifiers is to benefit from their non-overlapping classifications and the potential that these classifications retain some complementary information [18]. Next we describe two fusion methods that we are using in this study, namely the highest rank and the *w-score* method [19].

The highest rank method operates on a rank level. In the method each matched candidate is assigned the highest rank of the set of rankings computed by different recognition algorithms. The updated rank for user k is

$$r'_k = \min_j r_{k,j}, \tag{1}$$

where $r_{k,j}$ is the rank assigned to user k by the algorithm j. In the method, ties are solved randomly. The method is said to be useful only when the number of enrolled users is large compared to the number of matching algorithms. If this is not the case, the probability of having a tie increases raising the risk of the re-ranking to be uninformative [17]. In practical biometric systems, as well as in our work, the number of enrolled users is much bigger than the number of algorithms making the highest rank a potential solution for fusion.

To meet the requirements of robust fusion, a method called *w-score* normalization technique was introduced in [19]. The method is based on modelling a non-match distribution, in other words, it measures a probability that a particular score is not drawn from the non-match distribution. Among the key insights of *w-score* is that analysis is often done for a single input sample at a time, which is not, indeed, based on the overall match or non-match distributions. For example, in face identification, an input face produces at most one match score mixed in with a larger set of non-match scores, assuming the gallery contains only one template per subject. The idea is to model the tail of the non-match distribution, which is achieved by fitting a Weibull distribution to the top n matches. Given that, it does not matter what the underlying non-match distribution truly is, the only requirement is that the match scores are bounded, which is the case for most recognition systems. The details of the algorithm can be found from the original article [19]. Once the fitting has taken place the normalization can be completed. The *w-score* fusion can be defined as

$$s'_i = \sum_j s_{i,j}, \tag{2}$$

where $s_{i,j}$ is the normalized score for gallery class i using the algorithm j. In [19], robust fusion was defined to be a process that is insensitive to errors in its distributional assumptions on the data, has simple parameter estimation, and a high input failure tolerance.

IV. Experimental analysis

We conduct our descriptor examinations on unconstrained face identification by first evaluating the description methods separately and then by combining them using two fusion methods operating on different system levels. In this work, we utilize the Remote Face (Remote) [20] and the Point-and-Shoot Challenge (PaSC) [5] datasets. The evaluation protocol for both benchmarks is based on testing several separate query sets under different imaging conditions on a target frontal face set with varied number of templates per individual. The evaluation protocol for Remote follows the guidelines of the dataset package distributed by the authors in [20]. However, the protocol used in the related paper [20] is not the same than in the official distribution. As there are not any official specifications for how the PaSC dataset should be used in face identification experiments, we define our own protocol which is easily reproducible. Finally, we evaluate the given description methods on face verification using the PaSC dataset contributing to the recently announced face recognition challenge [5].

A. Setup

Remote contains 17 different individuals and 2,102 face images in total. The dataset is collected in challenging conditions covering images taken from long distances in an unconstrained outdoor maritime environment [20]. The face images of the database are organized into seven folders which are listed in Table I with detailed description and the number of images. The images are organized so that the folders *low_reso*, *pose_frontal*, and *pose_nonfrontal* are non-overlapping, whereas the rest are subsets of *pose_frontal* and *pose_nonfrontal* folders. Using the given sets, one is able to examine how the image descriptors perform under specific imaging conditions.

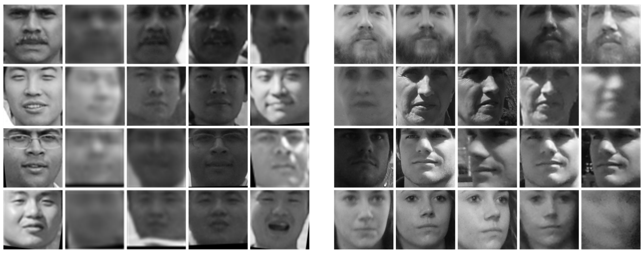
PaSC contains 293 different individuals and 9,376 faces in total. The dataset is organized into two folders referred as *query* and *target*, which both have equal number of samples. Both of these folders contain 16 images per an individual so that there are four samples taken in four conditions referred as *close_frontal*, *close_nonfrontal*, *distant_frontal*, and *distant_nonfrontal* explaining the distance and pose conditions. PaSC images were taken in nine different locations, both in- and outdoors.

TABLE I. The face image folders in the Remote database.

folder	description	no. of images
gallery	5 clear and frontal with good illumination	85
low_reso	Very low resolution	90
blur	Frontal and blurred	75
illum	Frontal with varying illumination	561
illum_blur	Frontal with blur and varying illumination	128
pose_frontal	All frontal	1,081
pose_nonfrontal	All non-frontal	846

For evaluating the local description methods and their fusion in face identification on these two benchmarks we adopt the following procedure: For both datasets we take a gallery set which contains only frontal face images. For query sets, we then take several sets that exhibit certain imaging conditions. For Remote, we use the *gallery* folder for the gallery, and for query sets we use the remaining six folders. With this kind of setting, the gallery images in this experiment are composed of clear and well illuminated faces. For PaSC, we use the *close_frontal* subset of the *target* folder as a gallery set, and for query sets we use all four subsets of the *query* folder. Unlike in the Remote experiment, gallery images in this experiment contain also blurred and poorly illuminated faces. In both experiments, the number of faces per subject in the gallery set is gradually increased, and each time the faces are chosen randomly. The experiments are repeated 20 times and the final recognition result is reported using the average of these permutations. For the face verification part we use only the PaSC dataset and follow the guidelines given by the recent face recognition challenge [5]. The problem is divided into two parts where (i) only all frontal query faces are matched against all frontal target faces and (ii) all query faces are compared against all target faces. The corresponding results are presented at FAR=1% like it was proposed in [5].

We first align each face with respect to eye coordinates provided by the databases. After that we resize each face into 66×66 pixels and apply the illumination normalization method proposed in [15]. The aligned face is then processed using the given local binary descriptor, and then, divided into 6×6 non-overlapping cells from which the descriptor labels are collected into histograms which are finally concatenated forming the overall description of the face. The classification is performed using nearest neighbor classifier with χ^2 distance metric. Fig. 4 depicts some geometrically normalized face samples from the used face datasets.



(a) (b)

Fig. 4. A random set of normalized face samples from (a) Remote and (b) PaSC databases. Each row represents images taken from the same individual. The images demonstrate some of the major challenges (such as lighting, motion blur and poor focus) that are common in unconstrained face recognition.

All of the local descriptors used here are parametric with respect to the spatial support (neighborhood size) and to the number of quantizations (or filterings) to be performed. We found 7×7 neighborhoods with eight-bit binary coding to perform consistently well for all methods for the given face size. For LBP, we used circular topology with eight sampling points using a radius of 3. For LPQ, we used the common four-frequency-point coding of the phase information in the 7-by-7 neighborhood, and for BSIF, a bank of eight 7×7 size of filters. With these settings, the length of the concatenated histogram representation of the face using any of these methods yields 9216 ($6\times6\times256$) elements. For LPQ, we also used the decorrelation scheme as described in [13]. For fitting Weibull distributions in score-level fusion used in face identification, we found that the tail size of 5 performs consistently well for all descriptors.

B. Results

In the Remote experiment, we first evaluate each descriptor separately performing rank-1 analysis by gradually increasing the number of gallery samples per individual. The idea of this experiment is to examine how well these descriptors perform in exact recognition and how well one is able to alleviate the task by simply increasing the number of gallery samples per individual. The results, shown in Table II, point out that there is no great difference between the methods in their performance. While having only one gallery sample per individual, BSIF seems to be superior compared with others. Moreover, except with low resolution images, BSIF seems to be slightly the best on all folders. It can be seen that the performance generally increases as a function of the number of gallery samples per individual.

TABLE II. RANK-1 ANALYSIS ON THE REMOTE DATABASE USING LBP, LPQ, AND BSIF DESCRIPTORS (EACH COLUMN IS FOR DIFFERENT NUMBER OF GALLERY SAMPLES PER INDIVIDUAL).

method	1	2	3	4	5
			low_reso		
LBP	9.0	12.2	11.2	11.7	12.4
LPQ	9.0	10.1	11.9	13.1	13.5
BSIF	10.3	11.0	10.3	11.4	11.2
			blur		
LBP	40.3	42.3	45.4	45.9	45.9
LPQ	41.2	44.9	51.1	55.4	58.1
BSIF	42.8	46.0	53.9	59.3	62.2
			illum		
LBP	57.5	67.2	74.1	76.8	78.4
LPQ	58.6	66.8	73.7	76.9	79.5
BSIF	60.0	68.7	75.0	77.6	79.3
			illum_blur		
LBP	46.2	54.6	64.8	69.1	73.2
LPQ	49.9	59.6	67.5	71.7	75.6
BSIF	52.3	61.5	69.2	72.2	74.8
			pose_frontal		
LBP	49.4	57.5	63.5	66.1	67.9
LPQ	50.1	57.4	64.6	67.9	70.6
BSIF	51.5	59.0	65.4	68.2	70.1
			pose_nonfrontal		
LBP	33.1	39.8	44.0	47.0	49.9
LPQ	35.8	40.6	44.1	45.9	47.5
BSIF	35.9	41.8	45.3	47.5	49.2

To have a deeper view of the performance we further performed rank-k analysis. The idea of this step is to examine how well the descriptors perform in rank-k recognition, which measures the recognition rate when it is enough that the correct match is among the first k candidates returned by the algorithm. The methods are evaluated in such a way that the gallery set contains only one sample per individual, which is because the fusion methods tested here are originally designed only to that kind of operation mode. We run the 20 permutations and report the average rates up to rank-5. The results on *pose_frontal* and *pose_nonfrontal* folders are shown in Table III. According to the table, LPQ and BSIF outperfom LBP, and BSIF seems to perform better compared with LPQ. Fusing all three descriptors or only BSIF and LPQ on the score level using *w-score* seems to give the best results.

In the PaSC examination we proceed like in the previous experiment. We first perform rank-1 analysis by gradually increasing the number of gallery images per individual. The results, shown in Table IV, resemble the results obtained from the previous experiment for the most part. First, the performance of each method improves when the number of gallery samples per individual increases. Moreover, BSIF seems to outperform

TABLE III. RANK-k ANALYSIS OF LBP, LPQ, AND BSIF METHODS WITH AND WITHOUT FUSION, WHERE H STANDS FOR *h-rank* AND W FOR *w-score*.

method		rank-1	rank-2	rank-3	rank-4	rank-5
pose_frontal						
LBP		49.4	60.6	68.4	74.4	79.1
LPQ		50.1	61.5	69.4	75.7	80.5
BSIF		51.5	62.2	69.9	75.8	80.7
LBP+LPQ	H	49.6	61.7	69.6	75.5	80.4
BSIF+LBP	H	50.5	62.1	69.7	75.6	80.3
BSIF+LPQ	H	50.9	62.6	70.3	76.2	81.2
ALL 3	H	49.6	62.3	70.0	76.1	80.7
LBP+LPQ	W	51.2	62.0	70.0	75.9	80.7
BSIF+LBP	W	51.9	62.8	70.4	76.4	80.7
BSIF+LPQ	W	51.7	62.9	70.6	76.6	**81.7**
ALL 3	W	**52.4**	**63.0**	**71.0**	**76.8**	81.4
pose_nonfrontal						
LBP		33.1	43.8	51.4	57.5	63.0
LPQ		35.8	47.3	55.3	61.2	66.7
BSIF		35.9	47.8	55.9	61.9	67.0
LBP+LPQ	H	34.7	46.2	54.0	60.1	65.1
BSIF+LBP	H	34.3	46.3	54.3	60.4	65.3
BSIF+LPQ	H	35.8	48.2	56.4	62.4	67.3
ALL 3	H	33.7	47.0	54.9	61.1	66.1
LBP+LPQ	W	36.0	47.1	54.5	60.3	65.4
BSIF+LBP	W	36.0	47.1	54.8	60.8	65.7
BSIF+LPQ	W	36.7	**48.8**	**56.9**	**62.8**	**67.6**
ALL 3	W	**37.0**	48.5	55.8	61.8	66.8

other methods with frontal and nonfronal face images captured from closer distances. It is noteworthy, however, to see the deteriorated performance of BSIF with respect to LBP and LPQ with face images captured at a distance.

TABLE IV. RANK-1 ANALYSIS ON THE PASC DATASET USING LBP, LPQ, AND BSIF DESCRIPTORS.

method	1	2	3	4
close_frontal				
LBP	25.1	35.3	43.4	49.4
LPQ	26.4	36.7	44.3	49.1
BSIF	27.0	37.4	45.6	51.0
close_nonfrontal				
LBP	6.5	9.3	11.6	13.7
LPQ	6.5	9.3	11.4	12.7
BSIF	6.9	9.8	12.7	14.9
distant_frontal				
LBP	21.9	30.5	37.0	41.2
LPQ	23.4	32.3	38.8	42.9
BSIF	21.8	29.7	36.6	41.1
distant_nonfrontal				
LBP	12.9	18.9	23.3	26.6
LPQ	13.6	19.5	24.6	28.9
BSIF	13.1	18.8	23.2	26.0

In the rank-k analysis we compare all methods in the same manner like in the Remote experiment, but now in the four scenarios defined by the PaSC database. The results are shown in Table V. Evaluating on PaSC, the problem turns out be more difficult compared with the Remote experiment, which is because of the increased number of individuals (17 vs. 293), but also, because the gallery set in PaSC contains blurred and poorly illuminated faces. It seems that all methods are quite invariant to imaging distance on frontal faces, but nonfrontal faces from close are harder to recognize than nonfrontal faces at a distance, which may be due to the inherent characteristics of the database. Finally, score-level fusion of all three descriptors using *w-score* seems to offer the highest accuracy throughout the folders.

In the end, we make our contribution to the recent face recognition challenge using the PaSC dataset by evaluating the given descriptors on a face verification mode. The results, shown in Table VI, indicate that any of the descriptors performs better than the given baselines, CohortLDA and LRPCA. However, it can be seen that the margin between the algorithm developed by Pittsburgh Pattern Recognition (PittPatt) and the rest of the methods remains quite huge.

TABLE V. RANK-k ANALYSIS OF LBP, LPQ, AND BSIF METHODS WITH AND WITHOUT FUSION.

method		rank-1	rank-2	rank-3	rank-4	rank-5
close_frontal						
LBP		25.1	31.1	34.8	37.5	39.3
LPQ		26.4	32.2	35.8	38.3	40.6
BSIF		27.0	33.1	37.1	39.9	42.3
LBP+LPQ	H	25.7	32.7	36.5	39.4	41.7
BSIF+LBP	H	26.0	33.2	37.1	40.3	42.4
BSIF+LPQ	H	26.6	**33.7**	37.5	40.2	42.5
ALL 3	H	25.7	33.4	37.6	40.4	**42.9**
LBP+LPQ	W	26.0	32.2	36.4	39.4	41.8
BSIF+LBP	W	26.6	33.0	37.3	40.3	42.5
BSIF+LPQ	W	**27.4**	**33.7**	**37.7**	**40.5**	42.7
ALL 3	W	27.1	33.5	37.5	**40.5**	**42.9**
close_nonfrontal						
LBP		6.5	9.0	10.8	12.2	13.5
LPQ		6.5	8.9	10.7	12.1	13.5
BSIF		6.9	9.7	11.9	**13.7**	**15.2**
LBP+LPQ	H	6.5	9.1	11.0	12.6	13.8
BSIF+LBP	H	6.7	9.6	11.8	13.5	15.0
BSIF+LPQ	H	6.7	9.6	11.6	13.2	14.8
ALL 3	H	6.4	9.6	11.5	13.3	14.6
LBP+LPQ	W	6.8	9.3	11.3	12.7	14.0
BSIF+LBP	W	**7.1**	9.8	**12.0**	**13.7**	15.1
BSIF+LPQ	W	**7.1**	9.7	11.8	13.6	14.9
ALL 3	W	**7.1**	**9.9**	**12.0**	13.6	15.0
distant_frontal						
LBP		21.9	27.2	31.1	33.9	36.1
LPQ		23.4	29.1	32.8	35.4	37.6
BSIF		21.8	27.1	30.7	33.3	35.6
LBP+LPQ	H	22.6	29.3	33.1	36.1	38.4
BSIF+LBP	H	21.9	28.0	31.8	35.0	37.5
BSIF+LPQ	H	22.5	29.4	33.3	36.2	38.5
ALL 3	H	21.6	29.0	33.2	36.4	38.8
LBP+LPQ	W	**24.1**	29.8	33.6	36.5	38.7
BSIF+LBP	W	23.1	28.3	32.1	35.3	37.8
BSIF+LPQ	W	23.7	29.5	33.4	36.6	38.9
ALL 3	W	**24.1**	**30.0**	**34.0**	**37.0**	**39.4**
distant_nonfrontal						
LBP		12.9	16.9	19.5	21.8	23.7
LPQ		13.6	17.9	20.7	23.1	25.0
BSIF		13.1	17.2	19.9	22.1	24.4
LBP+LPQ	H	13.1	18.0	20.9	23.3	25.2
BSIF+LBP	H	13.0	17.5	20.5	22.9	24.9
BSIF+LPQ	H	13.5	18.2	21.2	23.5	25.5
ALL 3	H	12.7	17.9	21.0	23.6	25.5
LBP+LPQ	W	14.0	18.3	21.2	23.7	25.4
BSIF+LBP	W	13.6	17.8	20.8	23.2	25.2
BSIF+LPQ	W	**14.2**	**18.7**	21.6	24.0	26.0
ALL 3	W	**14.2**	**18.7**	**21.9**	**24.3**	**26.3**

TABLE VI. FACE VERIFICATION EXPERIMENT ON THE PASC DATASET USING LBP, LPQ, AND BSIF DESCRIPTORS. FOR EACH METHOD IN THE TABLE, VERIFICATION RATE IS REPORTED AT FAR=1%.

method	frontal	all
PittPatt [5]	55.0	41.0
CohortLDA [5]	22.0	8.0
LRPCA [5]	19.0	10.0
LBP	24.3	9.5
LPQ	23.1	13.2
BSIF	24.9	14.3

C. Discussion

Based on the face identification results, BSIF is the most efficient individual descriptor when matching faces containing enough resolution and sharpness. However, recognizing very low resolution faces of Remote and distant faces of PaSC, LPQ seems to be the most efficient. It is noteworthy, however, that the used BSIF filters are learnt using natural images that seemingly do not contain any inherent blur which may finally explain the deteriorated perfomance. Thus, for future work, it is of interest to see whether learning BSIF filters using blurred images could offer some improvement. Also, it is of interest to see whether it is possible to learn even more powerful filters using face images instead of natural ones.

Based on the results of fusing descriptors in face identification, score-level fusion works better than rank-level fusion. This may be because by analysing scores one takes into better consideration the relation of the ranked matches produced by each individual matcher. Either way, one evidently is able to gain benefit from fusion. We also saw that unconstrained face identification can be alleviated by using several gallery images per individual. However, it is highly expected that the overall perfomance could improve even more by introducing more sophisticated classifiers compared to nearest neighbor one, for example, by taking simply the k-nearest neighbor classifier. In that case, however, one must consider fusion methods that are suitable for processing with multi-sample galleries.

Comparing our achieved results to the ones in [20] on Remote it seems that by our methods one is able to obtain at least as good rank-1 rates on the different folders. However, because the evaluation protocol used here is slightly different from the one in [20], stronger conclusions are harder to make. Concerning the face identification experiments on PaSC, to the best of our knowledge, our work is the very first. What comes to the face verification, all of our methods outperformed the baseline methods, but still do not reach the best result that was reported for this particular dataset in [5].

V. Conclusion

In this work we investigated face recognition in challenging conditions using three local binary description methods evaluating them on two recently published benchmarks called the Remote Face and the Point-and-Shoot Challenge. The local binary descriptors evaluated here are *Local binary pattern*, *Local phase quantization*, and *Binarized statistical image features*. With each of them, we applied the widely used concatenated histogram based face representation and measured their performance using standard histogram distance metrics combined with simple nearest neighbor classifier. We also investigated the fusion of the descriptors on rank- and score-levels using the highest rank and the recent *w-score* method with promising results.

The motivation of our work is partly to offer face identification benchmark results on these new challenging face datasets, but also to evaluate the efficiency of recent local binary descriptor method (BSIF), and in general, to investigate the potential of fusing local image descriptors. We also discussed some potential means to further improve the results on unconstrained face identification. It can be seen that the obtained results are far from perfect, but as there are hardly any works examining the problem using the given new datasets we believe that our study works as a valuable reference point for future research.

Finally, we made our contribution to the recent Point-and-Shoot face recognition challenge by evaluating the given description methods on unconstrained face verification. We showed outperforming results compared to the announced baseline methods, but also observed that they are still inferior to commercial solutions.

Acknowledgement. The financial support of the Academy of Finland is gratefully acknowledged. The work was also partially supported by the FiDiPro program of Tekes and ICT COST Action IC1106.

References

[1] N. Kumar, A. C. Berg, P. N. Belhumeur and S. K. Nayar, "Describable Visual Attributes for Face Verification and Image Search", *IEEE TPAMI* **33** 10 (2011) pp. 1962–1977

[2] D. A. Vaquero, R. S. Feris, D. Tran, L. Brown, A. Hampapur and M. Turk, "Attribute-based People Search in Surveillance", *Proc. WACV'09* (2009) pp. 1–8

[3] H. K. Ekenel, J. Stallkamp, H. Gao, M. Fischer and R. Stiefelhagen, "Face Recognition for Smart Interactions", *Proc. ICME'07* (2007) pp. 1007–1010

[4] W. Zhao, R. Chellappa, J. Phillips, and A. Rosenfeld, "Face Recognition: A Literature Survey", *ACM Comput. Surveys* **35** 4 (2003) pp. 399–458

[5] J. R. Beveridge, P. J. Phillips, D. Bolme, B. A. Draper, G. H. Givens, Y. M. Lui, M. N. Teli, H. Zhang, W. T. Scruggs, K. W. Bowyer, P. J. Flynn and S. Cheng, "The Challenge of Face Recognition from Digital Point-and-Shoot Cameras", *Proc. BTAS'13* (2013) pp. 1–8

[6] J. C. Klontz and A. K. Jain, "A Case Study on Unconstrained Facial Recognition Using the Boston Marathon Bombings Suspects", Technical Report MSU-CSE-13-4 (Last Revised May 30, 2013)

[7] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments", Technical Report 07-49, University of Massachusetts, Amherst, October 2007.

[8] N. Pinto, J. DiCarlo, and J. Cox, "How Far Can You Get with a Modern Face Recognition Test Set Using Only Simple Features?", *Proc. CVPR'09* (2009) pp. 2591–2598

[9] P.J. Phillips, H. Wechsler, J. Huang, and P.J. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms", *IVC* **16** 5 (1998) pp. 295–206

[10] P.J. Phillips, P.J Flynn, T. Scruggs, K.W. Bowyer, C. Jin, K. Hoffman, J. Marques, M. Jaesik, and W. Worek, "Overview of the face recognition grand challenge", *Proc. CVPR'05* (2005) pp. 947–954 vol. 1

[11] M. Bereta, W. Pedrycz and M. Reformat, "Local Descriptors and Similarity Measures for Frontal Face Recognition: A Comparative Analaysis", *J. Vis. Commun. Image R.* **24** (2013) pp. 1213–1231

[12] T. Ahonen, A. Hadid and M. Pietikäinen, "Face Description with Local Binary Patterns: Application to Face Recognition", *IEEE TPAMI* **28** 12 (2006) pp. 2037–2041

[13] E. Rahtu, J. Heikkilä, V. Ojansivu and T. Ahonen, "Local Phase Quantization for Blur-insensitive Image Analysis", *IVC* **30** 8 (2012) pp. 501–512

[14] J. Kannala and E. Rahtu, "BSIF: Binarized Statistical Image Features", *Proc. ICPR'12* (2012) pp. 1363–1366

[15] X. Tan and B. Triggs, "Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions", *IEEE TIP* **19** 6 (2010) pp. 1635–1650

[16] G. Heusch, Y. Rodriguez and S. Marcel, "Local Binary Patterns as an Image Preprocessing for Face Authentication", *Proc. AFGR'06* (2006) pp. 6–14

[17] A. A. Ross, K. Nandakumar and A. K. Jain, *Handbook of Multibiometrics*, 1st ed. Springer

[18] J. Kittler, M. Hatef, R. P. W. Duin and J. Matas, "On Combining Classifiers", *IEEE TPAMI* **20** 3 (1998) pp. 226–239

[19] W. Scheirer, A. Rocha, R. Micheals and T. Boult, "Robust Fusion: Extreme Value Theory for Recognition Score Normalization", *Proc. ECCV'10* (2010) pp. 481–495

[20] R. Chellappa, J. Ni and V. M. Patel, "Remote Identification of Faces: Problems, Prospects, and Progress", *Pattern Recognition Letters* **33** (2012) pp. 1849–1859