HAL
open science

Enhanced Local Texture Feature Sets for Face Recognition under Difficult Lighting Conditions

Xiaoyang Tan, Bill Triggs

To cite this version:

Xiaoyang Tan, Bill Triggs. Enhanced Local Texture Feature Sets for Face Recognition under Difficult Lighting Conditions. AMFG - 3rd International Workshop Analysis and Modelling of Faces and Gestures, Oct 2007, Rio de Janeiro, Brazil. pp.168-182, 10.1007/978-3-540-75690-3_13 . inria-00548674

HAL Id: inria-00548674

https://inria.hal.science/inria-00548674

Submitted on 20 Dec 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Enhanced Local Texture Feature Sets for Face Recognition under Difficult Lighting Conditions

Xiaoyang Tan and Bill Triggs

INRIA & Laboratoire Jean Kuntzmann, 655 avenue de l'Europe, Montbonnot 38330, France
{xiaoyang.tan,bill.triggs}@imag.fr

Abstract. Recognition in uncontrolled situations is one of the most important bottlenecks for practical face recognition systems. We address this by combining the strengths of robust illumination normalization, local texture based face representations and distance transform based matching metrics. Specifically, we make three main contributions: (*i*) we present a simple and efficient preprocessing chain that eliminates most of the effects of changing illumination while still preserving the essential appearance details that are needed for recognition; (*ii*) we introduce Local Ternary Patterns (LTP), a generalization of the Local Binary Pattern (LBP) local texture descriptor that is more discriminant and less sensitive to noise in uniform regions; and (*iii*) we show that replacing local histogramming with a local distance transform based similarity metric further improves the performance of LBP/LTP based face recognition. The resulting method gives state-of-the-art performance on three popular datasets chosen to test recognition under difficult illumination conditions: Face Recognition Grand Challenge version 1 experiment 4, Extended Yale-B, and CMU PIE.

1 Introduction

One of the key challenges of face recognition is finding efficient and discriminative facial appearance descriptors that can counteract large variations in illumination, pose, facial expression, ageing, partial occlusions and other changes [27]. There are two main approaches: geometric feature-based descriptors and appearance-based descriptors. Geometric descriptors can be hard to extract reliably under variations in facial appearance, while appearance-based ones such as eigenfaces tend to blur out small details owing to residual spatial registration errors. Recently, representations based on local pooling of local appearance descriptors have drawn increasing attention because they can capture small appearance details in the descriptors while remaining resistant to registration errors owing to local pooling. Another motivation is the observation that human visual perception is well-adapted to extracting and pooling local structural information ('micro-patterns') from images [2]. Methods in this category include Gabor wavelets [16], local autocorrelation filters [11], and Local Binary Patterns [1].

In this paper we focus on Local Binary Patterns (LBP) and their generalizations. LBP's are a computationally efficient nonparametric local image texture descriptor. They have been used with considerable success in a number of visual recognition tasks including face recognition [1,2,20]. LBP features are invariant to monotonic gray-level changes by design and thus are usually considered to require no image preprocessing

before use[1]. In fact, LBP itself is sometimes used as a lighting normalization stage for other methods [12]. However, in practice the reliability of LBP decreases significantly under large illumination variations (*c.f.* table 3). Lighting effects involve complex local interactions and the resulting images often violate LBP's basic assumption that gray-level changes monotonically. We have addressed this problem by developing a simple and efficient image preprocessing chain that greatly reduces the influence of illumination variations, local shadowing and highlights while preserving the elements of visual appearance that are needed for recognition.

Another limitation of LBP is its sensitivity to random and quantization noise in uniform and near-uniform image regions such as the forehead and cheeks. To counter this we extend LBP to Local Ternary Patterns (LTP), a 3-valued coding that includes a threshold around zero for improved resistance to noise. LTP inherits most of the other key advantages of LBP such as computational efficiency.

Current LBP based face recognition methods partition the face image into a grid of fixed-size cells for the local pooling of texture descriptors (LBP histograms). This coarse (and typically abrupt) spatial quantization is somewhat arbitrary and not necessarily well adapted to local facial morphology. It inevitably causes some loss of discriminative power. To counter this we use distance transform techniques to create local texture comparison metrics that have more controlled spatial gradings.

To illustrate the effectiveness of our approach we present experimental results on three state-of-the-art face recognition datasets containing large lighting variations similar to those encountered in natural images taken under uncontrolled conditions: Face Recognition Grand Challenge version 1 experiment 1.0.4 ('FRGC-104') [19]; Extended Yale illumination face database B ('Extended Yale-B') [9,15]; and CMU PIE [22].

2 Related Work

As emphasized by the recent FRVT and FRGC trials [19], illumination variations are one of the most important bottlenecks for practical face recognition systems. Generally, one can cope with this in two ways. The first uses training examples to learn a global model of the possible illumination variations, for example a linear subspace or manifold model, which then generalizes to the variations seen in new images [5,3]. The disadvantage is that many training images are required.

The second approach seeks conventional image processing transformations that reduce the image to a more "canonical" form in which the variations are suppressed. This has the merit of easy application to real images and the lack of a need for comprehensive training data. Given that complete illumination invariants do not exist [7], one must content oneself with finding representations that are resistant to the most common classes of natural illumination variations. Most methods exploit the fact that these are typically characterized by relatively low spatial frequencies. For example, the Multiscale Retinex (MSR) method of Jobson *et al.* [13] normalizes the illumination by dividing the image by a smoothed version of itself. A similar idea (with a different local filter) is used by Wang *et al.* [23] in the Self Quotient Image model (SQI). More recently, Chen *et al.*

[1] One exception is Local Gabor Binary Pattern Histogram Sequences [26] whose Gabor magnitude mapping can be regarded as a special kind of preprocessing for LBP.

[8] improved SQI by using Logarithmic Total Variation (LTV) smoothing, and Gross & Brajovic (GB) [10] developed an anisotropic smoothing method that relies on the iterative estimation of a blurred version of the original image. Some comparative results for these and related works can be found in [21].

In this paper we adopt the "canonical form" philosophy, basing our method on a chain of efficient processing steps that normalize for various effects of the changing illumination environment. The main advantages of our method are simplicity, computational efficiency and robustness to lighting changes and other image quality degradations such as blurring.

We describe our LBP/LTP face descriptors and their distance transform based similarity metric in the next two sections, detailing our preprocessing method in §5 and concluding with experiments and discussion.

3 Local Ternary Patterns

3.1 Local Binary Patterns (LBP)

Ojala *et al*. [17] introduced the Local Binary Pattern operator in 1996 as a means of summarizing local gray-level structure. The operator takes a local neighborhood around each pixel, thresholds the pixels of the neighborhood at the value of the central pixel and uses the resulting binary-valued image patch as a local image descriptor. It was originally defined for 3×3 neighborhoods, giving 8 bit codes based on the 8 pixels around the central one. Formally, the LBP operator takes the form

$$LBP(x_c, y_c) = \sum_{n=0}^{7} 2^n \, s(i_n - i_c) \tag{1}$$

where in this case n runs over the 8 neighbors of the central pixel c, i_c and i_n are the gray-level values at c and n, and $s(u)$ is 1 if $u \geq 0$ and 0 otherwise. The LBP encoding process is illustrated in fig. 1.

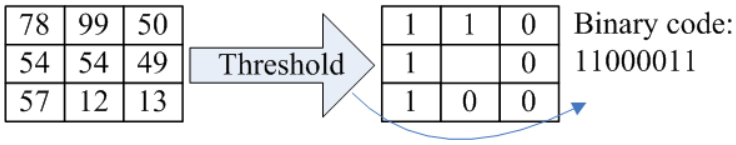


Fig. 1. Illustration of the basic LBP operator.

Two extensions of the original operator were made in [18]. The first defined LBP's for neighborhoods of different sizes, thus making it feasible to deal with textures at different scales. The second defined the so-called *uniform patterns*: an LBP is 'uniform' if it contains at most one 0-1 and one 1-0 transition when viewed as a circular bit string. For example, the LBP code in fig. 1 is uniform. Uniformity is an important concept in the LBP methodology, representing primitive structural information such as edges and corners. Ojala *et al*. observed that although only 58 of the 256 8-bit patterns are uniform, nearly 90 percent of all observed image neighbourhoods are uniform. In methods that

histogram LBP's, the number of bins can be thus significantly reduced by assigning all non-uniform patterns to a single bin, often without losing too much information.

3.2 Local Ternary Patterns (LTP)

LBP's are resistant to lighting effects in the sense that they are invariant to monotonic gray-level transformations, and they have been shown to have high discriminative power for texture classification [17]. However because they threshold at exactly the value of the central pixel i_c they tend to be sensitive to noise, especially in near-uniform image regions. Given that many facial regions are relatively uniform, it is potentially useful to improve the robustness of the underlying descriptors in these areas.

This section extends LBP to 3-valued codes, *Local Ternary Patterns*, in which gray-levels in a zone of width $\pm t$ around i_c are quantized to zero, ones above this are quantized to $+1$ and ones below it to -1, *i.e.* the indicator $s(u)$ is replaced by a 3-valued function:

$$s'(u, i_c, t) = \begin{cases} 1, & u \geq i_c + t \\ 0, & |u - i_c| < t \\ -1, & u \leq i_c - t \end{cases} \tag{2}$$

and the binary LBP code is replaced by a ternary LTP code. Here t is a user-specified threshold (so LTP codes more resistant to noise, but no longer strictly invariant to gray-level transformations). The LTP encoding procedure is illustrated in fig. 2. Here the threshold t was set to 5, so the tolerance interval is $[49, 59]$.

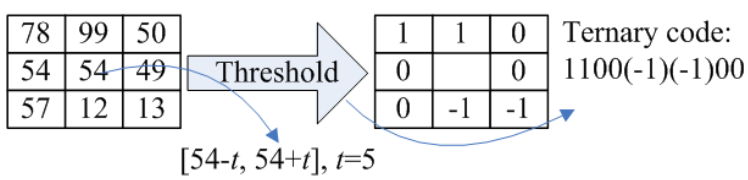


Fig. 2. Illustration of the basic LTP operator.

When using LTP for visual matching we could use 3^n valued codes, but the uniform pattern argument also applies in the ternary case. For simplicity the experiments below use a coding scheme that splits each ternary pattern into its positive and negative parts as illustrated in fig. 3, subsequently treating these as two separate channels of LBP descriptors for which separate histograms and similarity metrics are computed, combining these only at the end of the computation.

LTP's bear some similarity to the texture spectrum (TS) technique from the early 1990's [24]. However TS did not include preprocessing, thresholding, local histograms or uniform pattern based dimensionality reduction and it was not tested on faces.

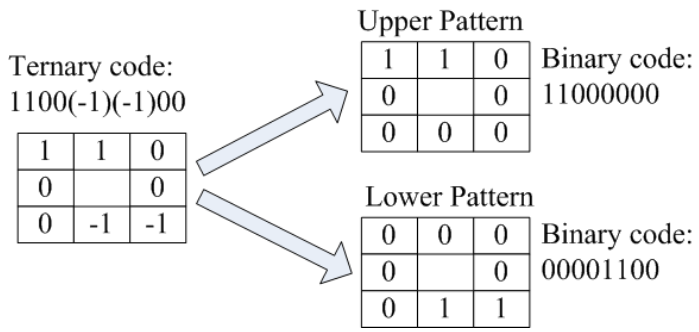


Fig. 3. An example of the splitting of an LTP code into positive and negative LBP codes.

4 Distance Transform based Similarity Metric

T. Ahonen *et al*. introduced an LBP based method for face recognition [1] that divides the face into a regular grid of cells and histograms the uniform LBP's within each cell, finally using nearest neighbor classification in the χ^2 histogram distance for recognition:

$$\chi^2(p,q) = \sum_i \frac{(p_i - q_i)^2}{p_i + q_i} \tag{3}$$

Here p, q are two image descriptors (histogram vectors). Excellent results were obtained on the FERET dataset.

Possible criticisms of this method are that subdividing the face into a regular grid is somewhat arbitrary (cells are not necessarily well aligned with facial features), and that partitioning appearance descriptors into grid cells is likely to cause both aliasing (due to abrupt spatial quantization) and loss of spatial resolution (as position within a grid cell is not coded). Given that the aim of coding is to provide illumination- and outlier-robust appearance-based correspondence with some leeway for small spatial deviations due to misalignment, it seems more appropriate to use a Hausdorff distance like similarity metric that takes each LBP or LTP pixel code in image X and tests whether a similar code appears at a nearby position in image Y, with a weighting that decreases smoothly with image distance. Such a scheme should be able to achieve discriminant appearance-based image matching with a well-controllable degree of spatial looseness.

We can achieve this using Distance Transforms [6]. Given a 2-D reference image X, we find its image of LBP or LTP codes and transform this into a set of sparse binary images b^k, one for each possible LBP or LTP code value k (*i.e*. 59 images for uniform codes). Each b^k specifies the pixel positions at which its particular LBP or LTP code value appears. We then calculate the distance transform image d^k of each b^k. Each pixel of d^k gives the distance to the nearest image X pixel with code k (2D Euclidean distance is used in the experiments below). The distance or similarity metric from image X to image Y is then:

$$D(X,Y) = \sum_{\text{pixels } (i,j) \text{ of } Y} w(d_X^{k_Y(i,j)}(i,j)) \tag{4}$$

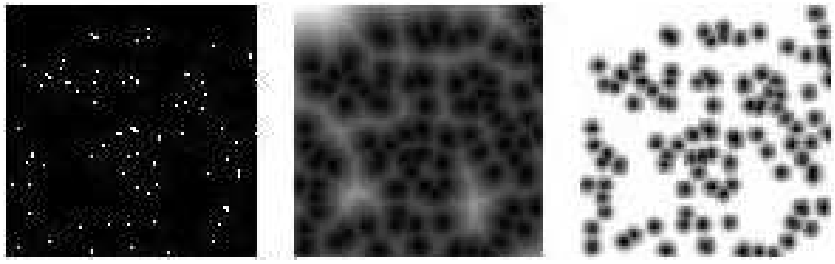


Fig. 4. From left to right: a binary layer, its distance transform, and the truncated linear version of this.

Here, $k_Y(i,j)$ is the code value of pixel (i,j) of image Y and $w()$ is a user-defined function[2] giving the penalty to include for a pixel at the given spatial distance from the nearest matching code in X. In our experiments we tested both Gaussian similarity metrics $w(d) = \exp\{-(d/\sigma)^2/2\}$ and truncated linear distances $w(d) = \min(d, \tau)$. Their performance is similar, with truncated distances giving slightly better results overall. The default parameter values (for 120×120 face images in which an iris or nostril has a radius of about 6 pixels) are $\sigma = 3$ pixels and $\tau = 6$ pixels.

Fig. 4 shows an example of a binary layer and its distance transforms. For a given target the transform can be computed and mapped through $w()$ in a preprocessing step, after which matching to any subsequent image takes $\mathcal{O}$(number of pixels) irrespective of the number of code values.

5 Illumination Normalization

5.1 The Processing Sequence

This section describes our image preprocessing method. It incorporates a series of steps chosen to counter the effects of illumination variations, local shadowing and highlights, while still preserving the essential elements of visual appearance for use in recognition. Although not done by design, the final chain is reminiscent of certain preprocessing stages found in the mammalian visual cortex. In detail, the steps are as follows.

Gamma Correction. This is a nonlinear gray-level transformation that replaces gray-level I with I^γ (for $\gamma > 0$) or $\log(I)$ (for $\gamma = 0$), where $\gamma \in [0,1]$ is a user-defined parameter. It has the effect of enhancing the local dynamic range of the image in dark or shadowed regions, while compressing it in bright regions and at highlights. The basic principle is that the intensity of the light reflected from an object is the product of the incoming illumination L (which is piecewise smooth for the most part) and the local surface reflectance R (which carries detailed object-level appearance information). We want to recover object-level information independent of illumination, and taking logs

[2] w is monotonically increasing for a distance metric and monotonically decreasing for a similarity one. In D, note that each pixel in Y is matched to the nearest pixel with the same code in X. This is not symmetric between X and Y even if the underlying distance d is, but it can be symmetrized if desired.

makes the task easier by converting the product into a sum: for constant local illumination, a given reflectance step produces a given step in $\log(I)$ irrespective of the actual intensity of the illumination. In practice a full $\log$ transformation is often too strong, tending to over-amplify the noise in dark regions of the image, but a power law with exponent γ in the range $[0, 0.5]$ is a good compromise. Here we use $\gamma = 0.2$ as the default setting.

Difference of Gaussian (DoG) Filtering. Gamma correction does not remove the influence of overall intensity gradients such as shading effects. Shading induced by surface structure is potentially a useful visual cue but it is predominantly low frequency spatial information that is hard to separate from effects caused by illumination gradients. High pass filtering removes both the useful and the incidental information, thus simplifying the recognition problem and in many cases increasing the overall system performance. Similarly, suppressing the highest spatial frequencies reduces aliasing and noise, and in practice it often manages to do so without destroying too much of the underlying signal on which recognition needs to be based. DoG filtering is a convenient way to obtain the resulting bandpass behaviour. Fine spatial detail is critically important for recognition so the inner (smaller) Gaussian is typically quite narrow ($\sigma_0 \leq 1$ pixel), while the outer one might have σ_1 of 2–4 pixels or more, depending on the spatial frequency at which low frequency information becomes misleading rather than informative. Given the strong lighting variations in our datasets we find that $\sigma_1 \approx 2$ typically gives the best results, but values up to about 4 are not too damaging and may be preferable for datasets with less extreme lighting variations. LBP and LTP seem to benefit from a little smoothing ($\sigma_0 \approx 1$), perhaps because pixel based voting is sensitive to aliasing artifacts. Below we use $\sigma_0 = 1.0$ and $\sigma_1 = 2.0$ by default[3].

We implement the filters using explicit convolution. If the face is part of a larger image the gamma correction and prefilter should be run on an appropriate region of this before cutting out the face image. Otherwise extend-as-constant boundary conditions should be used: using extend-as-zero or wrap-around (FFT) boundary conditions significantly reduces the overall performance, in part because it introduces strong gradients at the image borders that disturb the subsequent contrast equalization stage. If DoG is run without prior gamma normalization, the resulting images clearly show the extent to which local contrast (and hence visual detail) is reduced in shadowed regions.

Masking. If a mask is needed to suppress facial regions that are felt to be irrelevant or too variable, it should be applied at this point. Otherwise, either strong artificial gray-level edges are introduced into the convolution, or invisible regions are taken into account during contrast equalization.

Contrast Equalization. The final step of our preprocessing chain globally rescales the image intensities to standardize a robust measure of overall contrast or intensity variation. It is important to use a robust estimator because the signal typically still contains a

[3] Curiously, for some datasets it also helps to offset the center of the larger filter by 1–2 pixels relative to the center of the smaller one, so that the final prefilter is effectively the sum of a centered DoG and a low pass spatial derivative. The best direction for the displacement is somewhat variable but typically diagonal. The effect is not consistent enough to be recommended practice, but it might repay further investigation.

small admixture of extreme values produced by highlights, garbage at the image borders and small dark regions such as nostrils. One could, *e.g.*, use the median of the absolute value of the signal for this, but here we have preferred a simple and rapid approximation based on a two stage process:

$$I(x,y) \leftarrow \frac{I(x,y)}{(\mathrm{mean}(|I(x',y')|^a))^{1/a}} \tag{5}$$

$$I(x,y) \leftarrow \frac{I(x,y)}{(\mathrm{mean}(\min(\tau, |I(x',y')|)^a))^{1/a}} \tag{6}$$

Here, a is a strongly compressive exponent that reduces the influence of large values, τ is a threshold used to truncate large values after the first phase of normalization, and the mean is over the whole (unmasked part of the) image. By default we use $\alpha = 0.1$ and $\tau = 10$.

The resulting image is now well scaled but it can still contain extreme values. To reduce their influence on subsequent stages of processing, we finally apply a nonlinear function to compress over-large values. Here we use the hyperbolic tangent $I(x,y) \leftarrow \tau \tanh(I(x,y)/\tau)$, thus limiting I to the range $(-\tau, \tau)$.

5.2 Robustness and Computation Time

To illustrate the behavior of the proposed preprocessing method, we examine its effect on the LBP histogram feature set. Fig. 5 shows a matching pair of target and query images chosen randomly from the FRGC dataset (see below). LBP features are extracted from corresponding local regions of the two images (the white squares in fig. 5), both before and after illumination normalization. The resulting histograms are shown in fig. 6. It can be seen that without illumination normalization the descriptors of the two images are very different, but that this is significantly reduced by normalization. In fact, the χ^2 distance between the two spatial histograms reduces from 93.4 before normalization to 25.0 after it.

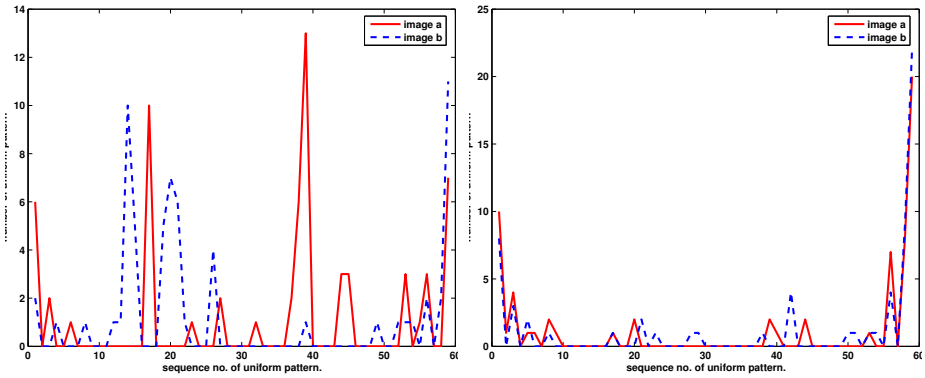
Since run time is a critical factor in many practical applications, it is also interesting to consider the computational load of our normalization chain. Our method uses only simple closed-form image processing operations so it is much more efficient than ones that require expensive iterative optimizations such as Logarithmic Total Variation [8] and anisotropic diffusion [10]. Our (unoptimized Matlab) implementation takes only about 50 ms to process a 120×120 pixel face image on a 2.8 GHz P4, allowing face preprocessing to be performed in real time. In comparison, the current implementation of LTV is about 300 times slower.

6 Experiments

We now present experiments that illustrate the effectiveness of our method using three publicly available face datasets with large illumination variations: Face Recognition Grand Challenge version 1 experiment 1.0.4 ('FRGC-104') [19]; Extended Yale Face Database B ('Extended Yale-B') [15]; and CMU PIE [22].

(a) (b)

Fig. 5. Two face images from the same subject. The LBP histograms from the marked region are shown in fig. 6.



(a) (b)

Fig. 6. LBP histograms for the marked region in fig. 5, before (a) and after (b) illumination normalization.

6.1 Experimental Settings

We use only frontal face views but lighting, expression and identity may all vary. All of the images undergo the same geometric normalization prior to analysis: conversion to 8 bit gray-scale images; rigid scaling and image rotation to place the centers of the two eyes at fixed positions, using the eye coordinates supplied with the original datasets; and image cropping to 120×120 pixels.

The default settings of the various parameters of our methods are summarized in table 1. Unless otherwise noted, these apply to all experiments. General guidelines on how to set these parameters can be found in section 5. We find that the proposed preprocessing method gives similar results over a broad range of parameter settings, which greatly facilitates the selection of parameters.

For comparison, we also tested the LTP/DT features with our preprocessing chain replaced with several alternative illumination normalization methods including standard histogram equalization (HE), Self-Quotient Image (SQI [23]), Multiscale Retinex

Table 1. Default parameter settings for our methods.

Procedure	Parameter	Value
Gamma Correction	γ	0.2
DoG Filtering	σ_0	1
	σ_1	2
Contrast Equalization	α	0.1
	τ	10
LTP	t	0.1-0.2
LBP/LTP χ^2 cell size		8×8

(a) (b)

Fig. 7. Examples of images from FRGC-104: (a) target images (upper row) and query images (lower row) without illumination preprocessing; (b) the corresponding illumination normalized images from the proposed preprocessing chain.

(MSR [13]), Logarithmic Total Variation (LTV [8]) and Gross & Brajovic's anisotropic smoothing (GB [10]). The implementations of these algorithms were based in part on the publicly available *Torch3Vision* toolbox (http://torch3vision.idiap.ch) with its default or recommended parameter settings. We would also like to thank Terrence Chen for making his implementation of LTV [8] available to us.

Although we do not show the results here owing to lack of space, our preprocessing method also gives substantial performance gains for the other image descriptors and recognition algorithms tested including eigen- and Fisher-faces under a range of descriptor normalizations and distance metrics, kernelized versions of these, and Gabor features.

6.2 Results on FRGC-104

We now show results on the dataset of the FRGC 1.0.4 experiment [19]. This dataset is challenging because, although the gallery images were obtained under carefully controlled conditions, the probe images were captured in uncontrolled indoor and outdoor settings. Fig. 7 shows some example images. There are large changes in illumination, appearance and expression. For the experiments reported here we use nearest-neighbour matching of each of the 943 probe images against each of the 943 gallery images. The training images are not used at all.

Fig. 8 shows the extent to which standard LBP can be improved by combining the three enhancements proposed in this paper: using preprocessing (PP); replacing

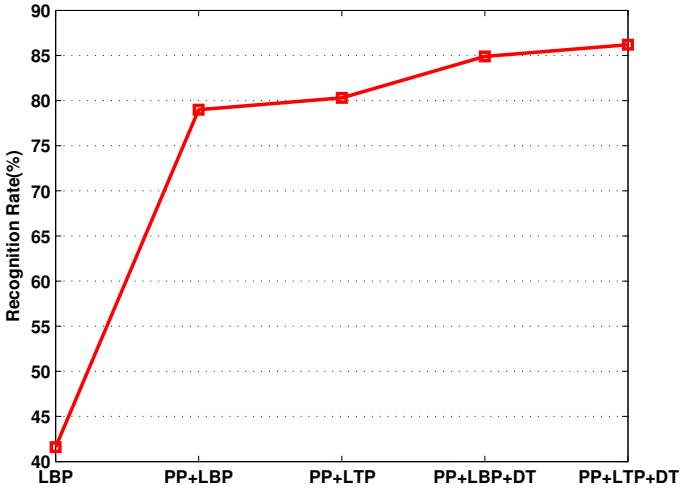


Fig. 8. Overall results for the proposed methods on FRGC-104.

LBP with LTP; replacing local histogramming and the χ^2 histogram distance with the Distance Transform based similarity metric (DT). Overall the absolute recognition rate is increased by about 45% relative to standard unpreprocessed LBP/χ^2. Preprocessing alone boosts the performance by 37% (from 41.6% to 79.0%). Replacing LBP with LTP improves the recognition rate to 80.4% and adding DT further improves it to 86.3%. LTP consistently outperforms LBP by a small margin, and the DT based similarity metric significantly improves on χ^2 histogram distance independent of the local region size. By way of comparison, the best previous performance that we are aware of on this dataset (unfortunately for a different and hence not strictly comparable experimental setup) is 78.0% [16].

To evaluate the performance of our preprocessing chain, we replaced it with several competing illumination normalization methods. The results are shown in fig. 9. Our method significantly outperforms the other ones tested. In particular, the performance of the sophisticated methods GB and LTV is disappointing on this dataset (*c.f.* their results on Yale-B below). It seems that on average they remove too much information during normalization. The reasons for this deserve further investigation. Note that besides large lighting variations, the dataset contains various other commonly encountered variations including ageing effects and image blurring.

6.3 Results on Extended Yale-B

The Yale Face Dataset B [4] containing 10 people under 64 different illumination conditions has been the de facto standard for studies of variable lighting over the past decade. It was recently updated to the Extended Yale Face Database B [15], containing 38 subjects under 9 poses and 64 illumination conditions. In both cases the images are divided

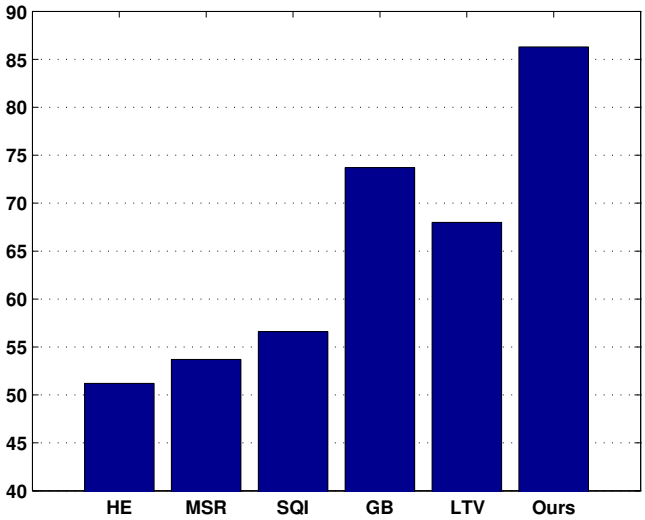
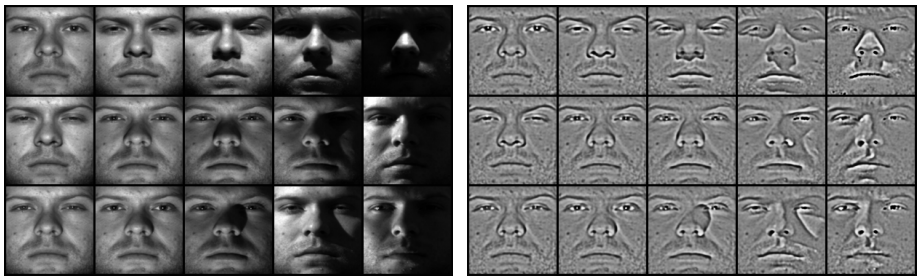


Fig. 9. Comparison of recognition rates of different preprocessing methods on the FRGC-104 database with LTP/DT features (%).



(a) (b)

Fig. 10. Examples of images of one person from the Extended Yale-B frontal database. The columns respectively give images from subsets 1 to 5. (a) input images. (b) the corresponding illumination normalized images from our preprocessing chain.

into five subsets according to the angle between the light source direction and the central camera axis ($12°, 25°, 50°, 77°, 90°$). Example images are shown in fig. 10. For our experiments, the images with the most neutral light sources ('A+000E+00') were used as the gallery, and all frontal images of each of the standard subsets 1–5 were used as probes (2414 images of 38 subjects for the Extended dataset, 640 of 10 for the standard one).

Our LTP/DT method gives perfect results on all 5 subsets of standard Yale-B[4]. On Extended Yale-B, the overall performance of our methods is shown in table 2 and the

[4] For comparison, on Yale-B subsets 2,3,4: Harmonic image Exemplars give 100, 99.7, 96.9% [25]; nine point of light gives 100, 100, 97.2% [14]; and Gradient Angle gives 100, 100, 98.6% [7]. These authors do not test on the most difficult set, 5.

Table 2. Overall recognition rates (%) on Extended Yale-B.

Methods	Subset No.(number of probes) 1 (363)	2 (456)	3 (455)	4 (526)	5 (714)
LBP/χ^2	100.0	100.0	97.6	65.2	44.4
PP+LBP/χ^2	100.0	100.0	99.8	97.3	87.5
PP+LTP/χ^2	100.0	100.0	100.0	98.1	97.1
PP+LBP/DT	100.0	100.0	100.0	**99.4**	95.2
PP+LTP/DT	**100.0**	**100.0**	**100.0**	99.2	**97.2**

Table 3. Recognition rates (%) with different preprocessing methods on Extended Yale-B database.

Methods	Subset No.(number of probes) 1 (263)	2 (456)	3 (455)	4 (526)	5 (714)
w/o	100.0	100.0	97.1	66.8	60.6
HE	100.0	100.0	97.4	76.8	66.8
MSR	100.0	99.8	96.7	93.6	82.1
SQI	100.0	99.8	97.8	94.1	82.4
GB	100.0	100.0	99.8	96.9	90.6
LTV	100.0	100.0	98.4	98.9	**97.6**
Ours	**100.0**	**100.0**	**100.0**	**99.2**	97.2

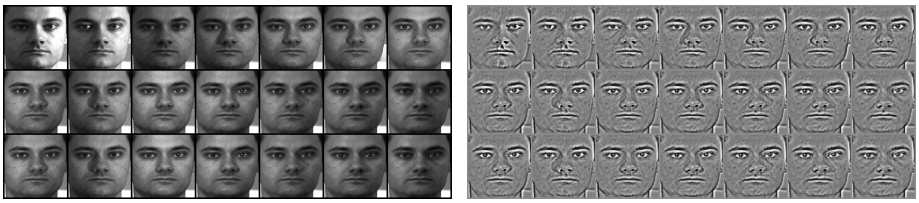
Methods	HE	MSR	SQI	GB	LTV	Ours
Accuracy	98.3	98.3	98.5	99.7	100.0	**100.0**

Table 4. Recognition Rates with different preprocessing methods on the CMU-PIE database (%)

effect of using different illumination normalization methods with LTP/DT is shown in table 3. In table 2, note that for the most difficult subset (5), including either LTP or the distance transform increases performance over PP+LBP/χ^2 by respectively about 10.0% and 8.0%. As the first row of table 3 indicates, even without image preprocessing the system performs quite well under the mild lighting changes of subsets 1–3 (*c.f.* fig. 10). However its performance drops significantly under the more extreme lighting conditions of subsets 4–5. In these cases, illumination normalizers such as LTV [8] and GB [10] significantly improve the accuracy. Our three-stage preprocessing chain achieves the top performance on subsets 3–4 and is beaten only marginally by LTV on set 5. As mentioned above, LTV is also about 300 times slower than our preprocessing method.

6.4 Results on the CMU PIE database

CMU PIE [22] is another dataset that is often used for studies of illumination variations. We tested a subset containing frontal images of 68 individuals illuminated from 21 directions. Images of one individual are shown in fig. 11. As before, the images with the most neutral lighting are used for the gallery and all of the remaining images are used as probes. When preprocessing is included, all of our LBP and LTP based schemes achieve 100.0% on this dataset. Comparative results for LTP/DT with different preprocessing methods are shown in table 4. As before, our preprocessing chain and LTV are the top performers.



(a) (b)

Fig. 11. Example images of one person from the CMU PIE database: (a) original images; (b) the corresponding normalized images obtained with the proposed preprocessing chain.

7 Conclusions

We have presented new methods for face recognition under uncontrolled lighting based on robust preprocessing and an extension of the Local Binary Pattern (LBP) local texture descriptor. There are three main contributions: (*i*) a simple, efficient image preprocessing chain whose practical recognition performance is comparable to or better than current (often much more complex) illumination normalization methods; (*ii*) a rich descriptor for local texture called Local Ternary Patterns (LTP) that generalizes LBP while fragmenting less under noise in uniform regions; and (*iii*) a distance transform based similarity metric that captures the local structure and geometric variations of LBP/LTP face images better than the simple grids of histograms that are currently used. The combination of these enhancements provides very promising performance on three well-known face datasets that contain widely varying lighting conditions.

Work in progress includes experiments on the much larger FRGC 2.0.4 dataset and tests against subspace based recognition methods.

References

[1] T. Ahonen, A. Hadid, and M. Pietikainen. Face recognition with local binary patterns. In *Proc. ECCV'04*, pages 469–481, Prague, 2005.

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

[3] R. Basri and D. Jacobs. Lambertian reflectance and linear subspaces. *IEEE TPAMI*, 25(2):218–233, February 2003.

[4] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE TPAMI*, 19(7):711–720, 1997.

[5] P. Belhumeur and D. Kriegman. What is the set of images of an object under all possible illumination conditions. *IJCV*, 28(3):245–260, 1998.

[6] G. Borgefors. Distance transformations in digital images. *Comput. Vision Graph. Image Process.*, 34(3):344–371, 1986.

[7] H. Chen, P. Belhumeur, and D. Jacobs. In search of illumination invariants. In *Proc. CVPR'00*, pages I: 254–261, 2000.

[8] T. Chen, W. Yin, X. Zhou, D. Comaniciu, and T. Huang. Total variation models for variable lighting face recognition. *IEEE TPAMI*, 28(9):1519–1524, 2006.

[9] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman. From few to many: illumination cone models for face recognition under variable lighting and pose. *IEEE TPAMI*, 23(6):643–660, 2001.

[10] R. Gross and V. Brajovic. An image preprocessing algorithm for illumination invariant face recognition. In *Proc. AVBPA'03*, pages 10–18, 2003.
[11] F. Guodail, E. Lange, and T. Iwamoto. Face recognition system using local autocorrelations and multiscale integration. *IEEE TPAMI*, 18(10):1024–1028, 1996.
[12] G. Heusch, Y. Rodriguez, and S. Marcel. Local binary patterns as an image preprocessing for face authentication. In *Proc. FGR'06*, pages 9–14, USA, 2006.
[13] D. Jobson, Z. Rahman, and G. Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE TIP*, 6(7):965–976, 1997.
[14] K. Lee, J. Ho, and D. Kriegman. Nine points of light: Acquiring subspaces for face recognition under variable lighting. In *Proc. CVPR'01*, pages I:519–526, 2001.
[15] K. Lee, J. Ho, and D. Kriegman. Acquiring linear subspaces for face recognition under variable lighting. *IEEE TPAMI*, 27(5):684–698, 2005.
[16] C. Liu. Capitalize on dimensionality increasing techniques for improving face recognition grand challenge performance. *IEEE TPAMI*, 28(5):725–737, 2006.
[17] T. Ojala, M. Pietikainen, and D. Harwood. A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition*, 29, 1996.
[18] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invarianat texture classification with local binary patterns. *IEEE TPAMI*, 24(7):971–987, 2002.
[19] P. J. Phillips, P. J. Flynn, W. T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. J. Worek. Overview of the face recognition grand challenge. In *Proc. CVPR'05*, pages 947–954, San Diego, CA, 2005.
[20] Y. Rodriguez and S. Marcel. Face authentication using adapted local binary pattern histograms. In *ECCV*, 2006.
[21] J. Short, J. Kittler, and K. Messer. A comparison of photometric normalization algorithms for face verification. In *Proc. AFGR'04*, pages 254–259, 2004.
[22] T. Sim, S. Baker, and M. Bsat. The cmu pose, illumination, and expression (pie) database of human faces. Technical Report CMU-RI-TR-01-02, Robotics Institute, Carnegie Mellon University, January 2001.
[23] H. Wang, S. Li, and Y. Wang. Face recognition under varying lighting conditions using self quotient image. In *Proc. AFGR'04*, 2004.
[24] L. Wang and D. He. Texture classification using texture spectrum. *Pattern Recognition*, 23:905–910, 1990.
[25] L. Zhang and D. Samaras. Face recognition under variable lighting using harmonic image exemplars. In *Proc. CVPR'03*, pages I: 19–25, 2003.
[26] W. Zhang, S. Shan, W. Gao, and H. Zhang. Local Gabor Binary Pattern Histogram Sequence (LGBPHS): A novel non-statistical model for face representation and recognition. In *Proc. ICCV'05*, pages 786–791, Beijing, China, 2005.
[27] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Survey*, 34(4):399–485, 2003.