Evaluation of descriptors and distance measures on benchmarks and first-person-view videos for face identification

Bappaditya Mandal*[,1], Wang Zhikai[2], Liyuan Li[1] and Ashraf A. Kassim[2]

[1]Visual Computing Department, Institute for Infocomm Research, Singapore
[2]Electrical and Computer Engineering, National University of Singapore
Email address: bmandal@i2r.a-star.edu.sg (*Contact author: Bappaditya Mandal);
a0080959@nus.edu.sg (Wang Zhikai); lyli@i2r.a-star.edu.sg (Liyuan Li);
ashraf@nus.edu.sg (Ashraf A. Kassim)

Abstract. Face identification (FI) has made significant amount of progress in the last three decades. Its application is now moving towards wearable devices (like Google Glass and mobile devices) leading to the problem of FI on first-person-views (FPV) or ego-centric videos for scenarios like business networking, memory assistance, etc. In the existing literature, performance analysis of various image descriptors on FPV data are little known. In this paper, we evaluate four popular image descriptors: local binary patterns (LBP), scale invariant feature transform (SIFT), local phase quantization (LPQ) and binarized statistical image features (BSIF) and ten different distance measures: Euclidean, Cosine, Chi square, Spearman, Cityblock, Minkowski, Correlation, Hamming, Jaccard and Chebychev with first nearest neighbor (1-NN) and support vector machines (SVM) as classifiers for FI task on both benchmark databases: FERET, AR, GT and FPV database collected using wearable devices like Google Glass (GG). Comparative analysis on these databases using various descriptors shows the superiority of BSIF with Cosine, Chi square and Cityblock distance measures using 1-NN as classifier over other descriptors and distance measures and even some of the current state-of-art benchmark database results.

1 Introduction

The rise of wearable technology has opened up numerous opportunities to further improve our lifestyles with technological advancements. Bulky medical equipments used to measure our vital statistics can be replaced with watches or handphones and heavy cameras replaced with GoPros [1] and Google Glass [2]. With facial recognition technology emerging in the past decade, wearable cameras such as the GG allow for amazing possibilities. These cameras can recognize daily activities and detect social interactions effortlessly; atomic actions such as turning left and right can be detected from first-person camera movement, while group activities can be recognized based on individual actions and pairwise context [3]. Faces can be used as a significant source of information as

their attention patterns play a huge role in identification, recognizing each other in social interactions, business networking, visual memory assistant and many other applications that are popular nowadays [4, 5].

There are some unique challenges of performing FI on FPVs data generated from wearable devices. One of them is that in FPV, both wearable camera and the subject are moving or jittering, so the images are often blurry in nature and mug shot (studio or controlled condition) image of the person is not readily available/possible. Also, it is difficult to obtain large number of images of the person to be recognized because the person might not stay in the view for a long time. Moreover, in wearable devices the computation resources are limited so the algorithm should be fast enough to be executed under constrained mobile environment. So it is important that we perform the evaluations of local descriptors on FPV (face data obtained from GG) as well as benchmark face image databases under the same framework. This will help us to understand what features along with the distance measures are beneficial for wearable devices and in general FPV data.

In this paper, we don't want to use any training/learning algorithms, but use raw local descriptor features for FI on FPV data. We evaluate various descriptors and distance measures in order to determine the optimal configurations for which we can obtain good FI accuracies in FPV video data. For us to achieve a fair comparison, we test 4 very popular descriptors: LBP, LPQ, BSIF and SIFT in computer vision on numerous benchmark databases including AR, FERET, GT databases, as well as our own collected data from wearable devices like the GG. We then rigorously test these descriptors in conjunction with the chosen distance measures, as well as support vector machines in every possible combination in order to obtain reliable results. Through this process, we hope to pave the way for future work which utilizes FI in wearable devices such as GoPros and the GG, by providing reliable and efficient descriptors and distance measures.

In the following subsection we first study the related work and then in Section 2, we describe the local descriptors and distance measures along with 1-NN and SVM as classifiers. In Section 3, we present the experimental results and analysis. In Section 4 we present the summary of the experimental results and finally conclude in Section 5.

1.1 Related Work

Face recognition (FR) is a very challenging problem. Despite extensive research over the last four decades, the problem is still far from solved in unconstrained environments [6, 7]. A systematic independent evaluation of recent face recognition algorithms from commercial and academic institutions can be found in the face recognition vendor test (FRVT) 2013 report [8]. Using it on the wearable devices like GG, makes the problem still more challenging. In addition to all the traditional problems of FR (like uneven illumination condition, pose, expression and aging) [6, 9], FPVs possess blurry or jittering and out of focus images. This is because both the camera and target (face) are always on moving or non-stationary platforms.

FI on wearable devices is gaining very popularity because of its wide range of applications like an assistant in social interaction, memory aid for people who cannot remember faces or unable to recall names of the person whom he/she meet before, business networking and profession cooperation [10, 11]. In such situations, keeping a log about the people you interacted with during the day to augment your memory and help you remember better is useful. Till date, all the FR studies have been focused on benchmark face images third person view (TPV) data but with the availability of wearable devices, many more FPV are generated, stored, used and shared by the users. So it is important that we evaluate various local features and distance measures for FPV videos.

Utsumi *et al.* proposed a wearable FR system in [12], which uses a course-to-fine recognition method. Their system requires a desktop PC because of the high computational cost of the various algorithms. Krishna *et al.* [13], developed an iCare Interaction Assistant device for helping visually impaired individuals for social interactions. Their evaluations are limited to only 10 subjects' face images captured under tightly controlled and calibrated face images using classical subspace methodologies like principal component analysis (PCA) [14], linear discriminant analysis (LDA) [15] and Bayesian interpersonal classifier (BIC) [16] and have not evaluated the performance using recent local descriptors.

Face detection and recognition involving various scales and orientation are proposed in [17]. They use a color camera and an infrared camera for capturing face images. They have used hidden markov model on a very small number of subjects to perform FR. Their system is bulky and cumbersome, far away from any practical system. Wang *et al.* proposed a FR system for improving social lives of Prosopagnosics (people with inability to recognize faces or distinguish facial features that differentiate people) [10]. They have used LBP features for development of their FR system, however, LBP is very sensitive to noises and blurry images. For this system also, the performance evaluation is limited to 20 subjects only. A well-known performance evaluation of local descriptors has been done in [18]. Their evaluations are performed on images of natural scenery, texts and buildings, etc and not on face images.

Many researchers have used LBP, SIFT, LPQ and BSIF for solving various computer vision problems. For example, LBP has been used for FR [19]. SIFT has been used for detecting and extracting scale invariant features for face classification [20]. LPQ is used for texture classification in blurry images [21]. Recently, Kannala *et al.* proposed BSIF in [22] to perform the texture recognition better than LBP and LPQ. To the best of our knowledge BSIF has never been used for FR. In this work, we evaluate these 4 popular local features with 10 different distance measures/classifier like SVM to find out which combination works best for improving FR accuracy in FPV videos and also for benchmark face image databases. In the next section, we first discuss each descriptors briefly and then perform their evaluations on various databases to study the effectiveness of these descriptors in FR.

2 Local Descriptors, Distance Measures and Classifiers

In order to compare and differentiate between faces belonging to different individuals, one can use descriptors to describe each face. There are 4 very common local descriptors being used in FR: LBP, SIFT, LPQ and BSIF and 10 different distance measures and SVM in the literature. Below we describe each of these local descriptors, distance measures and SVM classifier briefly.

2.1 Face Image Descriptors

Local Binary Patterns

Since faces can be seen as a composition of micropatterns, we can describe these micropatterns using the local binary patterns (LBP) operator. This operator assigns a label to every pixel of an image by thresholding the 3×3 (*i.e.* for a 3 by 3 sized filter; this can be predefined by the user) neighborhood of each pixel with the center pixel value and considering the result as a binary number. Finally, we obtain the histogram of the labels as a structure which can be used as a texture descriptor [19].

Scale Invariant Feature Transform

With wearable devices, we often have different views of the same subject, which can result in reduction of classification accuracy as a different view can cause a subject's identity to be misinterpreted. With Scale Invariant Feature Transform (SIFT), we have features that are invariant to scale and orientation and thus are highly distinctive of the subject. This allows us to extract features which provides us with reliable matching between varying views of the same subject. This operator first computes the locations of potential interest points in the image by obtaining the maxima and minima of a set of Difference of Gaussian filters applied at varying scaled throughout the image. Next, we discard points of low contrast in order to refine the locations. We then assign an orientation to each key point based on local image features. Finally, we compute a local feature descriptor at every key point, which is based on the local image gradient and transformed according to the orientation of the key point, allowing us to have orientation invariance [20].

Local Phase Quantization

In FPV, images are often blurry because of the camera motion and the target (face) object motion. Also, the images obtained are often out of focus because the wearable camera (such as GG) takes some time to adjust/focus to the object in view. Since image deblurring is difficult and introduces new artifacts, we use a blur insensitive descriptor, the local phase quantization (LPQ) operator [21]. This operator first decorrelates the image, as information is maximally preserved in scalar quantization if the samples to be quantized are statistically independent. Next, short-term Fourier Transform is performed on the image in

every 3×3 (*i.e.* for a 3 by 3 sized filter; this can be predefined by the user) neighborhood of each pixel. Again, we obtain the histogram of the result as a structure which can be used as a texture descriptor. The resultant code is insensitive to centrally symmetric blur [21].

Binarized Statistical Image Features

Unlike LBP and LPQ where we are required to manually define the filters, binarized statistical image features (BSIF) has pre-defined texture filters which we can utilize [22]. The BSIF operator first convolves the image with the pre-defined texture filters and binarizes the filter responses using a threshold of zero. The texture filters are learnt from a training set of natural image patches by maximizing the statistical independence of the filter responses. Research also suggests that the results of BSIF can be further improved if the pre-defined texture filters can be tailored to fit images that have unusual characteristics, such as certain medical images of specific sections of the human anatomy [22].

2.2 Distance Measures and Classifiers

In this work, we do not intend to use any training or learning mechanism. Rather, we want to evaluate the effectiveness of various local features with different distance measures for FI task. We evaluate two classifiers: (i) various distance measures with 1-nearest neighbor (1-NN) and (ii) features from local descriptors with SVM.

Distance Measures

We extract the features from face images using various descriptors and then use distance measures to define faces belonging to different and same people. This is done by computing the distance between two distinct faces. In this paper, we study and analyze Euclidean, Cosine (angle-based), Chi square, Spearman, Cityblock, Minkowski, Correlation, Hamming, Jaccard and Chebychev distance measures [23]. Each of these is rigorously tested with the 4 local descriptors previously mentioned and the results are provided in the Experimental Results and Analysis section.

Support Vector Machines

Another method for determining if faces belong to the same or different person using various descriptors is support vector machines (SVM) [24]. SVMs are a useful technique for data classification. A data classification task involves separating data into training and testing sets. Each instance in the training set contains one "target value" (i.e. the identity of the person which the specific face belongs to) and several "attributes" (i.e. the features of the specific face). SVM produces a model based on the training data which predicts the target values of the test data given only the test data attributes [24].

3 Experimental Results and Analysis

We evaluate the performance of 4 local descriptors using 10 distance measures and classifiers on 4 benchmark databases (TPV face data) and 1 our own collected wearable device database (FPV face data) for face identification purpose. The databases used to test these methods are 2 sets of Facial Recognition Technology (FERET) database [25], Aleix Martinez and Robert Benavente (AR) database [26], Georgia Tech (GT) database [27] and our own collected wearable device database [28]. In all the experiments, images are preprocessed following the CSU Face Identification Evaluation System [29]. For all the local descriptors, default parameters are set accordingly to the results reported in their respective references. They are kept same for all experiments across all the databases reported in this paper. Out of 10 distance measures mentioned before, we filter and present only top 6 best performing distance measures on each of the databases.

3.1 Results on FERET Database 1

There are 2,388 images comprising of 1,194 persons (two images FA/FB per person) selected from the FERET database [25]. Images are cropped into the size of 33×38 similar to [30–33]. We evaluate the performance of 4 local descriptors used in conjunction with 10 distance measures as well as SVMs. We use the first image of each subject as the gallery image and the second image as the probe image. So there are 1,194 gallery images and 1,194 probe images. After using the LBP and LPQ descriptors on the face images, we have an array of 1194 by 256 for both the gallery and probe image sets. For BSIF, we obtain an array of 1194 by 4096 for both the sets. We then use each descriptor on the images and match them by calculating the distance between the gallery image and all the probe images. Next, we apply a simple first nearest neighbor (1-NN) classifier to test the efficiency of the descriptors by calculating the recognition rates. We also apply SVM on the features obtained from the local descriptors to calculate the recognition rates. The recognition rates with top 6 best performing distance measures are recorded in Table 1.

Table 1. Recognition rates (%) using various image descriptors vs. different distance measures/classifier for face recognition on FERET database 1.

	Distance Measures with 1-NN as classifier						Classifier
Descriptors	Euclidean	Cosine	Chi square	Cityblock	Correlation	Spearman	SVM
LBP	43.2	43.2	50.8	50.1	42.5	7.7	18.7
LPQ	60.6	63.2	65.7	66.4	64.1	65.0	64.5
BSIF	80.2	84.8	90.6	91.0	86.8	86.9	73.0

We use the SIFT descriptor on the images and compare the features of two face images by finding the closest descriptor between the two and record the distance between the pair. Using this method, we calculate the distances between

the gallery images of each individual and all the test images. A simple first nearest neighborhood classifier (1-NN) is applied to test the face descriptors using SIFT. We also filter the matches for uniqueness by adding a threshold to the comparison algorithm ("the uniqueness is measured as the ratio of the distance between the best matching key point and the distance to the second best key point" [34]). This threshold improves the classification accuracy by rejecting matches which are too ambiguous [34]. The recognition rates against the threshold values are shown in Fig. 1.

Without tuning any parameters[1], both the raw SIFT and LBP features on this database with large number of subjects perform very poorly as shown in Fig. 1 and Table 1. Also, BSIF with SVM as classifier does not perform well on this database as shown in Table 1. This is probably because SVM, which was originally meant for binary classification, in general, does not perform well on databases with large number of classes or for large multi-class problem (this is also evident in other experimental results). However, Table 1 shows that BSIF with Cityblock distance measures and 1-NN as classifier can achieve around 91% accuracy, outperforming all other features with different distance measures.

3.2 Results on FERET Database 2

This database is a subset of the original FERET database [25], created by choosing 256 subjects with at least four images per subject. However, we use the same number of images (four) per subject for all subjects similar to that used in [35, 30, 36, 32]. We use the first image of every individual as the gallery/training set and the remaining three images as the probe/testing set. Since there are 256 people in the dataset, there are 256 images in the training set and 768 images in the testing set. For this database, we apply different feature extractors such as LBP, SIFT, LPQ and BSIF operators with the various distance measures. Using the LBP/LPQ descriptor on the images, we have an array of 256×256 for the training set and an array of 768×256 for the testing set as the LBP descriptor uses 256 features, whereas BSIF resulted in obtaining 4096 features from each of the face images. The results using various descriptors are shown in Table 2.

Similar to the previous experiment with SIFT, we use this operator on this database to obtain the features and then perform recognition of individuals with varying threshold values. The results are shown in Fig. 1.

It can be seen from Table 2 and Fig. 1 that LBP, LPQ and SIFT features performs similar. For SIFT features the thresholding plays an important role for this database. However, the highest recognition rates can be obtained using BSIF features with Chi square and Cityblock distance measures and 1-NN as

[1] A test was performed with LBP and the Chi square distance measure with a different filter size and 1-NN as classifier, producing a classification accuracy of 60%. The difference of 10% (in Table 1) showcases the significance of fine-tuning the parameters in LBP. However, in this work, we are not focusing on fine tuning the parameters for LBP, but use same default parameters for all the experiments. This is also same for all other descriptors including SIFT, LPQ and BSIF.

Table 2. Recognition rates (%) using various image descriptors vs. different distance measures/classifier for face recognition on FERET database 2.

	Distance Measures with 1-NN as classifier						Classifier
Descriptors	Euclidean	Cosine	Chi square	Cityblock	Correlation	Spearman	SVM
LBP	81.8	82.8	87.9	86.7	82.8	4.5	32.6
LPQ	46.5	92.4	93.2	93.8	92.0	89.3	64.1
BSIF	95.9	96.9	99.0	98.8	96.7	98.4	82.3

classifier. It is also notable that using BSIF features with such distance measures and classifier, the recognition rates outperform all other state-of-art methods on this database [30, 35, 36].

3.3 Results on AR Database

The AR database has frontal view faces with varying facial expressions and illumination conditions. The color images in AR database [26] are converted to gray scale and cropped into the size of 120×170, same as the image size used in [26, 37, 32]. In this database, we have 75 subjects, with 14 images each. We evaluate the performance of 4 local descriptors used in conjunction with 10 distance measures as well as SVMs. We store 7 images from each subject in the gallery set while the remaining 7 images per subject are used as probe images [26, 37, 32]. We then use each descriptor on the images and match them by calculating the distance between the gallery images and all the probe images. After using the LBP and LPQ descriptors on the face images, we have an array of 525 by 256 for both the gallery and probe image sets, which are then subjected to the distance measures as well as SVM. For BSIF, we obtain an array of 525 by 4096 for both the sets.

Table 3. Recognition rates (%) using various image descriptors vs. different distance measures/classifier for face recognition on AR database.

	Distance Measures with 1-NN as classifier						Classifier
Descriptors	Euclidean	Cosine	Chi square	Cityblock	Correlation	Spearman	SVM
LBP	63.2	64.4	73.5	72.8	64.4	19.6	70.1
LPQ	84.4	84.4	88.2	88.4	85.5	83.0	94.9
BSIF	92.8	94.7	98.7	98.7	94.7	97.5	99.8

We apply a simple first nearest neighbor (1-NN) classifier to test the efficiency of the descriptors by calculating the recognition rate. The recognition rates for various local features vs. different distance measures and SVM are recorded in Table 3. Similar to the previous experiments with SIFT, we use this operator on this database to obtain the features and then perform recognition of individuals with varying threshold values. The results are shown in Fig. 1.

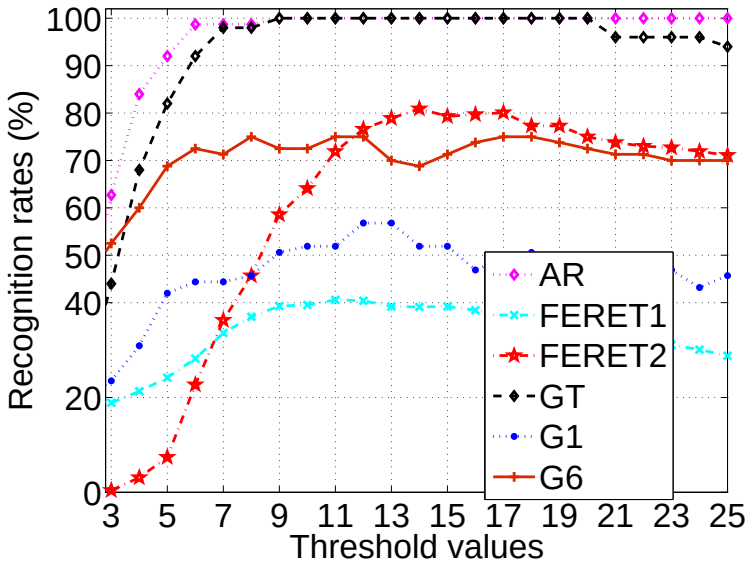

Fig. 1. Recognition rates vs Threshold values used in the matching on all six databases: AR, FERET 1, FERET 2, GT, G1 and G6 using SIFT features (best viewed in color).

From Fig. 1, it is evident that SIFT with thresholds 6 and above and BSIF features with Chi square and Cityblock distance measures with 1-NN as classifier in Table 3, perform best on this AR dataset. Also, BSIF features with SVM as classifier (in Table 3) perform very good probably because there are no changes in pose for this database. They also outperform the present state-of-art results of FR on this AR database [38, 30, 36].

It is also notable that the recognition accuracy does not vary much with change in the distance measures. The change in the accuracy occurs across different features obtained from the descriptors. Perhaps this is because the database has variations of illuminations and expressions with same frontal pose face images. Table 3 and Fig. 1 show that BSIF and SIFT are superior local descriptors of face images when there are no changes in pose as compared to LBP and LQP.

3.4 Results on GT Database

The Georgia Tech (GT) Face Database [27] consists 750 color images of 50 subjects (15 images per subject). These images have large variations in both pose and expression and some illumination changes. Images are converted to gray scale and cropped into the size of 92×112. The first eight images of all subjects are used in the training and the remaining seven images serve as testing images. This protocol is same as done in [30, 31, 39]. The testing results are numerically recorded in Table 4. Similar to the previous experiments with SIFT, we use this

operator on this database to obtain the features and then perform recognition of individuals with varying threshold values. The results are shown in Fig. 1.

Table 4. Recognition rates (%) using various image descriptors vs. different distance measures/classifier for face recognition on GT database.

	Distance Measures with 1-NN as classifier						Classifier
Descriptors	Euclidean	Cosine	Chi square	Cityblock	Correlation	Spearman	SVM
LBP	59.1	58.6	65.4	65.7	58.6	13.4	61.7
LPQ	75.4	76.3	76.6	76.6	74.0	73.1	84.0
BSIF	88.3	88.6	92.3	92.3	86.9	90.0	94.9

For this database, there are large changes in pose with varying expressions and lighting conditions. Local descriptor BSIF with SVM as classifier outperforms both LBP and LPQ with various distance measures with 1-NN and also their raw features with SVM as classifier. Since this database has only 50 subjects, it probably shows that when the number of subjects are few, SVM performs better than 1-NN classifier for all local descriptors, except LBP. Similar to the AR database, SIFT outperforms all other features with all distance measures and also SVM on BSIF features as recorded in Table 4 and shown in Fig. 1.

3.5 Wearable Device Database

This database is collected to study the problems of FR in wearable devices and it is publicly available at [28]. It contains faces of persons observed from FPV in natural social interactions, where people are involved in group meetings, indoor social interactions, business networking and all other activities in indoor office environment. There are large changes in poses, expressions, illuminations and jitters because of head and/or camera movement. Collected between Sep 2012 to Aug 2014, it comprises of 7075 images of 88 subjects (average 80.4 images per subject). Out of which 46 subjects are collected using head mounted Logitech C190 webcam connected to a tablet and rest 42 subjects by using first version of the Google Glass [2]. The database is composed of 9 females and 79 males across 9 races. Face and eye detections [40, 41] are applied to the color images captured by the wearable devices. Face images are then converted to gray scale and cropped into the size of 67×75. One sample image captured by GG and the extracted and normalized face images are shown in Fig. 2. The red box shown is the face image where both the eye coordinates are successfully detected and blue box shows a face in which either one of the eye coordinates is not detected.

Protocol

We evaluate the performance of 4 local descriptors using various distance measures for two applications of FR. In the first scenario, only 1 frontal face image

Fig. 2. Left, original image captured by Google Glass. Right, extracted normalized face images (red boxes: both eye coordinates are detected, blue box: either of the eyes is not detected). (Best viewed in color.)

per subject is available in the gallery, while remaining all face images are used as probes (termed as G1). This is similar to the commercial database of personal information containing only one mug shot image for each person. As mentioned previously and presented in the recent state-of-the-art wearable FR devices [13, 12], that keeping one mug shot image in the gallery may not be suitable for wearable FR for natural social interactions. However, in this work we perform experiments for such challenging scenarios. The FI accuracy with various features and distance measures are recorded in Table 5.

Table 5. Recognition rates (%) using various image descriptors vs. different distance measures/classifier for face recognition on wearable device database with G1 scenario.

	Distance Measures with 1-NN as classifier						Classifier
Descriptors	Euclidean	Cosine	Chi square	Cityblock	Correlation	Spearman	SVM
LBP	72.8	65.4	72.8	59.3	70.4	60.5	32.3
LPQ	63.0	61.7	71.6	66.7	69.1	58.0	32.0
BSIF	76.5	71.6	71.6	75.3	67.9	69.1	35.2

In the second scenario, 6 images with varying pose, expression and illumination per subject are stored in the gallery, while remaining all images are used as probes (termed as G6). The recognition rate of various features with different distance measures/classifier are presented in Table 6.

Similar to the previous experiments with SIFT, we use this operator on our wearable device database to obtain the features and then perform recognition of individuals with varying threshold values. We tested the threshold values over

Table 6. Recognition rates (%) using various image descriptors vs. different distance measures/classifier for face recognition on wearable device database with G6 scenario.

	Distance Measures with 1-NN as classifier						Classifier
Descriptors	Euclidean	Cosine	Chi square	Cityblock	Correlation	Spearman	SVM
LBP	82.7	81.5	85.2	86.4	82.7	82.7	58.7
LPQ	76.5	80.2	82.7	84.0	81.5	80.2	61.2
BSIF	86.4	87.7	87.7	86.4	84.0	84.0	59.4

a range and obtained the results on both the above scenarios G1 and G6. The recognition rates against the threshold values are shown in Fig. 1.

For this database, all the descriptors in G6 scenario have better performances than G1 scenario because of the availability of more number of samples in the gallery. It is evident from Fig. 1 that the SIFT features do not perform good on this unconstrained FPV face images. This is because the noises and artifacts in FPV face images are far more than face images in the standard benchmark databases (like FERET, AR and GT). This probably shows that SIFT features are very sensitive to blurry and jittery images, like that in the FPV videos.

In general, BSIF with various distance measures and 1-NN as classifier, has superior performances as compared to LBP and SIFT on this wearable device database for both the G1 and G6 scenario, while LPQ's performance is comparable to LBP. This shows that the features from BSIF are less sensitive to various unconstrained face image conditions, such as blurry, jittery and out of camera focus face images (in addition to the traditional FR problems such as pose, illumination and expression). BSIF with Euclidean, Cosine and Cityblock distance measures outperform all other features for this database for both G1 and G6 scenarios as shown in Tables 5 and 6.

4 Summary of the Experimental Results

We have performed comprehensive evaluation of 4 local descriptors in combination with 10 distance measures with 1-NN and SVM as classifiers on 6 databases. Without using any learning/training mechanism, these raw descriptors are evaluated for FI (multi-class classification) task. Each of these databases has its own challenges for performing FI. BSIF features when used with SVM as classifier, is observed to perform well on databases with small number of subjects, such as the AR and GT databases. Also, this is same for SIFT features, which performs well when the number of subjects in the database is small as shown in Fig. 1. One notable fact is that, the recognition performance is largely dependent on features that are selected rather than the distance measures. For example, in GT database (Table 4), LPQ outperforms LBP largely because of nature of the features. The performance does not change much with varying distance measures. Similar observations are noted between BSIF and LPQ, as shown in Table 4 and FERET databases in Tables 1 and 2.

In FERET database 2 (Table 2), BSIF using Chi square and City block distance measures with 1-NN as classifier outperform all other state-of-art FR results on this database [30, 35, 36]. Also, on AR database (Table 3 and Fig. 1), SIFT with thresholds 6 and above and BSIF features with Chi square and Cityblock distance measures with 1-NN as classifier and raw BSIF features with SVM as classifier, outperform the present state-of-art results of FR on this AR database [38, 30, 36].

From Tables 1-6, it is evident that both BSIF and LPQ, in general, perform better with different distance measures on most of the databases as compared to LBP. This shows that unlike LPQ and BSIF, LBP is very much sensitive to its parameters tuning (which is not done in this work). It is also evident that BSIF with Cosine, Chi square and Cityblock distance measures, in general, outperforms all other features on all databases. For wearable device database, it seems SVM does not provide good results as the face images are captured in unconstrained environment. BSIF is shown to be more robust to blurry, jittery and out of camera focus face images (in addition to the traditional FR problems such as pose, illumination and expression), as it exhibits superior performance to all other local descriptors in all the databases with 1-NN as classifier.

5 Conclusions

In the past few decades, many researchers have evaluated local descriptors like LBP, LPQ and SIFT for different computer vision problems including FI. To the best of our knowledge, BSIF has never been used for FI task. Also, the evaluations of these four local descriptors for FI in FPV or ego-centric views data are largely unknown in the literature. In this paper, we have evaluated local descriptors using various distance measures and classifiers 1-NN and SVM on wearable devices and benchmark face databases. This helps us to understand the performance of various local descriptors for FI task under the common framework. Through this process, we hope to pave the way for future work which utilizes FI in wearable devices such as GoPros and the GG, by providing reliable and efficient descriptors and distance measures. Among these descriptors, BSIF with Cosine, Chi square and Cityblock distance measures using 1-NN as classifier are superior to all other descriptors and distance measures on both benchmark and FPV video data.

References

1. GoPro. http://gopro.com/ (2014)
2. Google: Google glass. http://www.google.com/glass/start/ (2014)
3. Mandal, B., Eng., H.L.: 3-parameter based eigenfeature regularization for human activity recognition. In: IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP). (2010) 954–957
4. TedBlog: The future of facial recognition: 7 fascinating facts. http://blog.ted.com/2013/10/17/the-future-of-facial-recognition-7-fascinating-facts/ (2014)

5. Mandal, B., Eng, H.L.: Regularized discriminant analysis for holistic human activity recognition. IEEE Intelligent Systems **27** (2012) 21–31
6. Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face recognition: A literature survey. ACM Computing Surveys **35** (2003) 399–458
7. Phillips, P.J.: Face & ocular challenges. Presentation: http://www.cse.nd.edu/BTAS_10/BTAS_Jonathon_Phillips_Sep_2010_FINAL.pdf (2010)
8. Grother, P., Ngan, M.: Face recognition vendor test (frvt) performance of face identification algorithms. Technical Report: http://biometrics.nist.gov/cs_links/face/frvt/frvt2013/NIST_8009.pdf (2014)
9. Mandal, B., Jiang, X.D., Kot, A.: Multi-scale feature extraction for face recognition. In: IEEE International Conference on Industrial Electronics and Applications (ICIEA). (2006) 1–6
10. Wang, X., Zhao, X., Prakash, V., Shi, W., Gnawali, O.: Computerized-eyewear based face recognition system for improving social lives of prosopagnosics. Proceedings of the 7th International Conference on Pervasive Computing Technologies for Healthcare (2013) 77–80
11. Mandal, B., Ching, S., Li, L., Chandrasekha, V., Tan, C., Lim, J.H.: A wearable face recognition system on google glass for assisting social interactions. In: 3rd International Workshop on Intelligent Mobile and Egocentric Vision, ACCV. (2014)
12. Utsumi, Y., Kato, Y., Kunze, K., Iwamura, M., Kise, K.: Who are you?: A wearable face recognition system to support human memory. In: ACM Proceedings of the 4th Augmented Human International Conference. (2013) 150–153
13. Krishna, S., Little, G., Black, J., Panchanathan, S.: A wearable face recognition system for individuals with visual impairments. In: ACM SIGACCESS Conf. on Computer and Accessbility. (2005) 106–113
14. Turk, M., Pentland, A.: Eigenfaces for recognition. Journal of Cognitive Neuroscience **3** (1991) 71–86
15. Swets, D.L., Weng, J.: Using discriminant eigenfeatures for image retrieval. IEEE PAMI **18** (1996) 831–836
16. Moghaddam, B., Jebara, T., Pentland, A.: Bayesian face recognition. Pattern Recognition **33** (2000) 1771–1782
17. Singletary, B.A., Starner, T.E.: Symbiotic interfaces for wearable face recognition. In: In HCII2001 Workshop On Wearable Computing. (2001)
18. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE PAMI **27** (2005) 1615–1630
19. Ahonen, T., Hadid, A., Pietikainen, M.: Face description with local binary patterns: Application to face recognition. IEEE PAMI **28** (2006) 2037–2041
20. Aly, M.: Face recognition using sift features. CNS/Bi/EE report **186** (2006)
21. Ojansivu, V., Heikkil, J.: Blur insensitive texture classification using local phase quantization. Image and Signal Processing **5099** (2008) 236–243
22. Kannala, J., Rahtu, E.: Bsif: Binarized statistical image features. In: ICPR. (2012) 1363–1366
23. Perlibakas, V.: Distance measures for pca-based face recognition. Pattern Recognition Letters **25** (2004) 711–724
24. Hsu, C., Chang, C., Lin, C.: A practical guide to support vector classification (2010)
25. Phillips, P.J., Moon, H., Rizvi, S., Rauss, P.: The feret evaluation methodology for face recognition algorithms. IEEE PAMI **22** (2000) 1090–1104

26. Martinez, A.M.: Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. IEEE PAMI **24** (2002) 748–763
27. Nefian, A.V.: Georgia tech face database. http://www.anefian.com/research/face_reco.htm (2014)
28. Mandal, B., Ching, S., Li, L.: Werable device database. https://sites.google.com/site/bappadityamandal/human-detection-and-fr (2014)
29. Beveridge, R., Bolme, D., Teixeira, M., Draper, B.: The csu face identification evaluation system users guide: Version 5.0. Technical Report: http://www.cs.colostate.edu/evalfacerec/data/normalization.html (2013)
30. Jiang, X.D., Mandal, B., Kot, A.: Eigenfeature regularization and extraction in face recognition. IEEE PAMI **30** (2008) 383–394
31. Jiang, X.D., Mandal, B., Kot, A.: Face recognition based on discriminant evaluation in the whole space. In: IEEE 32nd International Conference on Acoustics, Speech and Signal Processing (ICASSP 2007), Honolulu, Hawaii, USA (2007) 245–248
32. Mandal, B., Jiang, X., Eng, H.L., Kot, A.: Prediction of eigenvalues and regularization of eigenfeatures for human face verification. Pattern Recognition Letters **31** (2010) 717–724
33. Mandal, B., Jiang, X.D., Kot, A.: Dimensionality reduction in subspace face recognition. In: IEEE ICICS. (2007) 1–5
34. VLFEAT: Vlfeat open source. www.vlfeat.org/overview/sift.html#tut.sift.param (2014)
35. Lu, J., Plataniotis, K.N., Venetsanopoulos, A.N., Li, S.Z.: Ensemble-based discriminant learning with boosting for face recognition. IEEE TNN **17** (2006) 166–178
36. Jiang, X.D., Mandal, B., Kot, A.: Complete discriminant evaluation and feature extraction in kernel space for face recognition. Machine Vision and Applications, Springer **20** (2009) 35–46
37. Park, B.G., Lee, K.M., Lee, S.U.: Face recognition using face-arg matching. IEEE Trans. Pattern Analysis and Machine Intelligence **27** (2005) 1982–1988
38. Geng, C., Jiang, X.: Fully automatic face recognition framework based on local and global features. Machine Vision and Applications **24** (2013) 537–549
39. Mandal, B., Jiang, X.D., Kot, A.: Verification of human faces using predicted eigenvalues. In: 19th International Conference on Pattern Recognition (ICPR), Tempa, Florida, USA (2008)
40. Viola, P., Jones, M.: Robust real-time face detection. In: IJCV. Volume 57. (2004) 137–154
41. Yu, X., Han, W., Li, L., Shi, J., Wang, G.: An eye detection and localization system for natural human and robot interaction without face detection. TAROS (2011) 54–65