FACE-GRAB: Face Recognition with General Region Assigned to Binary Operator

Archana Sapkota[1], Brian Parks[1], Walter Scheirer[1,2], Terrance Boult[1,2]
[1] University of Colorado at Colorado Springs, [2] Securics, Inc.
{asapkota,bparks,walter,tboult}@vast.uccs.edu

Abstract

Face recognition in unconstrained environments is one of the most challenging problems in biometrics. One vexing problem in unconstrained environments is that of scale; a face captured at large distances is considerably harder to recognize than the same face at small distances. Several methods have been proposed to tackle unconstrained face recognition in a robust fashion, including face descriptors such as LBP and its many variations, of which some are less sensitive to scale variations than others. In this paper, we present a novel operator called General Region Assigned to Binary (GRAB), developed as a generalization of LBP. We demonstrate its performance for face recognition in both constrained and unconstrained environments and across multiple scales. Unlike prior work, the GRAB descriptor accounts for multiple scales and resolutions through the size and choice of its neighborhood and is evaluated with respect to varying scales. We show that GRAB significantly outperforms LBP in cases of reduced scale on subsets of two well-known published datasets of FERET and LFW, introducing useful subsets of these datasets for recognition system evaluation.

Figure 1. Example of our problem combining scaling issues with challenges in face recognition in uncontrolled environment, using images from the Labeled Faces in the Wild (LFW) Dataset. From left to right: Original images from LFW dataset, face images of the subject in different sizes. Beyond the challenges of pose and lighting we explore the challenge due to the large variation in the scale. We show down-sampled images of sizes 52x60, 39x45, 26x30 and 13x15 respectively. These, plus the original face region of about 130x150, are the scales we have used in our experiments.

1. Introduction

Face recognition has been one of the most commonly used biometrics for recognition and verification, because of its ease of acquisition, availability and its strong performance in the constrained environment and with frontal faces. However, in the unconstrained environment, face recognition still poses a lot of challenges due to changes in scale, the face's complex 3D structure, and motion, as well as other environmental variables like atmospheric blur and illumination. One of the challenges in face recognition is the extraction of features which are sufficiently discriminative in addition to being invariant to the aforementioned variables. This paper addresses one of these key elements: the impact of varying scales on face recognition.

As a motivating example, consider the use of face in facility protection or surveillance system, such as that described in [7] The authors describe a system for cataloging faces and relating identity and location, which clearly must work over a wide range of scales. 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 feet and one at 40 feet is a 10x change in scale/resolution. In Figure 1, we show two example images and a range of scales, equivalent to a 10x range of scales.

A lot of work has been done in the past in describing meaningful and distinctive features from images for object recognition. Scale Invariant Feature Transform (SIFT) is a popular method in object recognition [13, 12]. 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 [2] whereby 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 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 [9], SIFT features are combined in a mixed local-global strategy supporting a recognition-from-parts approach to address occlusion.

Local Binary Pattern (LBP) is another operator which was originally used to extract a texture based description from imagery and is widely used in face recognition. The operator assigns a label to every pixel of an image by thresholding the 3x3-neighborhood of each pixel with the center pixel value and considering the result as a binary number [1, 21]. 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 micro-structure as well as macro-structure 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 in the image.

Prior extensions to produce "multi-resolution" LBP [14], simply use a larger neighborhood "circle", but still sample the raw pixels on that circle. While it does consider pixels at greater distances, sampling does not properly model changes in resolution or scale, which result in pixels being combined and not sampled. Consider what happens on a region with a fine binary texture, where sampling chooses from one of the two binary colors, but changes in scale actually mix the values into new shades/colors. In [11], this multi-resolution LBP is combined with novel color representations which combine RGB, YCbCr, and YIQ color spaces. The results did improve performance on FRGC data, but that does not not actually contain multiple resolutions so sampling artifacts in color space would impact those experiments.

More recently, studies have introduced the concept of a Multi-scale Block Local Binary Pattern (MB-LBP) to provide a more robust operator than LBP [10]. In MB-LBP, the average sum of image intensity is computed in each sub region around a center sub-region. 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 8 ordinal rectangle features". While MB-LBP does improve recognition by representing a mixture of micro-structure as well as macro-structure of the image pattern, they did not study the impact of scale but rather focused on improving recognition at a fixed scale. To do this, they increased the operator dimensionality for a given resolution and redefined the sampling used, introducing the concept of Statistically Effective MB-LBP (SEMB-LBP) based on the percentage in distributions, instead of the number of 0-1 and 1-0 transitions as in the uniform LBP. They train a mapping from their MB-LBP data to a reduced set of 63 indices on FRGC data, then they use Ada-boosting to learn a 2-class classifier for matching (same identity) vs non-matching faces. They test on FRGC experiment 1 and 2, which contains all high-resolution frontal faces with more than 4x-400x pixels on faces compared to the experiments herein. They outperform a standard LBP operator, but did not perform as well as the best performing algorithm. Multiple scales is not considered in their paper.

LBP features have also been used in the past for face detection. [6] considered LBP features as a facial representation and build a face detection system using SVM as a classifier. This idea was recently extended to Multi-Block Local Binary Pattern features for face detection. [20]. But in real time face recognition systems, a face if detected should be able to be recognized as well and the scale of images plays a huge role here.

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 problem. Some of the previous works 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 (MHLVP) [22] is one such method which described face images as the concatenation of the local spatial histogram of local variations patterns computed from multi-resolution Gabor features.

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. [4] 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 paper we GRAB 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. However, future work may explore their approach for dimensionality reduction and study it with respect to scale changing.

Gabor features are another interesting set of features which are highly applied in face recognition [18, 17]. The gabor representation of face images incorporates multiscale 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. For face recognition using Gabor features, the gabor jets are extracted from a predefined set of points on the face images.

In this paper, we present a new description of facial images, 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 these features for facial recognition in images of varied scales and resolution. As shown in Figure 1, face images captured in the wild have varied scales. It is also interesting to mention that our GRAB-based face recognition performs very well even with the smallest scale image shown in Figure 1.

2. General Region Assigned to Binary (GRAB)

GRAB (Generalized Region Assigned to Binary) 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 actual neighbor pixel. For each pixel c, we can define the generalized binary representation as:

$$GR(c) = \sum_{j=1}^{n} g_j(c, j) \cdot 2^j. \tag{1}$$

where GRAB is the same as the standard LBP *iff*

$$g_j(p_c, p_j) = (p_c > p_j) \tag{2}$$

In the standard LBP pattern, the neighboring pixel is defined a label, 0 or 1 comparing its value with the center pixel value. If the neighboring pixel is greater than the center pixel value then a label 1 is assigned to it. But with the GRAB operator we define a threshold, whereby, if the difference in the neighboring pixel value and the center pixel value is more than the threshold, the neighboring pixel is assigned a label of 1. This gives us a binary representation of the center pixel corresponding to the pattern of comparisons in the neighborhood of the center pixel. However, our proposed GRAB method does not use a single uniform definition as in Local Binary Patterns, but it combines, in a more meaningful way, multiple different neighborhood

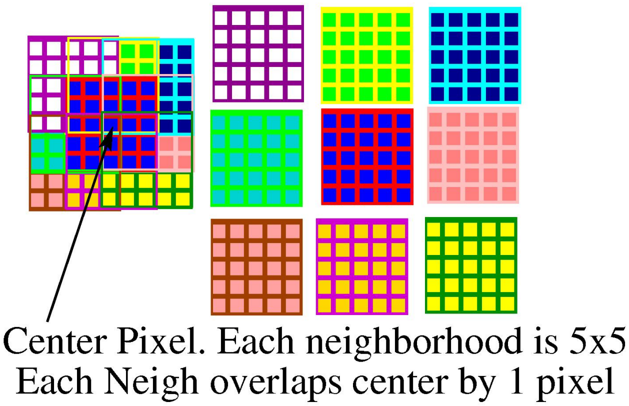

Figure 2. GRAB Representation for a GRAB-5. Each 5x5 region computes the average in that region (average over rectangles shown on right). Note that each region is displaced to just overlap the center pixel . If the center average is significantly different than average for neighbor k, then set bit k to 1, else set to 0. The blurring and displacement of the neighborhoods more accurately models the resolutions/scale chances in an image.

rules. GRAB is a generalization of LBP designed to overcome its limitations on scaling and orientation. One of the generalizations of LBP found in the literature was Elongated Local Binary Patterns with with Average Maximum Distance Gradient Magnitude [19]. In this work, the local features are defined considering the elliptical neighborhood. They also defined a feature called Average Maximum Distance Gradient Magnitude which includes the intensity difference between the center pixels and neighborhood pixels. They have shown successful results on FERET dataset with low resolution images. Our GRAB operator can be used as a generalization of ELBP in the sense the block averages around the center pixel can be arranged in circular or elliptical fashion.

Here are the major modifications we have made in describing the GRAB patterns.

2.1. GRAB as scale invariant operator

The first, and most significant, change is the use of the windowed operators for the neighborhoods. In standard LBP, the comparison is that of a pixel directly with it's neighbors. The prior extensions to produce "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 any shape, use of rectangular regions allows use of summed area tables [5], 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 the figure 2. The regions use N x 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 x 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 allow 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.

2.2. GRAB as noise tolerant operator

Standard LBP uses a simple comparison between the center pixel and its neighborhood, which makes it sensitive to noise. We introduce the idea of threshold in defining the bit pattern around a pixel:

$$g(c, j) = \|S(j) - (S(c)\| < e_N) \tag{3}$$

where e is threshold defined from a statistical analysis of expected level of noise for the sensor data when summed/blurred to level N. The rationale for the added test is that only "statistically significant" differences should be considered to produce a bit in the resulting binary number, or the resulting representation will be significantly impacted by noise. This makes GRAB more stable to minor variations and noise, which is common in intensified imagery, than the basic LBP operators.

2.3. GRAB as variations tolerant and rotation invariant operator

Another change in the definition of GRAB patterns is the labeling of the bits around a pixel. We define NxN block average around a pixel and we should still order the blocks to define the binary representation. The labeling in a standard LBP is spatially shown on the left of Figure 3 with our new labeling on the right. A small variation in the local edge direction in the standard pattern switches from bit 1 to bit 8, i.e. a very large change. Our new pattern mode provides a more refined 8-neighbor labeling, where any two neighboring directions in the image are never more than a factor of 4 away in resulting encoding. This increases stability if there are minor variations in the edge features. We also have proposed a new approach to solving the orientation problem, using the pixel values at a larger scale (larger window operator) to anchor the "orientation" for each pixel. When the local neighborhood rotates, the larger scale's blurring will also rotate, and the GRAB features become relatively orientation independent.

1	2	3
8		4
7	6	5

1	2	4
3		6
5	7	8

Figure 3. Left shows standard neighborhood numbering of pixels for LBP, right shows alternative numbering for GRAB, which ensure that small variation in edge angles cause smaller variations in the binary representation.

3. GRAB-based Recognition

As mentioned in Section 2, GRAB operator assigns a label to every pixel in the image by thresholding the center pixel with the pixel value of NxN block average by 8 neighbors of NxN block average. The pattern thus obtained is considered as a binary number and thus every pixel in the image is assigned such a number. Also, using the neighbor as a NxN block average does not affect the idea of uniform pattern. We can still make use of the uniform pattern which according to [1, 21] 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 (0 transitions), 01110000 (2 transitions) and 11001111 2 transitions) are uniform whereas the patterns 11001001 (4 transitions) and 01010011 6 transitions) are not. We continued to use uniform pattern in our representation because it accounts for a larger percentage of the image representation in FERET dataset [1, 21] 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.

(a) (b) (c)

Figure 4. An image at various stages in the GRAB process. (a) clean image (b) lower optic resolution image (c) image after GRAB processing with 64 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

Table 1. Performance of Nearest-Neighbor classification on FERET240 and LFW610 with weighted regions.

While we could work in the space of smaller images, scaling down the windows, its easier to conceptualize, and implement, when we scale the different resolution image 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 paper were done using ImageMagick's convert function. Figure 4(c) shows an original and the up-sampled low-scale image aligned.

For face description using GRAB features we use the same approach as LBP features because they represent the local as well as global description of the face image. Geometrically normalized images, which are all 130 pixels wide and 150 pixels high, are divided into 64 regions (8 rows and 8 columns) as shown in Figure 4(c). 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 GRAB window operator. For example, for GRAB-3-5-7, the binary pattern was computed taking the block average of 3x3, 5x5, 7x7 neighbors, we 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.

We also verified the performance of LBP on the standard FERET partitions as mentioned in [1] 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 the way the images are normalized.

We chose to use an SVM-based classification method to take advantage of the performance increases it offers over approaches traditionally used with LBP, such as nearest neighbor [1] and because anyone looking to deploy either LBP or GRAB should be using more advanced machine learning. We note the SVM process used improves the performance of both LBP and GRAB, but the choice of machine learning classifier is not the critical aspect of this paper. Refer Tables 1, 2 and 3 to see the performance gain due to SVM over Nearest Neighbor.

While the underlying models for the matching algorithm differ between our implementation and the standard LBP implementations, the processing of the images to generate a representative feature vector (as described in Section 3 for 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 multiclass Support Vector Machine (SVM), while the latter set is subsequently tested against the trained model. In particular we train the SVM with the the LBP or concatenated GRAB histograms as feature vectors, with each subject's gallery image being a positive example for the multiclass SVM, implemented via PyML. We then test with similar feature vectors obtained from the probe images.

4. Evaluation Protocols and Experiments

We test our proposed GRAB operator on subsets of two published data-sets. The FERET (Face Recognition Technology) [16] set was chosen due to extremely common use, allowing readers to do comparisons with many algorithm. It is, however relatively constrained nature in: 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 Labeled Faces in the Wild (LFW) [8]. This set is relatively unconstrained and is generally considered one of the most difficult published set for facial analysis.

To use our SVM-based classification method and address some of the pose issues we require a gallery of more than a single image. 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 data sets: subjects for whom the data set 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 data set. 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 particular no relation to standard stets 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. In comparison, LFW's original protocol is written from the perspective of a verification problem, which not could be be directly modified into the context of a recognition problem.

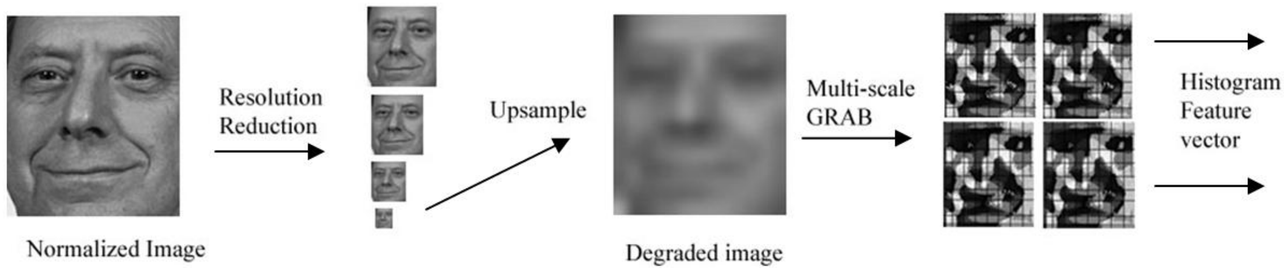

Figure 5. Extraction of GRAB Face Descriptors. From left to right, original image of a subject geo-normalized to the size of 130x150, images of the same subject at sizes 52x60, 39x45, 26x30,13x15, image of size 13x15 up-sampled to 130x150. The right shows a group of GRAB operators applied on the image using GRAB window operator of size 3, 5, 7, and 9. The histogram features are extracted from the 64 different regions of the GRABBed images and are concatenated to form a global histogram descriptor. Global descriptors from multiple scale GRABBed images are extracted and concatenated to form a multi-scale histogram feature vector.

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 ([17]). When using that algorithm with the above protocols, including the 3 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 it 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 with the process described in in Figure 5. In order to reduce the variables contributing to recognition score differences, so as to 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 [3]. This resulted in images of a uniform size containing faces oriented approximately the same way. Although the images are preprocessed to have 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 component "gallery" and "probe" subsets. Each image in the probe subset was then down-sampled to 10%, 20%, 30%, and 40% of its original size (face dimensions of 13x15 pixels, 26x30 pixels, 39x45 pixels, and 52x60 pixels, respectively), thus generating four new sets of probes for our experiments. As shown in Figure 5 we compute 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 preprocessing phase subsequently uses eye location to scales the probes (and the gallery images) to have consistent eye locations and overall face dimensions of 130 pixels width and 150 pixels 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.

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

Table 2. Rank 1 Recognition Rate of GRAB , LBP and V1-like algorithm with the percentage improvement of GRAB over LBP and V1-Like. This is on FERET240 dataset with Gallery and Probe images at difference scales. The width of the probe images are in pixels in the table. The gallery image size is 130x150. Probe and gallery images have the same aspect ratio.

5. Results

We ran experiments using the aforementioned protocols, to compare GRAB and standard LBP on images of various scales. Table 2 summarizes the results obtained in FERET240 set. We performed similar experiments with the LFW610 dataset and the results are shown in Table 3. 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 2 and 3 that our proposed GRAB method outperforms LBP in extremely low scale images even of 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. The percentage improvement of GRAB over LBP on less degraded images is consistent, but not very large. For

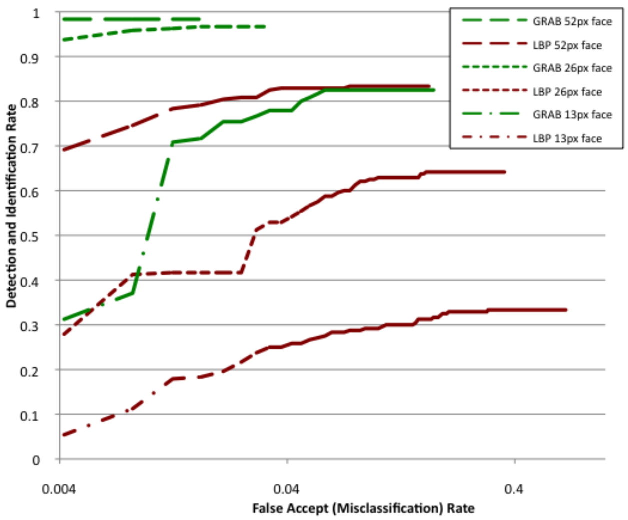

Figure 6. Detection and Identification Rate vs. False Accept Rate on FERET240 set for selected demonstrative scales. Both LBP and GRAB are shown, with GRAB vastly outperforming LBP.

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

Table 3. Rank 1 Recognition Rate of GRAB, LBP and V1-Like algorithm with the percentage improvement of GRAB over LBP and V1-Like. This is on LFW610 dataset with Gallery and Probe images at difference scales. The width of the probe images are in pixels in the table. The gallery image size is 130x150. Probe and gallery images have the same aspect ratio.

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

Table 4. Impact of defined combination of scales on GRAB performance on the FERET240 dataset. Results for GRAB-Best are obtained by using the ground-truth information, where we know the difference in scales in 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 all the scales but 1.

the images of size 52x60, the improvement is 4%, and for 39x45 images it is 10%. As the degradation increases, the percentage improvement increases. For the images which are scaled down by 80% with image size of 26x30, we get the improvement of around 29%, while for 90% degraded images and size reduced to 13x15 we get 80% improvement over LBP. This clearly shows that GRAB is tolerant with respect to changes in resolution while LBP still suffers when there is a significant change in scale.

In addition, an analysis such as that shown in Figure 6 further demonstrates the superiority of GRAB over LBP, especially on more degraded images. The technique for this analysis is based on a method for visualizing performance on open recognition problems presented in [15] that generalizes a ROC curve. It may seem unintuitive that the curves do not reach the full rank-1 recognition rate shown in 2; however, our data exhibited the property that the worst score over all false accepts was still better than several true accepts. Thus, the continuation of the line (if any) could not be accurately extrapolated from our data.

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 overall recognition on LFW610 is actually considerably high, considering the difficulty of the dataset, with 55.9% recognition on clean images, outperforming the V1-like algorithm of [17] which only achieves 41% rank-1 recognition when applied to this dataset in a recognition scenario. LBP achieves 53.28% on the geometrically normalized, unscaled LFW610 dataset. Looking across scales, the percentage improvement of GRAB over LBP is 12.2%, 12.4%, 28.8%, 105.6% respectively for probe image sizes of 52x60, 39x45, 26x30 and 13x14 . This is a significant improvement in the performance on a reasonably unconstrained dataset.

For each experiment with a probe image of particular scale, we tried different combination of GRAB window operator. For example, we use the combination of histogram feature vectors obtained by using GRAB window operator of size 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 4. 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. We analyzed the results to determine if we can simply use the combination of multiple scale GRAB features and and still obtain comparable results. We observed that combination of multiple scales sometimes decreases results for a wide scale variance but then it is more robust to combine multiple scales than to use a single scale like LBP. We also observed that the combination of larger GRAB window operators works better for a large decrease in scale while the combination of smaller GRAB window operators works better for less scale variance in probe and gallery images, but that the difference is not that much. For example, we considered the combination of 3 different GRAB window operator 3, 7, and 9 to see the performance over all the levels of resolution reduction. We call it GRAB-379 because it combines GRAB window operator of size 3, 7, and 9 and the results are summarized in Table 4. Rank-1 Recognition results for the GRAB-Best, GRAB-379, and GRAB-3579 are not very different, while they still perform significantly better than LBP.

6. Conclusion

In this study, we have presented the serious problem in face recognition of size and optic resolution variation due to scale and reviewed various pre-existing techniques that have attempted to overcome these obstacles. Based on the best performing of these methods, 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. Due to the nature of GRAB as a generalization of LBP, future work will revolve around evaluation of the many other generalizations defined by GRAB, and their ability to address additional issues in unconstrained face recognition.

7. Acknowledgments

This work was supported by ON STTR (Award Number N000014-07-M-0421) and ONR MURI (Award Number N00014-08-1-0638). We would also like to thank Chris Eberle, Charles Beckmann, and R. C. Johnson for their contributions to the software during the initial phase of implementation.

References

[1] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE TPAMI*, 28(12):2037–2041, 2006. 2, 4, 5

[2] M. Bicego, A. Lagorio, E. Grosso, and M. Tistarelli. On the use of sift features for face authentication. In *CVPRW '06: Proc. 2006 CVPR Workshops*, page 35, Washington, DC, USA, 2006. IEEE Computer Society. 2

[3] D. Bolme, R. Beveridge, M. Teixeira, and B. Draper. The CSU Face Identification Evaluation System: Its purpose, features and structure. In *ICVS*, pages 304–311, 2003. 6

[4] C.-H. Chan, J. Kittler, and K. Messer. Multi-scale local binary pattern histograms for face recognition. In S.-W. Lee and S. Z. Li, editors, *ICB '09: Proc. of the Int. Conference on Advances in Biometrics*, volume 4642 of *Lecture Notes in Computer Science*, pages 809–818. Springer, 2007. 2

[5] F. C. Crow. Summed-area tables for texture mapping. In *SIGGRAPH '84: Proceedings of the 11th annual conference on Computer graphics and interactive techniques*, pages 207–212, New York, NY, USA, 1984. ACM. 3

[6] A. Hadid, M. Pietikäinen, and T. Ahonen. A discriminative feature space for detecting and recognizing faces. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 2:797–804, 2004. 2

[7] A. Hampapur, S. Pankanti, A. Senior, Y.-L. Tian, L. Brown, and R. Bolle. Face cataloger: Multi-scale imaging for relating identity to location. In *IEEE Conference on Advanced Video and Signal Based Surveillance,*, pages 13–19. IEEE Computer Society, 2003. 1

[8] G. 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, 2007. 5

[9] D. Kisku, M. Tistarelli, J. Sing, and P. Gupta. Face recognition by fusion of local and global matching scores using ds theory: An evaluation with uni-classifier and multi-classifier paradigm. In *CVPRW '09: Proc. 2009 CVPR Workshop on Biometrics*, pages 60–65. IEEE Computer Society, 2009. 2

[10] S. Liao, X. Zhu, Z. Lei, L. Zhang, and S. Z. Li. Learning multi-scale block local binary patterns for face recognition. In *Int. Conf. on Biometrics*, pages 828–837, 2007. 2

[11] Z. Liu and C. Liu. Robust face recognition using color information. In *ICB '09: Proc. of the Int. Conference on Advances in Biometrics*, pages 122–131. Springer-Verlag, 2009. 2

[12] D. G. Lowe. Object recognition from local scale-invariant features. *IEEE International Conference on Computer Vision*, 2:1150, 1999. 1

[13] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:91–110, 2004. 1

[14] T. Ojala, M. Pietikäinen, and T. Mäenpää. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(7):971–987, 2002. 2

[15] P. J. Phillips, P. Grother, and R. Michaels. Evaluation methods in face recognition. In *Handbook of Face Recognition*, chapter 14, pages 329–348. Springer New York, 2005. 7

[16] P. J. Phillips, H. Moon, P. J. Rauss, and S. Rizvi. The FERET evaluation methodology for face recognition algorithms. *IEEE TPAMI*, 22(10), 2000. 5

[17] N. Pinto, J. J. DiCarlo, and D. D. Cox. How far can you get with a modern face recognition test set using only simple features? In *IEEE CVPR*, 2009. 3, 6, 7

[18] L. Shen and L. Bai. A review on gabor wavelets for face recognition. *Pattern Anal. Appl.*, 9(2):273–292, 2006. 3

[19] A. C. S. C. Shu Liao. Face recognition by using elongated local binary patterns with average maximum distance gradient magnitude. *Asian Computer Vision Conf. ACCV 2007*, pages 672–679, 2007. 3

[20] L. Zhang, R. Chu, S. Xiang, S. Liao, and S. Z. Li. Face detection based on multi-block lbp representation. *Advances in Biometrics*, pages 11–18, 2009. 2

[21] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang. Local gabor binary pattern histogram sequence (lgbphs): a novel non-statistical model for face representation and recognition. In *IEEE ICCV*, volume 1, pages 786–791, oct. 2005. 2, 4

[22] W. Zhang, S. Shan, H. Zhang, W. Gao, and X. Chen. Multiresolution histograms of local variation patterns (mhlvp) for robust face recognition. *Pattern Analysis & Applications*, pages 937–944, 2005. 2