Displacement Template with Divide-&-Conquer Algorithm for Significantly Improving Descriptor Based Face Recognition Approaches

Liang Chen[1,2], Ling Yan[2], Yonghuai Liu[3], Lixin Gao[1], and Xiaoqin Zhang[1]

[1] Wenzhou University, China
[2] University of Northern British Columbia, Canada
[3] Aberystwyth University, UK

Abstract. This paper proposes a displacement template structure for improving descriptor based face recognition approaches. With this template structure, a face is represented by a template consisting of a set of piled blocks; each block pile consists of a few heavily overlapped blocks from the face image. An ensemble of blocks, one from each pile, is taken as a candidate image of the face. When a descriptor based approach is used, we are able to generate a displacement description template for the face by replacing each block in the template with its local description, where a concatenation of the local descriptions of the blocks, one from each pile, is taken to be a candidate description of the face. Using the description template together with a divide-and-conquer algorithm for computing the similarities between description templates, we have demonstrated the significantly improved performance of LBP, TPLBP and FPLBP templates over original LBP, TPLBP and FPLBP approaches by the experiments on benchmark face databases.

Keywords: Descriptor Approach, Template, Face Recognition.

1 Introduction

Face recognition is a specific pattern recognition problem, for which numerous approaches have been developed and published, although we have to confess that we understand very little of the actual processes our brains use in performing such a task. Among all the approaches, (local feature) descriptor based approaches [1] are recently very popular, with local binary pattern (LBP) based approach as a typical instance[2]. The LBP approach has established a unique position in face analysis research, and numerous variations have been developed for a variety of tasks.

A local feature descriptor based approach uses local texture descriptors to build local descriptions of a face and then concatenate them into a global description. As a typical instance, the LBP based approach, or its variants, works as follows: face images are first partitioned into several windows on which local binary pattern histograms are generated, then these local histograms of local windows are concatenated into a single feature vector as the face representation.

A. Fitzgibbon et al. (Eds.): ECCV 2012, Part V, LNCS 7576, pp. 214–227, 2012.

The advantage, which is proved by experiments, is that such a representation codifies both texture and spatial information. In this representation, three different levels of patterns are involved: the pixel-level patterns of LBP labels, the histogram of the LBP labels over each window, and the global representation concatenated from the window level histograms. In this paper, for the convenience of discussion, the histogram of a local window is called window level local description; the global representation is called the (global) description of the face image.

The alignment of two face images to be matched is an important issue to the success of descriptor based approaches, just as most of the other face recognition approaches.[1] It should be noted that "bad" alignments can seriously lower the performance of face recognition algorithms. An efficient way to generate a relatively precise alignment becomes a key issue when developing a face recognition approach while it is known that an alignment of faces cannot be precise, manually or automatically. However, since human faces are nonrigid objects, a "perfect" alignment is almost impossible when face images to be compared are not exactly identical. We can easily notice that the alignment difficulties also exist in general image retrieval and classification problems, where an effective compensating strategy is to allow pixel level or region level displacement when calculating the similarities between location relevant features of different images [3,4,5]. Region level displacement has also been used in face recognition [6,7].

Enlightened by the success of a language translation tutoring system [8,9], perhaps the only translation tutoring online system presently available that allows free inputs, where templates are used to describe translations of specific sentences and each path represents one possible correct or model translation, we propose in this paper a local displacement template structure for human face image representation for the application of face recognition.

In the content of language tutoring system, a template is a loop-free graph where each arc is labeled by a word or a word phrase, and a concatenation of the words/word phases along a path from the start point(s) to the end point(s) suggests a possible translation. Within the displacement template concept for face recognition, a face is represented by a template consisting of a set of piled blocks, and each block pile consists of a few (heavily) overlapped blocks from a face image. An arbitrary ensemble of blocks, one from each pile, is taken as a candidate image of the face.

In consideration of an descriptor approach, with a displacement face template, by replacing every block in every pile with its block level description, we are able to generate a displacement description template for the face image. A concatenation of block level descriptions, one from each pile, is now taken as a candidate description of the face. [2] Noting that, a description of an image is a concatenation of window level descriptions; when there is no window across

[1] We follow the locating-alignment-recognition pipeline for face recognition, where face recognition normally refers to the last stage, but not locating or alignment.

[2] It is almost equivalent to the description of corresponding candidate image of the face, excluding the pixel level descriptions of pixels near the boundaries of the blocks.

a boundary of a block, a global description of an image can equivalently be constructed in two steps: first concatenate window level descriptions into block level descriptions and then concatenate block level descriptions into the global level description. The similarity between two images is now measured by the similarity of their description templates which is defined as the largest similarity among the similarities between all pairs of candidate descriptions of the two templates. We must note that, the window level descriptions are simple statistics of pixel level patterns in a window and therefore do not code any spatial information; only block level and global descriptions contain spatial information. We can see that, using the multiple candidate descriptions for a single face image is likely to reduce the divergence caused by registration difficulties of human faces.

There is a time cost for adopting the displacement template structures. With a description template consisting of m block piles, each contains k blocks, there are k^m candidate descriptions for a face image. Assuming the time cost for computing the similarity between the standard descriptions (without using template structures) of two face images is $O(T)$, the total time complexity computing the similarity between the description templates of two face images will be $O(k^{2m}T)$ if we compare all pairs of candidate descriptions of these two face images by enumeration. We prove that, a divide-and-conquer approach can be applied to significantly reduce the time complexity to $O(mk^2T)$. In practice, as shown in Section 2.2, we can use a regular description, i.e., a single level description template for a probe face image, to further reduce the time complexity to $O(mkT)$.

Before going to the details, we note here: The motivation of this paper, as indicated by its title, is to show that displacement template structure can improve descriptor approaches. Besides descriptor approaches, there are many non-descriptor approaches, and many mixed approaches which are the combinations of descriptor approaches and some machine learning techniques. Since we are not to claim that our approach is the best of all algorithms, but only to show that our displacement template can improve descriptor approaches; it would be a little bit out of focus to compare our descriptor based template with non-descriptor or mixed approaches. However, to further demonstrate the applicability of our approach, we will also compare our results with a few recent mixed approaches with similar settings.

The remaining of this paper is organized as follows: The displacement template structure and displacement description template are described in Section 2. The algorithm for template matching is also shown in Section 2. The experiments verifying the advantages of our approach are shown in Section 3. Conclusions and future research are discussed in Section 4.

2 Displacement Template Structure, Description Template and Algorithm

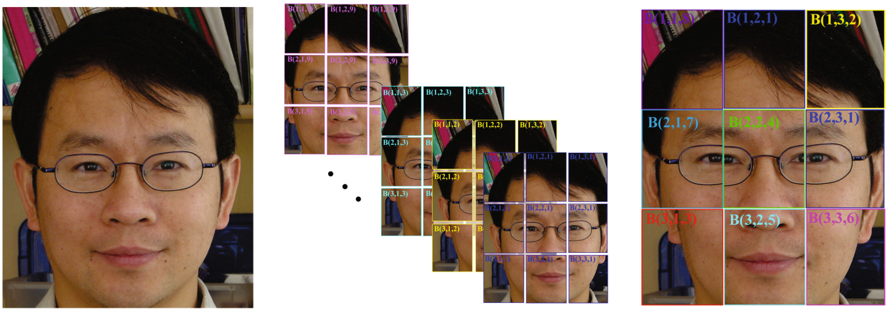
2.1 Image Template and Description Template

Definition 1. *Assuming we have a face image X of width $w + 2s$ and height $h + 2s$, we generate a* **template structure** *for X as follows:*

1. We first obtain $(2s+1)^2$ face images of size $w \times h$ by removing $s-i$ pixels from the top, $s+i$ pixels from the bottom, $s-j$ pixels from the left, and $s+j$ pixels from the right, where $-s \le i, j \le s$ [3]*. Such an image of size $w \times h$ (denoted by $I_{i,j}$) is called a layer in a pile;*
2. We then partition each image $I_{i,j}$ into $l \times c$ blocks and pile up the corresponding blocks to develop the template structure.

As an example, for the face image in Figure 1(a), we construct its template structure in Figure 1(b), where $s = 1$, $l = c = 3$.

As a further exploitation of the concept behind this alignment/displacement, note that step lengths in i and j do not necessarily have to be 1 pixel, neither do the numbers of pixels removed from the top-bottom side and the left-right side have to be equal. Blocks in each image do not have to be rectangular and do not have to be of the same size either, but the corresponding blocks of the same pile must be of the same shape and size. Within a template structure, a face is represented by a template consisting of a set of piled blocks, and each pile consists of a few (heavily) overlapped blocks from a face image.



(a) Face Image (b) Template Structure (c) A Candidate Image

Note: (1) Windows inside blocks are not shown in this figure. (2) $B(i, j, k)$ represents the block from pile (i,j) in the kth layer.

Fig. 1. A Face and Its Template Representation

Definition 2. *An ensemble of blocks in the displacement template of a face, one from each pile, is taken as a* **candidate image** *of the face.*

For the face image template shown in Figure 1(b), we can generate 9^9 candidate images out of it. The image in Figure 1(c) is one of the candidate images. Note that this candidate image is an ensemble of blocks from different layers.

It is important to note that, the concept of blocks is different from the concept of windows in a descriptor based approach. Blocks are units to construct

[3] In some cases, to reduce the size of each pile, we may also restrict $|i| + |j| \le t$, where t is an integer less than $2s$.

an image; windows are units of an image where local histograms of pixel level patterns are generated to be concatenated into the description of the image. Therefor, theoretically, a block can have many windows, and a window may cross the boundaries of blocks. To simplify the computation, this paper requires that no window cross the boundary of a block.

Our description template is a structure to be used to improve the performances of descriptor based approaches. For a descriptor based approach, as we know, the description of an image is a concatenation of the histograms of pixel level patterns of all windows. The histogram of pixel level patterns of a window is here referred to as the window level description. We now introduce the concept of a block level description as follows.

Definition 3. *The concatenation of window level descriptions of all windows within a block is called a block level description.*

It is not difficult to see that the description of an image is the concatenation of local descriptions of all the blocks.

Definition 4. *By replacing each block in each pile in the template structure of a face by its (block level) description, we are able to generate a* **(displacement) description template** *for the face.*

Definition 5. *An ensemble of blocks in the displacement description template of a face, one from each pile, is taken as a* **candidate description** *of the face.*

When a description template of a face consists of m block piles each of which contains k blocks, there are k^m candidate descriptions for the face.

2.2 Similarity between Description Templates

Definition 6. *The similarity between two images is measured by the similarity of two description templates, which is defined as the largest similarity among the similarities between all pairs of candidate descriptions, one from each template.*

The traditional approaches to compute similarity between a gallery image and a probe image in a descriptor based representations $L = \{L_1, L_2, \cdots\}$ and $M = \{M_1, M_2, \cdots\}$ are Euclidean Distance [4], Histogram Intersection [2], Log-likelihood statistic and Chi square statistic:

$$\text{Euclidean Distance:} \qquad E(L, M) = -\sum_i (L_i - M_i)^2. \tag{1}$$

$$\text{Histogram intersection:} \qquad H(L, M) = \sum_i \min(L_i, M_i). \tag{2}$$

$$\text{Log-likelihood statistic:} \qquad L(L, M) = \sum_i L_i \log M_i. \tag{3}$$

$$\text{Chi square statistic:} \qquad \chi^2(L, M) = -\sum_i \frac{(L_i - M_i)^2}{L_i + M_i}. \tag{4}$$

[4] We will use a squared version of Euclidean Distance for the simplicity in calculation.

Assuming the description of a block (in a block pile) $B(i, j, k)$ ($B(i, j, k)$ represents the block from pile (i, j) in the kth layer) is $DB(i, j, k)$, then the representation of an image consisting of blocks $B(1, 1, k_{1,1}), B(1, 2, k_{1,2}), \cdots$, should be $DB(1, 1, k_{1,1}), DB(1, 2, k_{1,2}), \cdots$.

It seems that all combinations of blocks with one from each pile should be enumerated in order to find the largest similarity between two templates. However, it is easy to note that, when using any of the above similarity measures, the similarity between description templates DB_1 and DB_2 can be formulated as:

$$\begin{aligned}\mathrm{Sim}(DB_1, DB_2) &= \max_{1 \le k_{1_{i,j}}, k_{2_{i,j}} <= p} \sum_{i,j} \mathrm{Sim}(DB_1(i, j, k_{1_{i,j}}), DB_2(i, j, k_{2_{i,j}})) \\ &= \sum_{i,j} \max_{1 \le k_{1_{i,j}}, k_{2_{i,j}} <= p} \mathrm{Sim}(DB_1(i, j, k_{1_{i,j}}), DB_2(i, j, k_{2_{i,j}})), \quad (5)\end{aligned}$$

where p represents the number of blocks in each pile.

Equation 5 indicates that, the similarity between the two description templates can be obtained by accumulating the locally best similarities between corresponding block piles. Therefore, not all candidate descriptions have to be tested in order to find the largest similarity between two description templates; the total comparisons can be reduced from $(2s+1)^{2lc}$ to $lc(2s+1)^2$ by adopting a divide and conquer strategy.

As was pointed out in [2], the log-likelihood measure is not appealing to face recognition. Therefore, we shall not use it as a similarity measure in this paper. It is clear that histogram intersection and Chi measures work better than most premier Euclidean distance measures; but we still include Euclidean distance as one of the similarity measures in order to demonstrate a bottom level performance of our approach when an improper similarity measure is used, and to show that even when using a "not-good" metric, our displacement template approach can still perform comparably to many complicated state-of-art approaches.

To further reduce computing complexity, for an input probe face X, we use only the regular description, that is, a single level description template: we remove s elements from each side, and partition it into $l \times c$ blocks, with each block forming one pile in the template. With this single level template, we can obtain the global description by concatenating the descriptions of all blocks.

2.3 Divide-and-Conquer Algorithm

Assuming we have a gallery of face images, we shall first represent each face image using a template structure: remove a total of $2s$ pixels from topmost and bottommost sides and $2s$ pixels from leftmost and rightmost sides to obtain $(2s+1)^2$ images; partition each image into $l \times c$ equal sized blocks [5]; generate the description for each block; and obtain the template description by piling up the descriptions of corresponding blocks.

[5] Almost equal, indeed, due to the round-off operations.

For an input probe image, we construct its single level template description by removing s pixels from each side and then partitioning it into $l \times c$ equal sized blocks, each constructing one block pile.

We then calculate the similarity between an input image description template and the gallery image description template by *locally* finding the description of a block from each pile in the gallery template that maximizes its local similarity value with the corresponding block of the input probe face.

The complete detailed divide-and-conquer algorithm, which is now straightforward from above explanations, is included in the supplementary package.

3 Experiments

We embed the original LBP approach of [10], TPLBP and FPLBP approaches of [11] into our description template structure. These description templates are called LBP description template, TPLBP description template and FPLBP description template respectively.

Besides the window sizes which we will discuss later, when LBP, TPLBP and FPLBP are used, a few parameters are involved:
1. For LBP: the circle of radius R, the number of sampling points P distributed on a circle of radius R, and whether non-uniform patterns are labeled differently.
2. For TPLBP: patch size $w \times w$, ring radius r of circles, the number S of the additional patches distributed in the ring, and the number α of patches that two patches should be apart from each other along the circle during code generation.
3. for FPLBP: patch size $w \times w$, radii r_1 and r_2 of two rings, the number S of the patches spreading out on each ring, and the number α of patches that two patches should be apart from each other along the outer ring circle during code generation.

In our experiments, for LBP approach, we choose radius to be 1 or 2, and number of sampling points to be 8 as suggested in [10]. We use only uniform patterns and label all remaining patterns with a single label as in [10]. For TPLBP, we choose $w = 3$, $r = 2$, $S = 8$ and $\alpha = 5$; for FPLBP, $r1 = 4$, $r2 = 5$, $S = 8$, $w = 3$ and $\alpha = 1$. These parameters are suggested values from their original paper as well as the software package the authors provided [11].

Although windows in LBP and T/FPLBP approaches are not restricted to be rectangular in shape and can overlap, we use non-overlapping rectangular windows as in [10]. The best results using LBP have been reported in paper [10]. TPLBP and FPLBP achieve their best results when the window numbers are in the range of 20×20 and 35×35. We report here their results with 25×25 windows. It was noted in [2,11] that the LBP and T/FPLBP representations were robust against the selection of parameters.

For our description template, to ensure that a window is always entirely within one block, we define that the numbers of windows in each row and each column of each block are whole numbers. (Indeed, in our experiment programs, as shown in the supplementary package, we first partition the pattern label map of a face image into windows, then ensemble neighboring windows into blocks.)

In our experiments on FERET[12] and FRGC [13], we divide images into 5×5 blocks, each with 7×7 windows. For the experiments on LFW [14], since the image sizes are significantly smaller, we divide images into 4×4 blocks, each with 5×5 windows.

For our approach, there is another parameter to set: s as in Definition 1. We here use $s = 3$. That is, in the first step of generating the image template (see subsection 2.1), we allow the removal of $3 - i$ pixels from the top, $3 + i$ pixels from the bottom, $3 - j$ pixels from the left, $3 + j$ pixels from the right, where $-3 \leq i, j \leq 3$. This will generate a total number of 49 blocks in a pile. To further reduce the pile size, and based on that we believe there is a high probability that the relative shift between the gallery image and the probe image is small, we restrict $|i| + |j| \leq 4$.

3.1 Experiments on FERET

We carry our experiments on FETET database[12]. The FERET database consists of 14051 gray-scale images from 1199 individuals in total. The images vary in lighting conditions, facial expressions, poses, etc. Following the work in [10], we here list the five sets for experiments: *Fa* gallery image set that contains images of 1196 subjects, one image for each subject; *Fb* that contains 1195 face images of the subjects in *Fa* but with alternative facial expressions; *Fc* probe set that contains 194 face images taken under different illumination conditions on the same day as the gallery image was taken; *Dup1* set that contains 722 face images taken anywhere between one minute and 1031 days after the gallery image was taken; *Dup2* set being a subset of *dup1* that contains 234 face images taken at least 18 months after the gallery image was taken. All faces are first normalized into a standard size 150×130 pixels (150 pixels per column, 130 pixels per row), where the distance between the centers of the two eyes is 56 pixels and the line between two eyes lies on the 53rd pixel below the top boundary. The grey histogram of each image is equalized before constructing descriptions or template descriptions.

We take *Fa* as the gallery set, and *Fb*, *Fc*, *Dup1* and *Dup2* as the probe sets. The standard 150×130 elliptical mask from FERET data collection is used to exclude non-face areas from the images. In our experiments, we reduce the elliptical mask size by removing 3 pixels from each side since the images to be partitioned into blocks in a template structure are indeed of size 144×124 (See the first step of generating image templates in subsection 2.1) .

Following [2], permutation test [15] with 95% confidence level is also carried out using the image list, list640.srt, in the CSU face identification evaluation system package [15]. list640.srt contains 4 images each for 160 subjects in FERET. 10000 permutations are tested, with each containing one image per subject in the gallery set and another in the probe set.

We summarize in Table 1 the results of our LBP, T/FPLBP description template approaches with the parameters listed at the beginning of this section. The best results of standard LBP description approach, the weighted LBP approach reported in [10], the T/FPLBP approaches, and our LBP and T/FPLBP

Table 1. The recognition rates of original LBP and weighted LBP, the LBP Template, original T/FPLBP, and T/FPLBP Template for the FERET probe sets, the mean recognition rate of the Fb+Fc+Dup1, and results of permutation test with a 95% confidence level

Method		Fb	Fc	Dup1	Dup2	Fb,Fc & Dup1	Permutation Test lower	mean	upper
LBP, no weight [10]		93%	51%	61%	50%	78.20%	71%	76%	81%
LBP, weighted [10]		97%	79%	66%	64%	84.74%	76%	81%	85%
LBP Template	Euclidean Distance	98.49%	90.21%	70.50%	61.11%	88.16%	78.13%	83.26%	88.13%
	Histogram intersection	98.91%	92.78%	76.04%	68.38%	90.53%	83.13%	87.88%	92.50%
	Chi square statistic	98.74%	91.24%	75.62%	65.81%	90.15%	83.13%	87.61%	91.88%
TPLBP	Euclidean Distance	94.64%	74.23%	62.33%	55.98%	81.71%	68.13%	74.12%	80.00%
	Histogram intersection	96.44%	86.08%	74.65%	69.23%	88.04%	80.00%	85.06%	90.00%
	Chi square statistic	95.98%	86.08%	74.79%	69.66%	87.83%	79.38%	84.50%	89.38%
TPLBP Template	Euclidean Distance	98.16%	86.60%	74.79%	72.22%	89.10%	78.75%	83.71%	88.75%
	Histogram intersection	98.16%	94.33%	81.86%	80.34%	92.23%	85.00%	89.56%	93.75%
	Chi square statistic	97.99%	94.33%	81.30%	80.34%	91.95%	84.38%	89.31%	93.75%
FPLBP	Euclidean Distance	95.73%	69.59%	64.13%	54.70%	82.52%	72.50%	78.07%	83.13%
	Histogram intersection	96.65%	74.23%	67.45%	56.84%	84.60%	75.94%	81.19%	86.25%
	Chi square statistic	96.65%	74.23%	67.73%	56.41%	84.70%	75.63%	81.16%	86.25%
FPLBP Template	Euclidean Distance	98.33%	75.77%	69.60%	61.97%	86.43%	78.75%	83.60%	88.75%
	Histogram intersection	98.58%	81.44%	72.71%	65.38%	88.16%	80.63%	85.60%	90.63%
	Chi square statistic	98.33%	79.90%	72.99%	64.96%	87.97%	80.63%	85.62%	90.63%

description template approaches are all shown in Table 1. It is clear that, our LBP displacement description template approach and T/FPLBP description template approaches achieve much better performances than the original LBP approach and original T/FPLBP description approaches respectively. These description template approaches also perform better than the weighted LBP approach.

As we mentioned at the beginning of this paper, we are only interested to see how our template structure can be used to improve descriptor approaches and we are not trying to claim that our approach itself will reach the best performance of all known approaches. As it is easy to see that most approaches in face recognitions with extremely high performances are "mixed" approaches, we can see that some machine learning and image processing techniques can be "added" onto our approaches to boost the performances in applications.

We understand that, for many practitioners, it is important for them to compare the results directly with "best" results so as to see if the contribution of this paper is significant enough in comparison to so called state-of-the-art results [6]. In order to do so, we here first add a "preprocessing" stage, including gamma correction, DoG filtering and Contrast Equalization, as suggested by Tan et al [16] to boost the performance of our approach. The face images are first preprocessed, using the same parameters as [16] suggested, before applying our descriptor template approaches. We also adopt the"weighting" strategy in the original LBP paper [2] to assign weights to blocks further improve the performance of our description template approach: The subfc [2], which contains 97 images in *Fa* as gallery and 97 images in *Fc* as probes, was used for training

[6] How many excellent new ideas have been killed by this simplified pragmatism during peer reviewing process?

Table 2. The recognition rates of the LBP Template approach boosted by preprocessing and weighting schemes on the FERET probe sets, and a few known approaches

Method		Fb	Fc	Dup1	Dup2
LBP Template with Preprocessing	Euclidean Distance	98.49%	98.45%	84.07%	82.05%
	Histogram intersection	99.00%	98.97%	88.23%	86.75%
	Chi square statistic	99.00%	98.45%	88.23%	86.75%
LBP Template with Preprocessing & Weighting	Euclidean Distance	98.91%	100.00%	84.90%	85.04%
	Histogram intersection	99.16%	100.00%	89.89%	87.61%
	Chi square statistic	99.16%	100.00%	90.03%	88.46%
LGBPHS[17]		98.0%	97.0%	74.0 %	71.0%
HGPP[5]		97.6%	98.9%	77.7 %	76.1%
SIS [18]		91.0%	90.0%	68.0 %	68.0%
Schwartz [19]		95.7%	99.0%	80.3 %	80.3%

purpose; experiment on only one block of a training set was carried each time; the recognition rates of corresponding blocks on left and right sides of the face were averaged and then sorted in nondecreasing order; the blocks with recognition rates above the top 10th percentile and above the top 20th percentile, were assigned with weights 4 and 2 respectively. For those below bottom 20th percentile, we assign weights 0, the remaining blocks took weight 1. We show the results of our LBP template approach boosted by preprocessing only and by both preprocessing and weighting schemes[7] in Table 2, together with a few known approaches which do not apply a training process.

3.2 Experiments on LFW

"Labeled Faces in the Wild" (LFW) [14], which is available via the LFW official site `http://vis-www.cs.umass.edu/lfw/results.html`, focuses on the face recognition task of pair matching. All the face images were taken in unconstrained environments, exhibiting " 'natural' variability in pose, lighting, focus, resolution, facial expression, age, gender, race, accessories, make-up, occlusions, background, and photographic quality" [14]. We test the performance of our approach on the 10 folds of view 2. In this task, given two face images, the goal is to decide whether two images are of the same person. This is a binary classification problem, with two possible outcomes: "same" or "different". LFW view 2 provides 10 folds of face sets where the sets of people in different folds are disjoint; when testing on one fold, the other nine folds can be used for training. Results of various approaches have been reported on the LFW official website.

[7] (1) Note that the former does not have a training process, and the "training" process of the latter only involves the "weights" therefor is really minor.
(2) In this experiment, we notice that the pixels near the boundaries of each images do not have LBP labels, therefore, we remove these pixels before partitioning into windows so that the size of windows near the boundaries are smaller than others.

Table 3. The accuracies of LBP Template and a few no-training approaches for LFW

Approach		Accuracy
SD-MATCHES		0.6410 ± 0.0062
H-XS-40		0.6945 ± 0.0048
GJD-BC-100		0.6847 ± 0.0065
LARK unsupervised		0.7223 ± 0.0049
LBP Template	Euclidean	0.6905 ± 0.0235
	Histogram intersection	0.7428 ± 0.0144
	Chi square statistic	0.7417 ± 0.0143
LBP Template, with Dummy Set	Euclidean	0.7352 ± 0.0180
	Histogram intersection	0.7633 ± 0.0152
	Chi square statistic	0.7613 ± 0.0172
LBP Template, with Dummy Set & Preprocessing	Euclidean	0.7508 ± 0.0146
	Histogram intersection	0.7730 ± 0.0157
	Chi square statistic	0.7727 ± 0.0162

We use LFW-a version of images (the images aligned using a commercial face alignment software) [20]. The images are of size 250×250. We first crop them into images of size 90×78 (by removing 88 pixel margins from top, 72 from bottom, and 86 pixel margins from both left and right sides). Note that, there were errors in the alignment of many images; we just keep them as they were (so some of the final cropped faces indeed are not correctly aligned).

For each pair of images, we first use the first one as a "gallery" image and the second as a "probe" image to compute the similarity, then use the second as a "gallery" image and the first as a "probe" image. We take the average of them as the final similarity value.[8]

The average of the best accuracies over the 10 folds are shown in Table 3. It should be compared to other unsupervised/no-training approaches as suggested on the LFW website; therefore we also include the results of all the unsupervised/no-training approaches available on the LFW website as of July 31st 2012: SD-MATCHES (L & R system with SIFT descriptors and MATCHES flavour), H-XS-40 (Histogram of LBP features with Chi Square similarity measure and 40 windows), GJD-BC-100 (Gabor Jets Descriptors with Borda Count measure and 100 reference images), and LARK representation without supervision [21]. We can see that the LBP displacement templates with histogram intersection and Chi Square statistic perform better than all other approaches.

Although this paper is intended to show only that displacement template can improve regular description approach; of course, we can further improve its performance, if some other strategies are employed. Simply following the suggestions of [1], where the result of GJD-BC-100 are obtained, by putting a few "reference faces" to find the relative ranking, we can improve the performance easily. Here, we use a dummy set of images: for the experiments in the i-th fold, we use the first images (named "***..-0001.jpg") of the first 10 individuals in the $(i-1)$th fold (when $i-1=0$, we use the 10th fold) as the dummy set. To compare a pair of images x and y, we use the dummy set as a gallery

[8] Due to space limit, we do not report the detailed results of T/FPLBP templates – they are slightly worse than LBP template, but better than the other approaches.

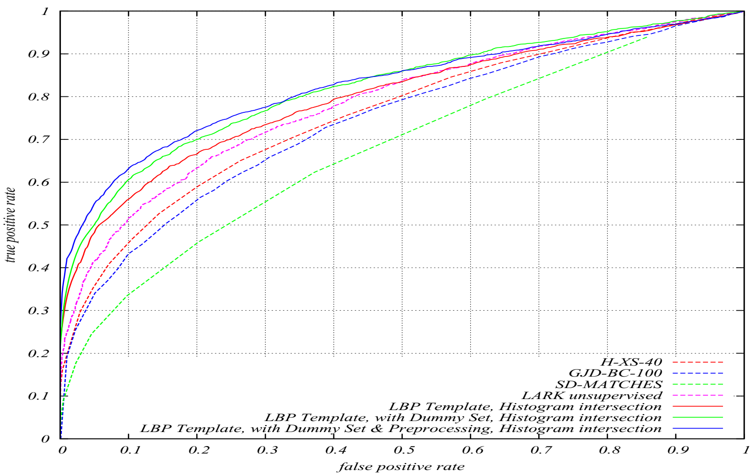


Fig. 2. ROC curves over View 2 of LFW

and calculate the similarity between each of these two images and each of the gallery images, then we define the relative similarity value as the similarity value between y and x divided by the sum of all the similarities between these two images and the dummy images. We can further improve the results by using the preprocessing process of [16]. All these results are shown in Table 3. We have also shown the ROC curves of all those unsupervised approaches aforementioned, and the ROC curves of our LBP template with histogram intersection (including the original one, the one using dummy set and the one using both dummy set & preprocessing) in Figure 2.

3.3 Experiments on FRGC

We also carry out the experiment 104 of FRGC version 1 [13]. This experiment is considered the most challenging in this dataset [19]. It requires to recognize 608 uncontrolled faces from 152 controlled gallery faces. Although standard protocol allows the usage of a training set, our approach does not use it. We normalize the face images into size 150×130 as we did for FERET experiments; the standard mask from FREET set is also used here. We report our results of LBP template with and without the "preprocessing" of [16] in Table 4. The best published result, which was done by Schwartz et al [19], is also listed. (Note: Schwartz et al listed a few results of other approaches in [19], none of which was better than Schwartz's approach. Due to space, we think it is not necessary to list those results as well as the corresponding references.) Therefore, we can see that LBP template with Chi square statistics seems to be the best of all. We should also note here that, our approach is an unsupervised approach while many known approaches for this problem are supervised with a training process.

Although we reach the best performance (as far as we know), we want to emphasize here: Just as most of face recognition algorithms, we may further

Table 4. Recognition rates of LBP Template approaches and the best published result on FRGC Experiment 104

LBP[22]	LBP Template			Schwartz[19]
	Euclidean	Histogram intersection	Chi square statistics	
28.1%	42.94%	47.37%	47.20%	78.2%
LBP with Preprocessing[22]	LBP Template with Preprocessing			
	Euclidean	Histogram intersection	Chi square statistics	
58.1%	74.01%	85.86%	86.18%	

improve the performances by adjusting perimeters, such as image sizes, the block sizes and pile sizes–and by doing so, we do reach better results for this dataset. We trust that, such adjustment, although can be used to improve accuracies to show off, does not contribute much academic values. Therefore, we do not discuss such adjustment here due to space limit.

4 Conclusions and Future Research

We have developed a local displacement description template for face recognition tasks. When LBP, TPLBP and FPLBP are used as the basic descriptors, our experiments have shown that LBP, TPLBP and FPLBP description templates can improve the performances of the original LBP, TPLBP and FPLBP approaches significantly. We expect that all other descriptors can be applied to our template structure. It is still interesting to see if the template approach using other descriptors can work better than LBP, TPLBP and FPLBP description templates.

It was shown in [6] that, when holistic algorithms are adopted, a "hard-combination" strategy ("Electoral College"), where the best matching is first selected from the corresponding blocks/regions of all gallery images and then a simple majority rule is used to make a final decision, works well for face identification. It would be interesting to investigate such a "hard-combination" strategy within our displacement description templates.

Acknowledgment. This work is partly supported by NSERC Discovery Grant (Grant No. 261403-2011 RGPIN), NSFC (Grants No. 61100147) and Zhejiang Provincial Natural Science Foundation (Grants No. LY12F03016).

References

1. Ruiz-del Solar, J., Verschae, R., Correa, M.: Recognition of faces in unconstrained environments: A comparative study. EURASIP Journal on Advances in Signal Processing, Article ID 184617, pages 19 (2009)
2. Ahonen, T., Hadid, A., Pietikäinen, M.: Face Recognition with Local Binary Patterns. In: Pajdla, T., Matas, J(G.) (eds.) ECCV 2004. LNCS, vol. 3021, pp. 469–481. Springer, Heidelberg (2004)
3. Torralba, A., Fergus, R., Freeman, W.: 80 million tiny images: A large data set for nonparametric object and scene recognition. PAMI 30(11), 1958–1970 (2008)

4. Liu, C., Yuen, J., Torralba, A.: Sift flow: Dense correspondence across scenes and its applications. PAMI 33(5), 978–994 (2011)
5. Zhang, B., Shan, S., Chen, X., Gao, W.: Histogram of Gabor phase patterns (HGPP): A novel object representation approach for face recognition. IEEE Transactions on Image Processing 16, 57–68 (2007)
6. Chen, L.: Stability as performance metric for subjective pattern recognition. In: Proc. of 19th ICPR, Tampa, FL, December 8-11 (2008)
7. Chen, L.: Pairwise macropixel comparison can work at least as well as advanced holistic algorithms for face recognition. In: BMVC, pp. 5.1–5.11 (2010)
8. Tokuda, N., Chen, L.: A new KE-free online ICALL system featuring error contingent feedback. Computer Assisted Language Learning 17(2), 177–201 (2004)
9. Tokuda, N., Chen, L.: An online intelligent tutoring system for language translation. IEEE Multimedia 8(3), 46–55 (2001)
10. Ahonen, T., Hadid, A., Pietikäinen, M.: Face description with local binary patterns: Application to face recognition. PAMI 28(12), 2037–2041 (2006)
11. Wolf, L., Hassner, T., Taigman, Y.: Descriptor based methods in the wild. In: Real-Life Images Workshop at ECCV (October 2008)
12. Phillips, P., Moon, H., Rizvi, S.A., Rauss, P.: The FERET evaluation methodology for face-recognition algorithms. PAMI 22(10), 1090–1104 (2000)
13. Phillips, P.J., Flynn, P.J., Scruggs, T., Bowyer, K.W., Chang, J., Hoffman, K., Marques, J., Min, J., Worek, W.: Overview of the face recognition grand challenge. In: CVPR, San Diego, vol. I, pp. 947–954 (June 2005)
14. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst (October 2007)
15. Beveridge, J.R., She, K., Draper, B., Givens, G.H.: A nonparametric statistical comparison of principal component and linear discriminant subspaces for face recognition. In: CVPR, pp. 535–542 (December 2001)
16. Tan, X., Triggs, B.: Enhanced Local Texture Feature Sets for Face Recognition Under Difficult Lighting Conditions. In: Zhou, S.K., Zhao, W., Tang, X., Gong, S. (eds.) AMFG 2007. LNCS, vol. 4778, pp. 168–182. Springer, Heidelberg (2007)
17. Zhang, W., Shan, S., Gao, W., Chen, X.: Local gabor binary pattern histogram sequence (LGBPH): A novel non-statistical model for face representation and recognition. In: ICCV, pp. 786–791 (2005)
18. Liu, J., Chen, S., Zhou, Z., Tan, X.: Single image subspace for face recognition. In: AFMG, pp. 205–219 (2007)
19. Schwartz, W.R., Guo, H., Davis, L.S.: A Robust and Scalable Approach to Face Identification. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part VI. LNCS, vol. 6316, pp. 476–489. Springer, Heidelberg (2010)
20. Taigman, Y., Wolf, L., Hassner, T.: Multiple one-shots for utilizing class label information. In: BMVC (2009)
21. Seo, H., Milanfar, P.: Face verification using the lark representation. IEEE Transactions on Information Forensics and Security 6(4), 1275–1286 (2011)
22. Holappa, J., Ahonen, T., Pietikainen, M.: An optimized illumination normalization method for face recognition. In: Proc. of 2nd IEEE International Conference on Biometrics: Theory, Applications and Systems, September 29-October 1, pp. 1–6 (2008)
23. Chen, L.: Robustness Instead of Accuracy Should Be The Primary Objective for Subjective Pattern Recognition Research. To Appear in: Computational Intelligence (2012), doi: 10.1111/j.1467-8640.2012.00439.x