RESEARCH   **Open Access**

GRAB: generalized region assigned to binary

Archana Sapkota[1*] and Terrance E Boult[1,2*]

Abstract

Scale is one of the major challenges in recognition problems. For example, a face captured across large distances is considerably harder to recognize than the same face at small distances. Local binary pattern (LBP) and its variants have been successfully used in face detection, recognition, and many other computer vision applications. While LBP features are shown to be discriminative in face recognition, the pixel level description of LBP features is sensitive to the change in scale of the images. In this work, we extend the utility of a generalized variant of LBP feature descriptor called generalized region assigned to binary (GRAB), previously introduced in an article below, and show that it handles the challenges due to scale. The original LBP operator in another article is defined with respect to the surrounding pixel values while the GRAB operator is defined with respect to overlapping surrounding regions. This gives more general description and flexibility in choosing the right operator depending on the varying imaging conditions such as scale variations. We also propose a way to automatically select the scale of the GRAB operator (size of neighborhood). A pyramid of multi-scale GRAB operators is constructed, and the operator at each scale is applied to an image. Selection of operator's scale is performed based on the number of stable pixels at different levels of the multi-scale pyramid. The stable pixels are defined to be the pixels in the images for which the GRAB value remains the same even as the GRAB operator scale changes. In addition to the experiments in the former article, we apply basic LBP, Liao et al.'s multi-scale block (MB)-LBP, and GRAB operator on face recognition across multiple scales and demonstrate that GRAB significantly outperforms the basic LBP and is more stable compared to MB-LBP in cases of reduced scale on a subsets of a well-known published database of labeled faces in the wild (LFW). We also perform experiments on the standard LFW database using strict LFW protocol and show the improved performance of GRAB descriptor compared to LBP and Gabor descriptors.

Introduction

One of the theoretical challenges in recognition is the extraction of features, which are sufficiently discriminative in addition to being invariant to the variables like illumination, translation, rotation, scale, etc. This work presents a feature descriptor primarily to handle the challenges due to scale in addition to the challenges due to illumination and noise and applies the descriptor for face recognition at low-scale images. Scale is critical in unconstrained face recognition since, in general, subjects may be at different distances from the camera, and the difference between a subject at 4 ft and one at 40 ft is a 10-time change in scale.

In this work, we present a new description based on the original local binary pattern (LBP), which combines micro-structure and global structure, as well as the structure at multiple scales of the face images. We call this operator general region assigned to binary (GRAB) and use this operator to extract features for facial recognition in images of varied scales. The prior extensions to produce the 'multi-resolution' [1] LBP simply used a larger neighborhood 'circle' but sampled the raw pixels on that circle. While it did consider pixels at greater distances, sampling does not mimic changes in resolution or scale. Our neighborhood operator overcomes this limitation by defining the pixels in terms of varied sizes of overlapping regions.

What is the impact of scale on face recognition? We conducted a small experimental analysis to see the impact of scale on face images. To reduce the number of variables contributing to recognition score differences, we took a subset of images from labeled faces in the wild (LFW) database, normalized them to the size of 150×130, downscaled the images to multiple scales, and upscaled back to the same size. The gallery and the probe consisted of the same images from the same subjects. The only

---
*Correspondence: asapkota@vast.uccs.edu; tboult@vast.uccs.edu
[1] Department of Computer Science, University of Colorado at Colorado Springs, Colorado, USA
[2] Securics Inc, Colorado Springs, Colorado, USA

variable is the scale. Gallery consisted of images of size 150×130 while the probe consisted of images of size 15×13 and 30×26. We took basic LBP features [2,3], multi-block (MB)-LBP features [4], and GRAB features and used a support vector machine (SVM) classifier for classification. The gallery and probe consisted of 1,830 images from 610 subjects from the LFW dataset. We will discuss more about this subset of LFW database in the 'Experiments and results' section. We observed that GRAB features were more discriminative than the basic LBP features and MB-LBP features on low-scale images. At the scale of 30×26, 8 images were misclassified out of 1,830 images using LBP while with proper selection of GRAB scale, all 1,830 images were correctly classified. At the scale of 15×13, 252 images were misclassified out of 1,830 probe images while GRAB achieved 100% accuracy. At the scale of 150×130, LBP and GRAB both achieve 100% accuracy. Figure 1 shows the examples of misclassified samples using LBP while being correctly classified using GRAB. We also observed that the GRAB features are more stable across multiple scales compared to MB-LBP features. In [4], the authors did not use scale-selection algorithm and used boosting algorithm after extracting multiple MB-LBP features. Therefore, we compare our GRAB features with MB-LBP features across multiple scales. The results on Table 1 shows that GRAB is more stable than MB-LBP with change is scales. This analysis was done on a very small data with only scale as a variable. The impact of scale on the accuracy on bigger data with more variations can be huge.

This shows that choice of feature descriptor is critical on low-scale images.

Following are the main contributions of our work: 1. Definition of GRAB as a generalized operator for feature description; 2. Method for selection of operator's scale space; 3. Demonstration of higher accuracy of GRAB descriptor compared to existing methods on low-scale images.

Related work

A lot of work has been done in the past in describing meaningful and distinctive features in images that can be used for recognition. Local binary pattern (LBP) is an operator, which was originally used to extract a texture description from imagery and is widely used in face recognition. The operator assigns a label to every pixel of an image by thresholding the 3×3 neighborhood of each pixel with the center pixel value, resulting to a binary number [2,3]. The pixel level features thus obtained are combined in the form of histograms in various ways to generate the global features for the face description. LBP has been one of the best-performing descriptors as it contains the microstructure as well as the macrostructure of the face image. Despite its popularity, there are a number of shortcomings in the LBP approach, including sensitivity to noise, scale changes, and rotation of the image.

One of the extensions of LBP to produce the multiresolution LBP [1] uses a larger neighborhood circle but still samples the raw pixels on that circle. While it does consider pixels at greater distances, sampling does not

Figure 1 Impact of scale on face images. Misclassified faces using standard LBP and correctly classified using GRAB: Gallery and probe consisted of the same set of images with only difference in scale. Probe images are low-scale images which are resized to higher scale to match the size of gallery images as we are using the histogram-based method. Top row consists of probe images of size 30×26, and bottom row consists of probe images of size 15×13. Both gallery and probe images are resized to 13×150 for matching. Images on the left side are probe images, and the images on the right are gallery images. The images in the red box are the misclassified images using standard LBP, and the images in the green box are the correctly classified images using GRAB.

Table 1 Classification accuracy of LBP, MB-LBP, and GRAB on images

Features	150 × 130		30 × 26			15 × 13			
	G1, P1	G1, P1	G3, P1 *	G5, P1	G3, P3	G1, P1	G3, P3	G5, P3 *	G7, P7
GRAB	1	0.9956	1	1	1	0.8622	0.9685	0.9978	1
LBP	1	0.9956	-	-	-	0.8622	-	-	-
% Gain	*0*	*0*	*0.44*	*0.44*	*0.44*	*0*	*12.32*	*15.72*	*15.98*
MB-LBP	1	0.9956	0.9972	0.9945	1	0.8622	0.8950	0.9464	0.9994
% Gain	*0*	*0*	*0.28*	*0.55*	*0*	*0*	*8.21*	*5.43*	*0.06*

From a subset of LFW database with multiple scales. The gallery and probe images are the same; the only difference is the scale. All gallery images are of the size 130 × 150 whereas probe images are of the sizes 130 × 150, 30 × 26, and 15 × 13. The columns of the table show the multiple scales of the operators. For example, (G5, P1) means the scale of the operator is 5 for gallery and 1 for probe, which means gallery images are smoothed by window size of 5 to match the unknown smoothing present in the probe. The columns marked with asterisk are the operator scales automatically selected according to our scale-selection algorithm described later in this paper. Since there is no such selection mechanism in MB-LBP except boosting algorithm, we compared the algorithms on multiple scales. Since LBP does not allow the averaging operator, we mark those fields with hyphens. According to the results above, GRAB is more stable across scales compared to LBP and MB-LBP.

properly model changes in resolution or scale, which results in pixels being combined and not sampled. Consider what happens on a region with a fine binary texture, where sampling chooses one of the two binary colors but changes in scale actually mix the values into new shades/colors. In [5], this multi-resolution LBP is combined with novel color representations which combine RGB, YCbCr, and YIQ color spaces. The results did improve performance on the FRGC data, but that did not actually contain multiple resolutions so sampling artifacts in color space would impact those experiments.

Studies have introduced the concept of a MB-LBP to provide a more robust operator than LBP [4]. In MB-LBP, the average sum of image intensity is computed in each subregion around a center subregion. These average sums are then compared with the center block. They note that, 'MB-LBP can be viewed as a certain way of combination using eight ordinal rectangle features'. While MB-LBP does improve recognition by representing a mixture of microstructure and macrostructure of the image pattern, they did not study the impact of scale but rather focused on improving recognition at a fixed scale.

The more recently proposed BRIEF descriptors [6,7] use binary strings as the feature descriptors instead of using decimal value of binary strings as used in basic LBP and its other variants. The binary strings are defined on the smoothed patches. Binary tests between a pair of pixels are performed for the classification. Similar to our work, they highlight the importance of smoothing before extracting LBP-like features. However, they choose a fixed 9 × 9 window for the experiments. For face recognition, the limited pairs of sample points or test points, with a fixed smoothing window may not be sufficient. Our GRAB features provide sufficient information for face recognition across multiple scales.

LBP features have also been used in the past for face detection. The work in [8] used LBP features as a facial representation and built a face detection system using SVM as a classifier. Another example of the variant of LBP used for face detection is [9]. It uses multi-block local binary pattern features and the boosting algorithm for face detection [9].

Due to the peculiarities of the face shape and variability of several aspects of the face, the face recognition problem is different from the other object recognition problems. Some of the previous work used the combination of local as well as global representation of the face descriptors to solve this problem. Multi-resolution histograms of local variation pattern [10] is one such method which describes face images as the concatenation of the local spatial histogram of local variation patterns computed from multi-resolution Gabor features.

Gabor features are another interesting set of features which are highly applied in face recognition [11,12]. The Gabor representation of face images incorporates multi-scale feature extraction. The Gabor wavelet representation of an image is the convolution of the image with a family of Gabor wavelets at different scales; for example, Pinto et al. present a V1-like algorithm that considers 96 different Gabor filters. Local features are represented by the coefficient set, or Gabor jet, which orders the convolution results at different orientation and scales for a particular point.

Feature transform (SIFT) is a popular method in object recognition [13,14]. They extract the features of an image using the key points that are invariant to scale change. To detect such key points, they search the stable features across all possible scales using a scale space and such key points are associated with location, scale, and orientation information. To define the local image features, they sample the local image intensities around the key points at the appropriate scale of the key point. Bicego et al. used SIFT features for authentication in [15], wherein they used the distance between all pairs of keypoint descriptors in the two images to define the matching score. For face authentication, this type of algorithm was not as successful as it

was in other object recognition problems using SIFT-like features. Unfortunately, the planarity assumption underlying the theory of SIFT features and the highly non-planar and self-occluding nature of faces result in weak performance on face recognition tasks. In [16], SIFT features are combined in a mixed local-global strategy supporting a recognition-from-parts approach to address occlusion.

In this work, we present an operator called GRAB, developed as a generalization of LBP. While we will show the effectiveness of GRAB, like other multi-resolution approaches, there is likelihood that it will suffer the curse of dimensionality. There are techniques for reducing dimensionality. For example, Chan et al. [17] uses subspace techniques of LDA to help reduce the dimensionality of standard MLBP while maintaining or increasing the accuracy of the added dimensionality. In terms of added accuracy, they argue that, 'However, by directly applying the similarity measurement to the multi-scale LBP histogram, the performance will be compromised. The reason is that this histogram is of high dimensionality and contains redundant information'. While Chan et al. show impressive results, in this work, we use GRAB and scale-selection algorithm rather than MBLP to avoid sampling issues and will use SVMs for recognition, which remove the redundancy in a different, and generally more effective way. And again, our focus is on addressing recognition under scale changes, not just improving recognition rates.

GRAB

GRAB is developed as a basic operator for neighborhood modeling of a pixel. For the simple GRAB operator, with neighbors $j = 1, \ldots, n$, we let c stand for the center pixel and j for the neighbor pixel. For each pixel c, we can define the generalized binary representation as:

$$GR(c) = \sum_{j=1}^{n} g_j(p_c, p_j) \cdot 2^{(j-1)}$$

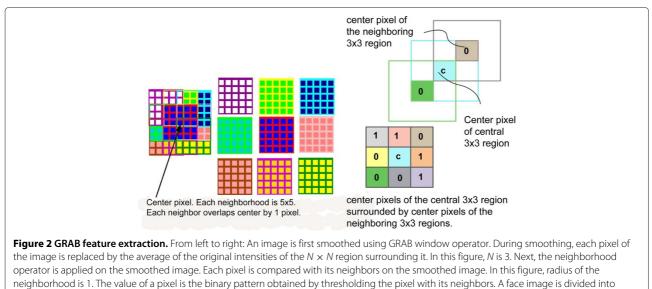$g_j(\cdot)$ is the generalized operator where,

$$g_j(\cdot) = 0 \ or \ 1 \qquad (1)$$

In this work, we consider a special case of the operator where:

$$g_j(p_c, p_j) = (p_j > p_c)$$
$$\text{and}$$
$$p_c = \frac{1}{N \times N} \sum_{i=1}^{N \times N} (p_i) \qquad (2)$$

We apply the abovementioned GRAB operator on a geo-normalized image as shown in Figure 2. First, an averaging operator is applied on the image with the window size $N \times N$. For each $N \times N$ region in the image, the center pixel of the region, p_c, is assigned an average value of that region resulting in a smoothed image. The implementation of this method is very efficient because of the use or integral image. Second, the neighboring operator is applied on the smoothed image. The neighboring pixel is given a label, 0 or 1, comparing its value with the center pixel value. If the neighboring pixel is greater than the center pixel value, a label 1 is assigned to it. At the borders of the image, where the neighbors do not exist, we set the values of pixels as zeros. Radius of the neighboring operator determines the overlap of the center region with its neighboring regions. The smaller the radius, the bigger is the overlap. For example, a 3×3 averaging operator with

center pixel of the neighboring 3x3 region

center pixels of the central 3x3 region

Center pixel of central 3x3 region

Center pixel. Each neighborhood is 5x5. Each neighbor overlaps center by 1 pixel.

center pixels of the central 3x3 region surrounded by center pixels of the neighboring 3x3 regions.

Figure 2 GRAB feature extraction. From left to right: An image is first smoothed using GRAB window operator. During smoothing, each pixel of the image is replaced by the average of the original intensities of the $N \times N$ region surrounding it. In this figure, N is 3. Next, the neighborhood operator is applied on the smoothed image. Each pixel is compared with its neighbors on the smoothed image. In this figure, radius of the neighborhood is 1. The value of a pixel is the binary pattern obtained by thresholding the pixel with its neighbors. A face image is divided into multiple regions and GRAB histograms are generated.

the neighboring radius of 1 has one-pixel overlap with the center pixel of the center region. The GRAB operator, the combination of averaging operator, and the neighboring operator, is shown in Figure 3.

This definition of GRAB does not use a single uniform definition as in local binary patterns, but it combines, in a more meaningful way, multiple different neighborhood rules. GRAB operator can be implemented as a generalization of ELBP [18] in the sense that the block averages around the center pixel can be arranged in circular or elliptical fashion. In this work, we consider a fast rectangular integral neighborhood definition.

GRAB as scale invariant operator

GRAB uses windowed operators for the neighborhoods instead of the pixels. In the standard LBP, the comparison is that of a pixel directly with its neighbors. The prior extensions to produce the multi-resolution LBP simply used a larger neighborhood circle but sampled the raw pixels on that circle. While it did consider pixels at greater distances, sampling does not mimic changes in resolution or scale. To address this, our neighborhood operators average the image over a region to define their values. We then define the averaging window and the idea of multi-scale GRAB. While the neighborhoods for averaging could be of any shape, use of rectangular regions allow use of summed area tables [19], also known as integral images, which allow very efficient computation of averages over rectangle regions.
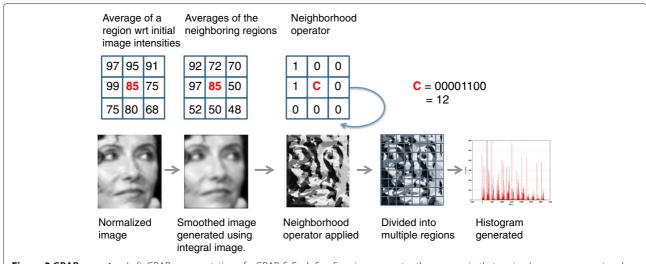
As an example, eight neighboring regions are labeled as in Figure 3. The regions use $N \times N$ rectangular average,
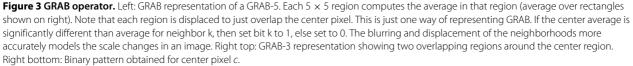
with one-pixel overlap where N is the size of GRAB window operator. For center pixel c, a region of size $N \times N$ is defined, and the average over the region is calculated. This value is assigned to the center pixel c. Similarly, for the neighboring regions of the same size, the average is computed. Now the average value of the central region, which is the value of the center pixel after the transform, is compared with the averages of the neighboring regions, and the threshold is applied to compute the labels of the neighboring pixels. The result is an 8-bit number representing one scale of neighborhoods around the point c. We can then compute a histogram, or partial histogram, of occurrence within the window. For face-based recognition we combine the histogram-based features for the multi-scale facial region description.
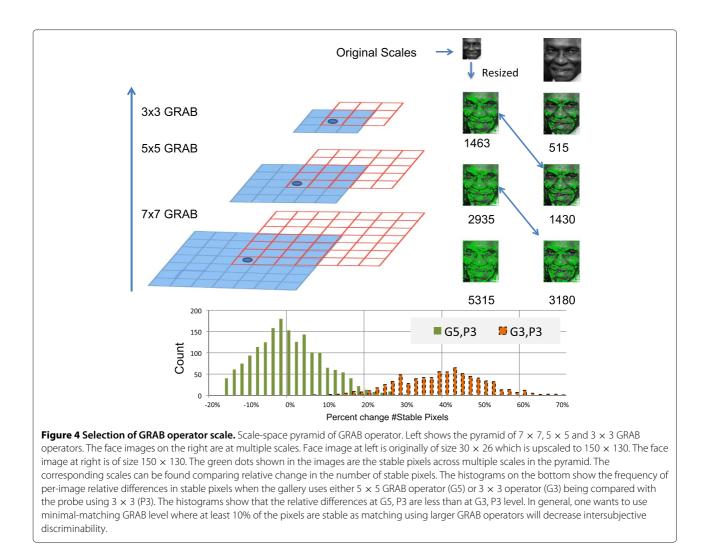
This multi-scale representation of GRAB descriptors allows it to account for the changes in spatial resolution in the images since we can store multiple scales at once. This makes facial recognition highly robust to changes in scale and also to changes in image quality.

Selection of GRAB scale

The previous work [20] did not explain the selection of GRAB operator scales. In this work, we propose a way of selecting the operator scale for matching images at multiple scales. A pyramid of GRAB operator at multiple scales is constructed as shown in Figure 4. For example, we start with the pyramid at 7×7 GRAB scale and move to 5×5 to 3×3. The choice of odd window regions is only for easy implementation. We define stable pixels as pixels for which the choice of GRAB operator scale does not

Figure 3 GRAB operator. Left: GRAB representation of a GRAB-5. Each 5×5 region computes the average in that region (average over rectangles shown on right). Note that each region is displaced to just overlap the center pixel. This is just one way of representing GRAB. If the center average is significantly different than average for neighbor k, then set bit k to 1, else set to 0. The blurring and displacement of the neighborhoods more accurately models the scale changes in an image. Right top: GRAB-3 representation showing two overlapping regions around the center region. Right bottom: Binary pattern obtained for center pixel c.

Figure 4 Selection of GRAB operator scale. Scale-space pyramid of GRAB operator. Left shows the pyramid of 7×7, 5×5 and 3×3 GRAB operators. The face images on the right are at multiple scales. Face image at left is originally of size 30×26 which is upscaled to 150×130. The face image at right is of size 150×130. The green dots shown in the images are the stable pixels across multiple scales in the pyramid. The corresponding scales can be found comparing relative change in the number of stable pixels. The histograms on the bottom show the frequency of per-image relative differences in stable pixels when the gallery uses either 5×5 GRAB operator (G5) or 3×3 operator (G3) being compared with the probe using 3×3 (P3). The histograms show that the relative differences at G5, P3 are less than at G3, P3 level. In general, one wants to use minimal-matching GRAB level where at least 10% of the pixels are stable as matching using larger GRAB operators will decrease intersubjective discriminability.

change the GRAB value at three different scales centering at that particular scale. For example, the stable pixels at 5×5 remain unchanged at scale 7×7, 5×5, and 3×3. We then compute the number of stable pixels for each scale. The change in number of stable pixels is tracked from one level of the pyramid to the next. Starting with the lower level in the pyramid (higher scale GRAB operator), when we track the GRAB values of image pixels which are stable across each pyramid levels, the change in the number of stable pixels across multiple scale operator decreases slowly in low-scale images and it decreases rapidly in high-scale images. This can be viewed as change in image gradient at multiple scales with respect to the neighborhood. During recognition, we match the probe and gallery image based on the relative change of stable pixels between image pairs. The match scales are considered to be the operator scales of probe and gallery images, at which the change in number of stable pixels are within a certain match criteria. While the larger GRAB operators produce more number of stable pixels, in general, one

wants to use minimal matching GRAB level where at least 10% of the pixels are stable as matching using larger GRAB operators will decrease intersubjective discriminability.

The operators at multiple scales can be automatically selected based on the number of stable pixels in the gallery model and the probe image. Figure 4 shows an example of scale-space pyramid of GRAB operators and the way of selection of scales based on the stable pixels.

Face description using GRAB

As mentioned in section 'GRAB', GRAB operator assigns a label to every pixel in the image by thresholding the center pixel with the pixel value of $N \times N$ block average by eight neighbors of $N \times N$ block average. The pattern thus obtained is a binary number and thus every pixel in the image is assigned such a number. Also, using the neighbor as a $N \times N$ block average does not affect the idea of uniform pattern. We can still make use of the uniform pattern which according to [2,3], is a binary pattern that contains at most two bit-wise transitions from 0 to 1 or vice versa,

when the bit pattern is considered circular. For example, the patterns 00000000 (zero transitions), 01110000 (two transitions) and 11001111 (two transitions) are uniform whereas the patterns 11001001 (four transitions) and 01010011 (six transitions) are not. We continued to use uniform pattern in our representation because it accounts for a larger percentage of the image representation in the face recognition technology (FERET) dataset [2,3], and we are using a subset of this dataset for our experiments. It also has the advantage of dimension reduction while using SVM. To represent the face image, the histogram of such patterns/binary numbers at different levels is used.

For face description using GRAB features, we use the same approach as LBP features because they represent the local and the global description of the face image. Geometrically normalized images, which are all 130-pixel wide and 150-pixel high, are divided into 64 regions (8 rows and 8 columns). GRAB-based histograms are computed in each region and are concatenated to form the global feature vector. To extend this idea to the multi-scale level, we actually compute GRAB histograms at different scales of the GRAB window operator. For example, for GRAB-3-5-7, the binary pattern was computed taking the block average of the 3×3, 5×5, and 7×7 neighbors. We then concatenate the histogram features of each scale to form the global histogram feature vector, which represents the local features and global features, as well as the features at different scales. While we could work in the space of smaller images, scaling down the windows, it is easier to conceptualize and implement, when we scale the different resolution images back to the same size, so all histograms are computed in the same manner and all 'window sizes' are in the same space with respect to facial geometry. All scale conversions for the work were done using ImageMagick's convert function.

We also verified the performance of LBP on the standard FERET partitions as mentioned in [2], achieving 96% on fafb, 47% on fafc, 57% on Dup1, and 48% on DupII without the weights assigned to the regions. The slight difference in the results could be due to the way the images are normalized.

We chose to use an SVM-based classification method to take advantage of the performance increase it offers over approaches traditionally used with LBP, such as the nearest neighbor [2]. We note the SVM used improves the performance of both LBP and GRAB, but the choice of machine learning classifier is not the critical aspect of this work (refer to Tables 2, 3, and 4 and section 'Experiments and results' later decribed in the paper to see the performance gain due to SVM over the nearest neighbor).

While the underlying models for the matching algorithm differ between our implementation and the standard LBP implementations, the processing of the images

Table 2 Performance of the nearest-neighbor classification on FERET240 and LFW610 with weighted regions

Image width	FERET240		LFW610	
	LBP	GRAB	LBP	GRAB
130	97.08	97.5	32.79	34.26
52	85.0	96.35	30.98	33.77
39	64.58	96.25	27.54	30.33
26	43.33	95.83	20.16	26.72
13	22.92	83.33	6.39	18.03

to generate a representative feature vector (as described in section 'Face description using GRAB') remains the same. Given feature vector representations for both training (gallery) and testing (probe) sets of images, the former set is used to train a multi-class SVM, while the latter set is subsequently tested against the trained model. In particular, we train the multi-class linear SVMs with default parameters($C = 1$) implemented via PyML. Concatenated LBP or GRAB histograms form the feature vectors, with each subject's gallery image being a positive example for the multi-class SVMs. We then test with similar feature vectors obtained from the probe images.

Experiments and results
LFW verification set
The labeled faces in the wild database provides the face images collected from the news articles on the web. It provides a protocol for face recognition where the recognition task is defined as a pair-matching problem. The database consists of 3,000 matched pairs and 3,000 non-matched pairs with 10-fold cross-validation. Each validation set consists of 5,400 training pairs, with 2,700 matched and non-matched pairs each and 600 testing pairs, with 300 match and non-match pairs. This is a binary classification problem where given a pair of images, decision is a 'match' or a 'non-match'. We use funneled version of the database [21], used the match and non-match sets provided by the database and followed 'Strict LFW' protocol. The original images are of the size 250×250. The face region is almost in the center in each image. We converted the images to gray scale and cropped to the size of 150×150 from the center using ImageMagick tool. We cropped the images such that the centers of 250×250 size images

Table 3 Performance comparison of LBP, Gabor, and GRAB on LFW database using strict LFW protocol

Feature descriptor	Performance
LBP	0.6625 +/- 0.0064
Gabor	0.6498 +/- 0.0066
GRAB	0.7090 +/- 0.0048

Table 4 Rank 1 recognition rate of GRAB , LBP, and V1-like algorithm

Image	GRAB	LBP	Gain	V1-like	Gain
130	99.17	98.75	0.4	97.5	0.17
52	98.83	94.58	4.49	89.17	10.83
39	98.83	88.75	11.35	69.17	42.87
26	96.67	75.0	28.89	26.25	268.2
13	83.33	46.25	80.17	0.42	19740

With the percentage improvement of GRAB over LBP and V1-like algorithm; this is on FERET240 dataset with gallery and probe images at different scales. The width of the probe images are in pixels in the table. The gallery image size is 130 × 150. Probe and gallery images have the same aspect ratio.

and 150 × 150 size images remain the same. This is to avoid the background information as much as possible while keeping the face region. We conducted experiments on LBP, Gabor [22], and GRAB features. The feature vector for an image pair $a - b$ in each set of experiment consists of $sqrt|f(a) - f(b)|$, where $f(a)$ is the feature vector from image a and $f(b)$ is the feature vector from image b. This experiment was conducted without applying our automatic scale selection algorithm. For both probe and gallery images, we use GRAB operators at 1×1, 3×3, and 7×7 scales and use linear SVMs for recognition.

Performance comparison on Table 5 shows the superior performance of GRAB over LBP and Gabor. While a lot of work has been proposed to improve the face recognition in unconstrained setting such as that of LFW, a choice of basic feature descriptor is critical, and for that, GRAB descriptor is a good alternative.

LFW610 and FERET240 subsets

We tested our proposed GRAB operator on subsets of two published datasets. The FERET [23] set was chosen due to extremely common use, allowing readers to do comparisons with many algorithms. It is, however, relatively constrained in nature: all images used were frontal and under fairly consistent lighting conditions. In order to provide a more robust, and realistic, set of experimental results for unconstrained face problems, the same tests were also run on a subset of LFW [24]. This set is relatively unconstrained and is generally considered one of the most difficult published set for facial analysis.

In our experiments, we use a model-based approach rather than a single-image-based approach. To reduce the potential for an outlier to have potentially disastrous effects on the training of the SVM, while still maintaining a relatively small gallery size and dealing with the limited number of views in the FERET protocol, we used three gallery images per subject.

Thus, the following protocol was designed and used for testing with both datasets: subjects for whom the dataset

contained fewer than four images were discarded. For each of the remaining subjects, a set of four images were chosen by an alphabetic sort on the names given in the original dataset. Of these four images, the first three comprised a subject's gallery; the last was used as a probe image. These subsets have been dubbed FERET240 and LFW610, respectively. For FERET, this ordering means the gallery generally included images from the FA and FB subsets while the probe is from the one of the more difficult sets (DUP1 or DUP2). For LFW, this ordering has no relation to standard sets or collection process. Because we use a multiple-image gallery for building the SVM, it was necessary to deviate from the published protocols for each data set. In addition, our effort is focused on recognition.
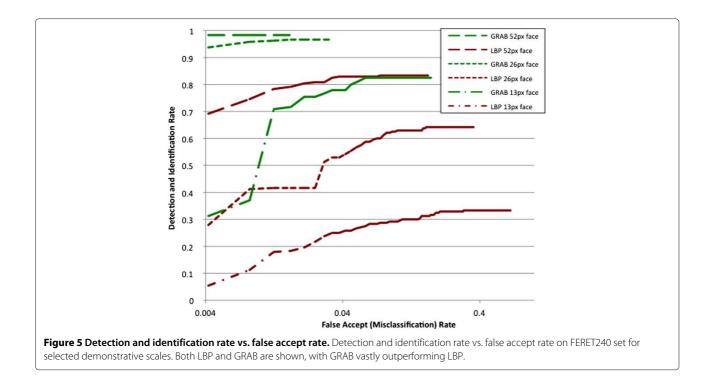
Because this protocol deviates so markedly from the published protocol for FERET and LFW, let us briefly mention the performance of Pinto et.al.'s V1 algorithm [12]. When using that algorithm with the above protocols, including the three image gallery training process, the V1-like algorithm achieves 97.5% accuracy (rank one recognition) on FERET240 and 41.3% on LFW610. The first thing to note is that, as one would expect, LFW is more difficult than FERET. The second and more important aspect of this comparison shows how much more difficult our LFW610 protocol is compared to the basic LFW verification protocol where the V1-like algorithms obtains nearly 80% accuracy following the standard LFW protocols.

To evaluate the impact of scale on the algorithms we generate several instances of reduced spatial resolution images. In order to reduce the variables contributing to recognition score differences, enabling us to better focus on the image degradation due to scale, images were first preprocessed using the standard geometric normalization process provided by the CSU face identification evaluation system [25] using the ground truth eye coordinates available with the databases. This resulted in images of uniform size containing faces oriented approximately the same way. Although the images are preprocessed to have

Table 5 Rank 1 Recognition rate of GRAB, LBP, and V1-like algorithm

Image	GRAB	LBP	Gain	V1-like	Gain
130	55.9	53.28	4.9	41.64	34.24
52	51.15	45.57	12.24	31.97	59.99
39	45.9	40.82	12.44	24.75	85.45
26	36.39	28.20	29.04	10.98	231.4
13	18.2	8.85	105.64	0.49	3610.4

With the percentage improvement of GRAB over LBP and V1-like algorithm; this is on LFW610 dataset with gallery and probe images at different scales. The width of the probe images are in pixels in the table. The gallery image size is 130 × 150. Probe and gallery images have the same aspect ratio.

Figure 5 Detection and identification rate vs. false accept rate. Detection and identification rate vs. false accept rate on FERET240 set for selected demonstrative scales. Both LBP and GRAB are shown, with GRAB vastly outperforming LBP.

the same pixel dimensions (and thus the same digital resolution), those whose original representation had fewer pixels in either dimension will still have reduced optic resolution due to the interpolation necessary to up-sample the image.

For individual experiments, each dataset was divided into its components gallery and probe subsets. Each image in the probe subset was then downsampled to 10%, 20%, 30%, and 40% of its original size (face dimensions of 13×15, 26×30, 39×45, and 52×60 pixels, respectively) thus generating four new sets of probes for our experiments. We computed the four scales, simulating degradation with respect to optic resolution. The image scaling resulted in a decrease in image size (both optic resolution and digital resolution as compared to the original image), which would complicate the data alignment issues. However, the geometric normalization of the pre-processing phase subsequently uses eye location to scale the probes (and the gallery images) to have consistent eye locations and overall face dimensions of a 130-pixel width and a 150-pixel height, regardless of input image size or optical resolution. Since the probe images were considerably smaller than the gallery images, the resulting preprocessed probes have considerably worse optic resolution than the preprocessed gallery images. This procedure was performed for both FERET240 and LFW610.

We conducted experiments using the aforementioned protocols, to compare GRAB and standard LBP on images of various scales. Table 4 summarizes the results obtained with the FERET240 set. We performed similar experiments with the LFW610 dataset, and the results are shown in Table 5. Since FERET is a highly constrained dataset, we get comparatively higher overall performance in FERET240 than in LFW610, which is a highly unconstrained dataset.

It is very clear from the results in Tables 4 and 5 that our proposed GRAB method outperforms LBP in extremely low-scale images, even the ones generated from simple controlled, mostly frontal images. The interesting results are when the images are degraded severely. The performance of LBP is highly impacted by decreases in scale while GRAB is far less susceptible. In addition, an analysis such as that shown in Figure 5 further demonstrates the superiority of GRAB over LBP, especially on more degraded images.

Table 6 Impact of selection of scales on GRAB performance on the FERET240 dataset

Image width	GRAB-best	GRAB-379	GRAB-3579
130	99.17	99.17	99.17
52	98.83	98.83	98.83
39	98.83	98.83	98.83
26	96.67	95.83	95.42
13	83.33	77.9	77.5

Results for GRAB-best are obtained using the ground-truth information, where we know the difference in scales between probe and gallery images and choose the appropriate scale operator. GRAB-3-7-9 is when we predefine the scale of GRAB operator to be 3, 7, and 9 and GRAB-3-5-7-9 is when we combine the scales 3, 5, 7, and 9.

We do a similar analysis for the results on LFW610 dataset as well, where the overall problem is much more difficult because of the greater natural variation in the data. The results show a significant improvement in the performance on a reasonably unconstrained dataset.

For each experiment with a probe image of particular scale, we tried a different combination of GRAB window operator. Tables 4 and 3 show the results for the best scales which were determined empirically. For example, we use the combination of histogram feature vectors obtained using GRAB window operator of sizes 1, 3, and 5. After performing several such experiments, we analyzed the best results we could obtain so far using GRAB, which we call 'GRAB-best' in Table 6. However, using this approach to recognize faces in the real world, where the difference between probe and gallery image scale is not known *a priori*, it would not be feasible and may not be computationally efficient to do so. Hence,for real world recognition scenarios, it is important to apply the scale selection method proposed in this paper.

Conclusions

In this study, we have presented the serious problem in face recognition of size and optic resolution variation due to scale, and we have reviewed various preexisting techniques that have attempted to overcome these obstacles. We have developed the novel GRAB operator and demonstrated its significant performance advantages over LBP in situations of severely decreased scale. While LBP's performance drops off sharply as resolution decreases, the performance of the GRAB operator remains high despite the radical loss of resolution. We also proposed a way to automatically select the GRAB scale based on the image scale and the number of stable pixels across multiple scales. Due to the nature of GRAB as a generalization of LBP, future work will revolve around evaluation of the many other generalizations, and their ability to address additional issues in unconstrained face recognition.

References

1. T Ojala, M Pietikainen, T Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans, Pattern Anal. Machine Intelligence. **24**(7), 971–987 (2002)
2. T Ahonen, A Hadid, M Pietikainen, Face description with local binary patterns: application to face recognition. IEEE Trans. Pattern Anal. Machine Intelligence. **28**(12), 2037–2041 (2006)
3. W Zhang, S Shan, W Gao, X Chen, H Zhang, in *Tenth IEEE International Conference on Computer Vision, 2005. (ICCV 2005)*. Local Gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition. Volume 1, 2005:786–791
4. S Liao, X Zhu, Z Lei, L Zhang, S Li, in *Advances in Biometrics*, ed. by Lee SW, Li S. Learning Multi-scale Block Local Binary Patterns for Face Recognition. Lecture Notes in Computer Science. Volume 4642 (Springer, Heidelberg, 2007), pp. 828–837. [http://dx.doi.org/10.1007/978-3-540-74549-5_87]
5. Z Liu, C Liu, in *Advances in Biometrics*, ed. by Tistarelli M, Nixon M. Robust face recognition using color information. Lecture Notes in Computer Science, Volume 5558 (Springer, Heidelberg, 2009), pp. 122–131. [http://dx.doi.org/10.1007/978-3-642-01793-3_13]
6. M Calonder, V Lepetit, C Strecha, P Fua, in *Computer Vision âĂŞ ECCV 2010*, ed. by Daniilidis K, Maragos P, and Paragios N. BRIEF: binary robust independent elementary features. Lecture Notes in Computer Science. Volume 6314 (Springer, Heidelberg, 2010), pp. 778–792. [http://dx.doi.org/10.1007/978-3-642-15561-1_56]
7. E Rublee, V Rabaud, K Konolige, G Bradski, in *2011 IEEE International Conference on Computer Vision (ICCV)*. ORB: an efficient alternative to SIFT or SURF, (2011), pp. 2564–2571
8. A Hadid, M Pietikainen, T Ahonen, in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004 (CVPR 2004)*. A discriminative feature space for detecting and recognizing faces. Volume 2, (2004), pp. II–797–II–804
9. L Zhang, R Chu, S Xiang, S Liao, S Li, in *Advances in Biometrics*, ed. by Lee SW, Li S. Face detection based on multi-block LBP representation. Lecture Notes in Computer Science. Volume 4642 (Springer, Heidelberg, 2007), pp. 11–18. [http://dx.doi.org/10.1007/978-3-540-74549-5_2]
10. W Zhang, S Shan, H Zhang, W Gao, X Chen, in *Audio- and Video-Based Biometric Person Authentication*, ed. by Kanade T, Jain A, and Ratha N. Multi-resolution histograms of local variation patterns (MHLVP) for robust face recognition. Lecture Notes in Computer Science. Volume 3546 (Springer, Heidelberg, 2005), pp. 937–944. [http://dx.doi.org/10.1007/11527923_98]
11. L Shen, L Bai, A review on Gabor wavelets for face recognition. Pattern Anal. Appl. **9**(2-3), 273–292 (2006). [http://dx.doi.org/10.1007/s10044-006-0033-y]
12. N Pinto, C DiJ,arlo, D Cox, in *IEEE Conference on Computer Vision and Pattern Recognition, 2009 (CVPR 2009)*. How far can you get with a modern face recognition test set using only simple features?, (2009), pp. 2591–2598
13. DG Lowe, Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vision. **60**(2), 91–110 (2004). [http://dx.doi.org/10.1023/B:VISI.0000029664.99615.94]
14. D Lowe, in *The Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999*. Object recognition from local scale-invariant features. Volume 2, (1999), pp. 1150–1157
15. M Bicego, A Lagorio, E Grosso, M Tistarelli, in *Conference on Computer Vision and Pattern Recognition Workshop, 2006 (CVPRW '06)*. On the use of SIFT features for face authentication, (2006), pp. 35–35
16. D Kisku, M Tistarelli, J Sing, P Gupta, in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, 2009 (CVPR Workshops 2009)*. Face recognition by fusion of local and global matching scores using DS theory: an evaluation with uni-classifier and multi-classifier paradigm, (2009), pp. 60–65
17. J Chan, CH Kittler, K Messer, in *Advances in Biometrics*, ed. by Lee SW, Li S. Multi-scale local binary pattern histograms for face recognition. Lecture Notes in Computer Science. Volume 4642 (Springer, Heidelberg, 2007), pp. 809–818. [http://dx.doi.org/10.1007/978-3-540-74549-5_85]
18. S Liao, A Chung, in *Asian Conference on Computer Vision âĂŞ ACCV 2007*, ed. by Yagi Y, Kang S, Kweon I, and Zha H. Face recognition by using elongated local binary patterns with average maximum distance gradient magnitude. Lecture Notes in Computer Science. Volume 4844 (Springer, Heidelberg, 2007), pp. 672–679. [http://dx.doi.org/10.1007/978-3-540-76390-1_66]
19. FC Crow, in *Proceedings of the 11th annual conference on Computer graphics and interactive techniques, SIGGRAPH '84*. Summed-area tables for texture mapping (ACM, New York, 1984), pp. 207–212. [http://doi.acm.org/10.1145/800031.808600]

20. A Sapkota, B Parks, W Scheirer, T Boult, in *the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. FACE-GRAB: face recognition with general region assigned to binary operator, (2010), pp. 82–89

21. G Huang, V Jain, E Learned-Miller, in *the IEEE 11th International Conference on Computer Vision, 2007 (ICCV 2007)*. Unsupervised joint alignment of Complex Images, (2007), pp. 1–8

22. J Zhu, M Vai, P Mak, in *Biometric Authentication*, ed. by D Zhang, A Jain. A new enhanced nearest feature space (ENFS) classifier for Gabor wavelets features-based face recognition. Lecture Notes in Computer Science. Volume 3072 (Springer, Heidelberg, 2004), pp. 124–130. [http://dx.doi.org/10.1007/978-3-540-25948-0_18]

23. P Phillips, H Moon, S Rizvi, P Rauss, The FERET evaluation methodology for face-recognition algorithms. IEEE Trans. Pattern Anal. Machine Intelligence. **22**(10), 1090–1104 (2000)

24. M Huang, GB Ramesh, T Berg, E Learned-Miller, Labeled faces in the wild: a database for studying face recognition in unconstrained environments. Tech. Rep. 07-49. (University of Massachusetts, Amherst, 2007)

25. M Bolme, DS Beveridge, JR Teixeira, BA Draper, in *Proceedings of the 3rd international Conference on Computer vision systems (ICVS'03)*. The CSU face identification evaluation system: its purpose, features, and structure (Springer, Heidelberg, 2003), pp. 304–313. [http://dl.acm.org/citation.cfm?id=1765473.1765507]