Pattern Recognition ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at SciVerse ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Face recognition using the POEM descriptor

Ngoc-Son Vu [a,b,*], Hannah M. Dee [c], Alice Caplier [a]

[a] GIPSA-lab, Grenoble Institute of Technology, BP 46, 38402 Saint Martin d'Hères, Cedex, France
[b] Vesalis, Clermont Ferrand, France
[c] Computer Science, Aberystwyth University, Penglais, Aberystwyth SY23 3DB, UK

ARTICLE INFO

Article history:
Received 31 May 2010
Received in revised form
21 November 2011
Accepted 20 December 2011

Keywords:
Face recognition
Face descriptors
FERET
LFW

ABSTRACT

Real-world face recognition systems require careful balancing of three concerns: computational cost, robustness, and discriminative power. In this paper we describe a new descriptor, POEM (patterns of oriented edge magnitudes), by applying a self-similarity based structure on oriented magnitudes and prove that it addresses all three criteria. Experimental results on the FERET database show that POEM outperforms other descriptors when used with nearest neighbour classifiers. With the LFW database by combining POEM with GMMs and with multi-kernel SVMs, we achieve comparable results to the state of the art. Impressively, POEM is around 20 times faster than Gabor-based methods.

1. Introduction

Face recognition is a highly active area of research with a wide variety of real-world applications, and in recent years a clearly defined face-recognition pipeline has emerged. This face-recognition pipeline has four main stages: *detection*, or finding where the faces are in an image; *alignment* ensuring the detected face(s) line up with a target face or a model; *representation* or feature description transforms the aligned faces into some representation emphasising certain aspects; and *classification*, which determines whether a certain face matches a target face or a model. The main contribution of this paper is in the *representation* stage, and thus our starting point is detected and aligned faces.

Within the domain of face recognition, facial feature description is a key aspect. If inadequate feature descriptors are used, even the best classifier will fail to achieve good recognition results. Good face representations are those which minimise intra-person dissimilarities (i.e., the differences between face images of the same person due to variations of illumination, pose) whilst enlarging the margin between different people. This is a critical issue, as variations of pose, illumination, age, and expression can be larger than variations of identity in the original face images. For real-world face recognition systems, we believe that a good representation should also be both fast and compact: if one is testing a probe face against a large database of desirable (or undesirable) target faces, the extraction and storage of the face representation has to be fast enough for any results to be quickly delivered to the end user. Thus, face representations have to satisfy a number of challenging criteria, including *robustness*, *distinctiveness*, *compactness*, and *speed*. In this paper we present a novel feature set named patterns of oriented edge magnitudes (POEM) for robust face recognition, a descriptor which we argue addresses these criteria. A preliminary version of this article has appeared in [1].

Once a representation has been decided upon, the next stage is to use this representation in classification. This usually involves pairing the descriptor with some kind of classifier—from simple "nearest neighbour" or distance in histogram space methods to more sophisticated techniques such as support vector machines (SVMs) [2] or hidden Markov models (HMMs) [3,4]. In this paper we couple our descriptor with a number of classifiers in order to demonstrate the strength of the descriptor itself.

The overall aim of our work is the production of face recognition systems for surveillance applications, which imposes extra constraints upon the design of both the descriptor and classification stage. Specifically:

1. *The system should be robust to variations in lighting, pose, image quality, and age.*
2. *Execution time should be fast*: The face representation must be fast to extract, and the classifier must also be fast.
3. *Extraction and matching must be automatic*: There should be no need for hand-labelling (of, for example, facial features).
4. *The solution must be scalable*: In particular, adding new faces to the database should not require re-training.

* Corresponding author at: GIPSA-lab, Grenoble Institute of Technology, BP 46, 38402 Saint Martin d'Hères cedex, France. Tel.: +33 476827134.
E-mail address: Ngoc-Son.Vu@gipsa-lab.grenoble-inp.fr (N.-S. Vu).

0031-3203/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2011.12.021

Fig. 1. Examples of images used in our tests: the first three columns are matching pairs of FERET images, and columns 4–6 are matching pairs from the LFW dataset. Note the wide range of qualities and unconstrained pose in the LFW images. Images have been converted to greyscale and cropped for display.

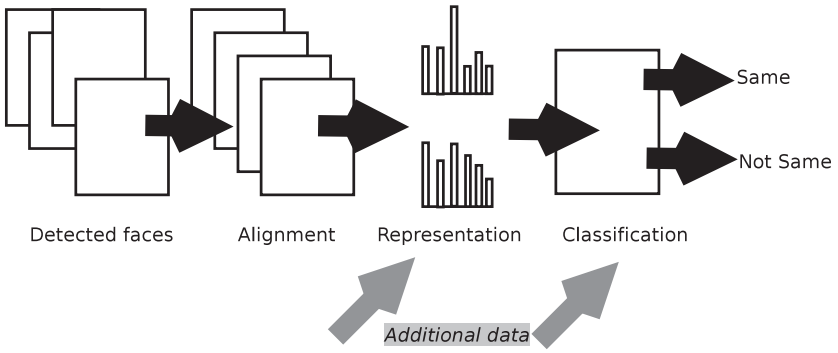

Fig. 2. The face-verification pipeline. Some methods – we call these *extended* methods – require additional data, such as faces with hand-labelled regions, or sets of "negative faces".

5. *The system should not rely on external data*: Systems which require a large set of "negative faces" (that is, faces which are guaranteed to not appear as a target) are impractical for surveillance purposes. We can never be 100% sure that a particular person will not appear in the surveillance domain.[1]
6. *The system should be able to work with just one gallery image*: Within a surveillance context, sometimes very limited information is available about targets. Therefore those systems which train models for each individual based upon multiple faces, such as [4] or [7], may not be feasible.

By combining our proposed face descriptor with simple and fast classifiers, we present results on two popular and challenging face recognition datasets: the FERET dataset which features a large number of high-quality face images; and the LFW (labelled faces in the wild) dataset, containing lower-quality face images captured "in the wild" (unconstrained images of celebrities downloaded from the Internet with great variation in pose, subject age, image size, image quality, and lighting). We present preliminary experiments and details of descriptor parameter choice using the FERET database, in which we use threshold on distance in descriptor space to determine facial identity. We also present recognition results on FERET showing that our descriptor outperforms many considered descriptors. The LFW dataset is more challenging, and therefore we have carried out more in-depth experiments using this set. Fig. 1 shows some sample images from each dataset, from which it is clear that the LFW faces are a more difficult test set. Within the LFW dataset and methodology, the question becomes a pair matching task: given a pair of face images, the classifier has to output *same* or *different*. We compare several classification methodologies: a simple threshold on histogram distance between descriptors; per-patch Gaussian mixture models (GMMs) modelling the likelihood of descriptors coming from the same person or different people; and support vector machines modelling image-level similarity between representations. We show that the POEM descriptor performs very well in this context. Perhaps more importantly, when compared with other systems, on both datasets, the POEM descriptor leads to a system of much lower complexity in both terms of computational time and storage requirements.

We discuss related work in Section 2 and describe the POEM descriptor in detail in Section 3. Experimental results are presented in Section 4 and conclusions are given in Section 5.

2. Related work

The face verification pipeline is illustrated in Fig. 2, and typically consists of alignment, then representation in some kind of feature space (for example, by using local descriptors to create a histogram of oriented edge values), and then classification as *same face* or *different faces* (for example, using an SVM). In our consideration of related work, we make a distinction between systems which follow this pipeline strictly using a pair matching protocol, and those which use additional information when making this decision. This additional input ranges from systems which use human annotation such as [8] to systems which incorporate additional negative examples such

[1] We make a distinction here between systems that use a separate training/testing split to build models of face distribution such as [5], and those that require a set of "negative faces" which are guaranteed to not overlap with the set of target faces under consideration such as [6].

Please cite this article as: N.-S. Vu, et al., Face recognition using the POEM descriptor, Pattern Recognition (2012), doi:10.1016/j.patcog.2011.12.021

as [6]. Our discussion here of prior work is therefore divided into two subsections: an overview of descriptors for face recognition; and a detailed look at methods applied on the LFW dataset. The LFW section is further subdivided into a consideration of those methods learned from training set of labelled image pairs (same/not same) and tested in a pair decision (same/not same) context, and those which use additional information (such as identity, additional databases, or multiple images of each face) and which we call *extended* methods.

It is worth noting that alignment is a major topic within the face recognition domain, and that the performance of alignment systems can greatly affect results (see, for example [9]). When comparing results this can make it hard to disentangle the contribution of alignment from the contribution of a descriptor or a classifier. Alignment methods vary from generic *image alignment* techniques such as congealing, a method which aims to minimise the entropy of a stack of images [10] to face specific techniques which either model the face as a whole, or detect specific facial features (eyes, nose, mouth, etc.) [11] and then align those with some image transformation. Indeed some authors argue that the recent success of "face averaging" techniques such as that of Jenkins and Burton [7] is due to the amelioration of alignment errors between target face and model [12,13].

It is also worth noting that the face recognition problem is intricately linked to the databases used to test the recognition. Historically, algorithms have been developed using test sets such as FERET involving high quality images taken under controlled conditions. However, recently databases such as LFW have come to prominence, gathered from the Internet using a standard face detector, e.g., [14]. These latter collections of images are more challenging to vision researchers as they include greater variation in terms of lighting, pose, age, and indeed image quality.

2.1. Face descriptors

There is an extensive literature on local descriptors and face recognition. Since this paper focuses on face descriptors in the context of face recognition and verification, we refer readers to [15] for an in-depth survey on the general use of local descriptors, and [16] for the broader face recognition literature. It maybe worth noting that there are two ways of using local descriptors: (i) approaches using only feature points—therefore a feature detection phase needs being carried out and (ii) approaches using all points (like the method proposed in this paper).

There are many representational approaches used within the domain of face recognition, including subspace based holistic features and local appearance features. Whilst some success has been demonstrated using global features (e.g., face recognition using whole-face pyramids of gradient orientations combined with an SVM [2], or the GradientFaces of [17]), it appears that traditional techniques such as Eigenfaces [18] which model the face as a whole are more sensitive to alignment and pose variation than local descriptor based methods, and therefore do not perform well on modern unconstrained face databases such as LFW. Heisele et al. [19] compare local and global approaches and observe that local systems outperform global systems for recognition rates larger than 60%.

One of the most successful local face representations are Gabor features [20–22] (or variants, such as truncated Gabor filters or Gabor "Jets"). Gabor features, which are spatially localised and selective to spatial orientations and scales, are analogous to the receptive fields of simple cells in the mammalian visual cortex [20]. Due to their robustness to local distortions, Gabor features have been successfully applied to face recognition. Indeed the FERET evaluation and the FRGC2004 contests have seen top performance from methods based upon Gabor features. Gabor features are typically calculated by convolving images with a family of Gabor kernels at different scales and orientations, which computationally speaking is a costly stage. Despite recent attempts at speeding this process up (e.g., the simplified Gabor wavelets of [23]) the process of extracting these features is still prohibitive for real-time applications.

More recently, the spatial histogram model local binary patterns (LBP) has been proposed to represent visual objects, and successfully applied to both texture analysis [24], human detection [25] and face recognition [26]. LBP is basically a fine-scale descriptor that captures small texture details, in contrast to Gabor features which encode facial shape and appearance over a range of scales. Using LBP, Ahonen et al. [26] have reported impressive results on the FERET database.

Although widely and successfully used in many types of computer vision application, such as object detection and recognition (e.g., [27]), the SIFT [28] and HOG [29] algorithms have not seen a great deal of use in face recognition. This is due to the fact that their strength lies in their ability to capture *generalities* in edge or local shape information (for example, detecting faces, people or heads) but not in the task of finding detailed local structure. Bicego et al. [30] evaluated the use of SIFT for face authentication and achieved good results on BANCA databases,[2] but no further results (on bigger databases) have been presented. In [31], HOG gives lower recognition rates than LBP or Gabor features on the FERET database.

A combination approach was introduced by Zhang et al. [22] extending LBP to LGBP (local Gabor binary pattern) by introducing multi-orientation and multi-scale Gabor filtering before using LBP. This additional stage greatly improves performance when compared with pure LBP. In a similar vein, they further proposed HGPP (histogram of Gabor phase patterns) [21] combining the spatial histogram and the Gabor phase information encoding scheme. Another fusion method proposed by Tan and Triggs [32] uses kernel PCA to combine LBP and Gabor features, and also presents impressive results on FERET, however this method suffers from the fact that the addition of a single new face requires retraining of the entire classifier. Zou et al. in [33] compare LBP, Gabor and PCA features using a subset of face regions (eyes, nose, mouth) extracted through manual labelling, and show that in this context Gabor features perform best on the FERET and AR datasets. Features are combined using the *Borda Count* [33] method. However, as mentioned earlier, Gabor wavelet based methods (including HGPP and LGBP) are computationally intensive, and are therefore impractical for real-time applications.

Boosting local features have been applied to the face-recognition problem by Jones and Viola in [34] (known as the MERL face recognition system). This achieves strong results on the FERET Fa and Fb sets (sets with expression variation but no age variation as images are taken on the same day); results are not provided for Dup1 and Dup2 (age variation) or Fc (illumination).

2.2. Recent techniques applied on the challenging LFW set

We now turn to techniques applied on the more challenging labelled faces in the wild dataset for unconstrained face recognition. This dataset consists of 13,233 images of celebrities, split into a variety of training and testing sets in such a way that for each training/testing split the people in the training set do not appear in the testing set. In the testing phase, researchers using LFW can be certain that they have *not* seen any images of the people in question during training. Two *views* of the database are provided—View 1 (which is to be used for model selection) contains a training set of 2200 face pairs and a testing set of

[2] http://www.ee.surrey.ac.uk/CVSSP/banca/.

Please cite this article as: N.-S. Vu, et al., Face recognition using the POEM descriptor, Pattern Recognition (2012), doi:10.1016/j.patcog.2011.12.021

1000 face pairs, each containing half matching and half non-matching pairs of faces. View 2 is for performance reporting, and consists of 10 non-overlapping sets which can be used for 10-fold cross validation of algorithms and parameters developed on View 1. Whilst this is a challenging dataset it is one that human observers have been shown to perform well on—in [8] Kumar et al. show that we can get 99% accuracy. We follow the standard procedure described in [35] and report the results in the form of mean classification accuracy $\pm$ standard error computed from 10 folds of the "image-restricted view 2" portion of the LFW set.

Wolf et al. [6] provide results for four different descriptor-based techniques in addition to an *extended* one-shot learning approach which will be described in the following section. The descriptors they cover are LBP, Gabor filters, and a novel *patch based LBP* in which the similarities between neighbouring patches are used to encode local information. The same/not-same decision is made by training a linear SVM. They also present combined results which are obtained by concatenating all of the descriptors, and training an SVM on the resultant 16 element vectors. Guillaumin et al. [36] use a similar approach concatenating several different descriptors (LBP, three-and four-patch LBP and also SIFT) and then applying a metric-learning approach.

The randomised tree approach of Nowak [37] has been applied to LFW by Huang et al. [38]. This involves learning randomised trees from simple image features (SIFT and geometric features), optimising the trees to generate same/not-same distinctions. This has also been combined with the MERL classifier, also described in [38] (as these classifiers appear to capture different qualities of the face a simple average of the two outputs improves results). Pinto et al. [39–41] use "simple" features in combination with SVMs and later multiple kernel learning SVMs (MKL-SVMs) and present good results on LFW. Their feature sets include pixels, V1 like features made by truncating Gabor wavelets, and what they call V1-like+, which consist of the V1-like features concatenated with various ad hoc features such as image histograms and a scaled down version of the original image. They do not discuss computational cost or run-time, but as they use many Gabor filters (around 1000) this suggests to us that this technique would be too slow to be applicable to surveillance applications.

Multi-region probabilistic histograms (MRH) are introduced by Sanderson and Lovell [5]. This technique is inspired by "bag-of-word" models (such as [42]) in which a training set is used to cluster features into a dictionary of *visual words*, and during subsequent processing, the closest visual words are used rather than the input features themselves. In [5] the features are extracted by dividing the image into small patches using DCT decomposition. Within the training set, these features are then clustered using a Gaussian mixture model (GMM) to create a dictionary of visual words. In testing, the face image is divided into regions and each region is represented by a probabilistic histogram, representing the probability that each of the visual words is present in that region. By using fairly large regions (just nine per face) and probabilistically modelling the presence or absence of a particular visual word, this method achieves a certain amount of robustness to noise and misalignment. The authors also present a method with improved results by using a normalised distance which relies upon the existence of a cohort of *negative faces*; according to our categorisation this variant is an extended method.

Also related to the "bag-of-word" technique is the work of Cao et al. [43]; in this technique pixel-level sampling in ring patterns is converted into a descriptor by first clustering using PCA-tree techniques, then dimensionality is reduced with a joint PCA-normalisation step. The authors argue that this method learns descriptors (they call the descriptor the LE descriptor) optimised for face recognition, and their results are very impressive. They couple this descriptor with a pose-estimation technique and facial feature-level matching and gain some of the best results to date on the LFW dataset. This work is in a similar vein to recent work from Winder, Brown and Hua on the learning of more general descriptors (SIFT/HOG style representations of images) as outlined in [44–46].

2.2.1. Extended methods

Wolf et al. [6] present a one-shot-learning method in which they use linear discriminant analysis to learn a same/not-same model for each of the images in a test face pair. This relies upon a large set of images of people who are definitely *not* either of the test faces; and is achieved by using one of the nine LFW training splits as negative examples. (The structure of the LFW dataset ensures there is no overlap in identity between the 10 splits.) The use of one-shot learning in this way improves the results markedly, however in real-life applications it may prove difficult to obtain a set of face images that are definitely *not* going to be in the test domain. In [47] the method is extended to include 2-shot learning and also a ranking of target faces against the background set of negative faces, further improving results. These techniques are effectively learning the relationship between the target faces and the distribution of face images in general. At the time of writing, Wolf et al.'s 2009 method [47] presented the best performance on the LFW benchmark. Taigman et al. in [48] use multiple one-shots using more than one image per individual (by making use of image labels during training—this is the LFW dataset's *image unrestricted training* protocol); they also present results using a hybrid descriptor (multiple patch based LBP with SIFT) and one-shot learning in a pair-matching context.

Kumar et al. [8] also present excellent results on the LFW dataset, using *attribute* and *simile* classifiers. Attribute classifiers are learned from a labelled dataset of faces—not merely labelled with the name of the person (as in LFW), but labelled with attributes such as "bags under eyes" or "Asian". This stage required the manual annotation of over 65,000 attributes using the Amazon Mechanical Turk system. Simile classifiers do not have this costly labelling stage but are learned from automatically extracted facial features (mouth, nose, eyes, etc.). Each simile classifier is trained on a particular region with positive examples coming from a set of images of the same person, and negative examples coming from the same face region but from different people. Simile classifiers capture the intuitive idea that whilst we might not be able to describe facial features, we can say whether they're similar ("*He's got eyes like Brad Pitt*", for example). Importantly, the training set for the simile classifiers cannot be in the set of images to be tested, and so this technique relies on a large external database of labelled faces (60 reference people, up to 600 positive face images per reference person, and ten times as many negative images). An SVM is used on the output of the attribute or simile classifiers for the pairs of images in the test set, in order to determine whether the images are from the same person or different people.

3. The POEM descriptor

We propose applying the LBP-based structure on oriented edge magnitudes to build a novel descriptor: patterns of oriented edge magnitudes (POEM). In order to calculate the POEM for one pixel, the intensity values in the calculation of traditional LBP are replaced by gradient magnitudes, calculated by accumulating a local histogram of gradient directions over all pixels of a spatial patch (*cell*). Additionally, these calculations are done across different orientations. We use the terms *cell* and *block*, as in [29], but with a slightly different meaning. Cells (large squares

Please cite this article as: N.-S. Vu, et al., Face recognition using the POEM descriptor, Pattern Recognition (2012), doi:10.1016/j.patcog.2011.12.021

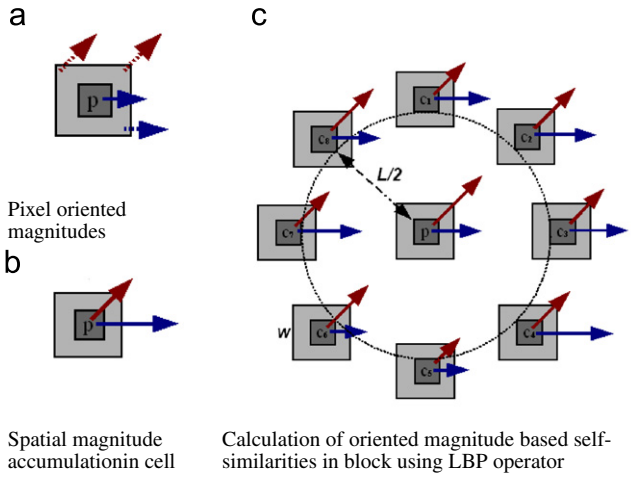

Fig. 3. Main steps of POEM feature extraction. (a) Pixel oriented magnitudes, (b) spatial magnitude accumulation in cell and (c) calculation of oriented magnitude based self-similarities in block using LBP operator.

in Fig. 3) refer to spatial regions around the current pixel where a histogram of orientation is accumulated and assigned to the cell's central pixel. Blocks (circular in Fig. 3) refer to more extended spatial regions, upon which the LBP operator is applied. Note that our use of oriented magnitudes is also different from that in [29] where HOG is computed in dense grids and then is used as the representation of cell. On the contrary, in POEM, *for each pixel*, a local histogram of gradient over all pixels of cell, centred on the considered pixel, is used as the *representation of that pixel*.

Many features have seen increasing use over the past decade [28,29,49], the basic idea being to characterise local object appearance and shape by the distribution of local intensity gradients or edge directions. The first step of our algorithm is inspired by this idea (i.e., characterising object by the distribution of edge directions) and in the second step we further apply the idea of a self-similarity calculation from LBP-based structures on these distributions since we believe that combining both the edge/local shape information and the relationship between this information in neighbouring cells can better characterise object appearance. As can be seen in Fig. 3, once the gradient image is computed, the next two steps are assigning the cell's accumulated magnitudes to its central pixel, and then calculating the block self-similarities based on the accumulated gradient magnitudes by applying the LBP operator.

3.1. The POEM descriptor in detail

The first step in extracting the POEM descriptor is the computation of the gradient image. The gradient orientation of each pixel is then evenly discretised over $0-\pi$ (*unsigned* representation) or $0-2\pi$ (*signed* representation). Thus, at each pixel, the gradient is a 2D vector with its original magnitude and its discretised direction (the continuous arrow emitting from pixel **p** in Fig. 3a).

The second step is to incorporate gradient information from neighbouring pixels (the discontinuous arrows in Fig. 3a) by computing a local histogram of gradient orientations over all cell pixels. At each pixel, the feature is now a vector of m values where m is the number of discretised orientations (i.e., number of bins). Vote weights of each pixel's contribution can either be the gradient magnitude itself, or some function of the magnitude: we use the gradient magnitude at each pixel, as in HOG [29]. To increase the importance of the central pixel, a weighted window can be used, such as a Gaussian filter or a binomial kernel,

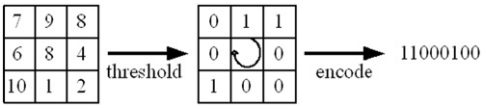

Fig. 4. LBP operator.

however we find this does not improve the discriminative power of the descriptor.

Finally, we encode the accumulated magnitudes using the LBP operator within a block. The original LBP operator labels the pixels of an image by thresholding the 3×3 neighbourhood surrounding the pixel with the intensity value of central pixel, and considering the sequence of eight result bits as a binary number (Fig. 4). Only uniform patterns, which are those binary patterns that have at most two transitions from 0 to 1, are typically used. This accelerates the method and decreases sensitivity to noise.

We apply this procedure on the accumulated gradient magnitudes and across different directions to build the POEM. Firstly, at the pixel position p, a POEM descriptor is calculated for each discretised direction θ_i:

$$POEM^{\theta_i}_{L,w,n}(p) = \sum_{j=1}^{n} f(s(m^{\theta_i}_p, m^{\theta_i}_{c_j}))2^j \quad (1)$$

where m_p, m_{c_j} are the accumulated gradient magnitudes of central and surrounding pixels p, c_j; $s(.,.)$ is the similarity function (e.g., the difference of two gradient magnitudes); L and w refer to the size of blocks and cells, respectively; n, set to 8 by default in this paper, is the number of pixels surrounding the considered pixel p; and f is defined as

$$f(x) = \begin{cases} 1 & \text{if } x \geq \tau \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where the value τ is slightly larger than zero to provide some stability in uniform regions, as in [6]. The final POEM descriptor at each pixel is the concatenation of these unidirectional POEMs at each of our m orientations:

$$POEM_{L,w,n}(p) = \{POEM^{\theta_1}, \ldots, POEM^{\theta_m}\} \quad (3)$$

3.2. Properties of POEM

POEM characterises not only local object appearance and shape, but also the relationships between this information in neighbouring regions. It has the following properties:

- POEM is an oriented feature. Since the number of discretised directions can be varied, POEM has the ability to capture image information in any direction, and is adaptable for object representation with different levels of orientation accuracy.
- Computed at different scales of cells and blocks, POEM is also a spatial multi-resolution feature. This enables it to capture both local information and global structure.
- Using gradient magnitudes instead of the pixel intensity values makes POEM more robust to lighting variations than intensity-based methods such as LBP.
- The computational cost of extracting the POEM descriptor is very low when compared with those methods based upon Gabor filters. Comparing unoptimised Matlab code, the complete extraction of a POEM descriptor is approximately 23 times faster than just the Gabor feature extraction stage of many other descriptor methods. The question of computational cost is considered in more depth in Section 3.4.

Please cite this article as: N.-S. Vu, et al., Face recognition using the POEM descriptor, Pattern Recognition (2012), doi:10.1016/j.patcog.2011.12.021

- Similarly, the data storage requirements of the POEM descriptor are low: using common settings each block can be represented as a 59×3 histogram, with 100 of these histograms per face.

Patch-based or multi-block LBP [6] considers relationships between regions in a similar way to the POEM descriptor. However, the use of gradients at multiple orientations rather than pixel magnitudes gives us greater descriptive power and robustness to lighting variations.

3.3. Parameter evaluation

In this section, we study how the parameters of the POEM descriptor influence final performance. Parameters to be chosen include the number m and type (unsigned or signed) of orientations, the cell size ($w \times w$), and block size ($L \times L$). Concerning cell/block geometry, two main geometries exist: rectangular and circular. In this paper we use circular blocks including bilinear interpolation for the values since these provide the relation between equidistant neighbouring cells as in [26]. Square cells are used, meaning that pixel information is calculated using its neighbourhood in a square patch.

The experiments described in this section were conducted on the FERET face database, following the standard evaluation protocol: Fa containing 1196 frontal images of 1196 subjects is used as Gallery, while Fb (1195 images of expression variations), Fc (194 images taking under different illumination conditions), Dup I (722 images taken later in time) and Dup II (234 images), are the Probe sets. Dup II consists of images that were taken at least one year after the corresponding gallery images. The classifier used to determine recognition rates is a simple nearest-neighbour classifier using histogram distance in feature space between pairs of concatenated POEM descriptors. We call the face representation created through concatenation of POEM descriptors POEM-HS (HS standing for histogram sequence).

As alignment, thanks to the available coordinates of eyes, all FERET facial images are geometrically aligned in such a way that centres of the two eyes are at fixed positions and images are resized to 110×110 pixels.

For the first classifier we consider, we use a similar procedure to Ahonen et al. [26] except that each pixel is characterised with the POEM descriptor instead of an LBP code (see Fig. 5). Whilst a very simple "classifier", we believe that these results are important as they show the strength of the POEM descriptor alone.

In practice, the oriented edge magnitude image (oriented EMI) is first calculated from the original input image (Section 3.1) and divided into m uni-oriented EMIs through gradient orientations of pixels. Note that the pixel value in uni-oriented EMIs is gradient magnitude. For every pixel on uni-oriented EMIs, its value is then replaced by the sum of all values on the cell, centred on the current pixel. These calculations are very fast (taking advantage of the integral image [50]). The resulting images are referred to as accumulated EMIs (AEMIs). LBP operators are applied on these AEMIs to obtain the POEM images (Fig. 5). The POEM images are spatially divided into multiple non-overlapping regions (10×10 regions in this paper), and histograms are extracted from each region. Finally, all the histograms estimated from all regions of all POEM images are concatenated into a single histogram sequence (HS) to represent the given face. This representation is called POEM-HS throughout the rest of this article.

3.3.1. Determining the number of orientation bins, and whether they should be signed or unsigned

In this experiment we consider nearly 600 cases, with recognition rates calculated for 3000+face images using different parameters: L={5,6,7,8,9,10,11}, w={3,4,5,6,7,8}, the number of discretised orientations are m={2, 3, 4, 5, 6, 7} in the case of an unsigned representation, and are doubled to m={4, 6, 8, 10, 12, 14} in the case of a signed representation. Cells can overlap, notably when blocks are smaller than cells, meaning that each pixel can contribute more than once. For each Probe set, the average rates are calculated through different numbers and types of orientation. Fig. 6 shows the recognition rates obtained on Probe sets Fb, Fc, Dup1, and Dup2.

Considering the question of using a signed or an unsigned representation, we find similar results to [29], in that including signed gradients decreases the performance of POEM-HS even when the data dimension is doubled to preserve the original orientation resolution. For face recognition, POEM-HS provides the

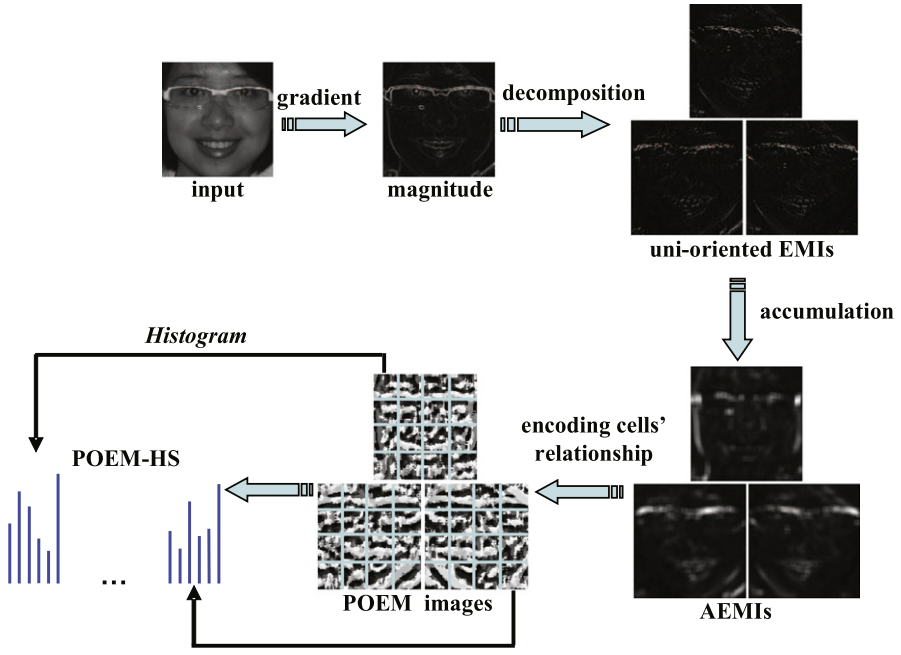

Fig. 5. Implementation of POEM histogram sequence (POEM-HS); EMIs: edge magnitude images (AEMIs: accumulated EMIs).

Please cite this article as: N.-S. Vu, et al., Face recognition using the POEM descriptor, Pattern Recognition (2012), doi:10.1016/j.patcog.2011.12.021

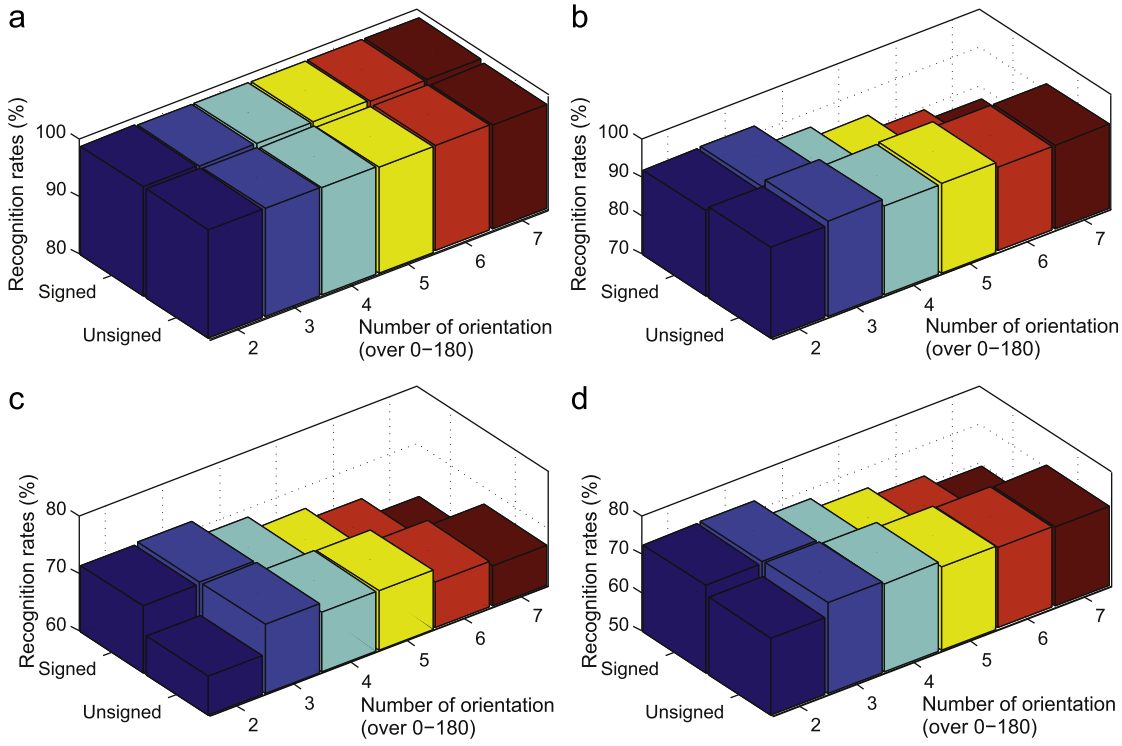

Fig. 6. Recognition rates obtained with different numbers of orientations on Probe sets Fb (a), Fc (b), Dup1 (c) and Dup2 (d). These rates are calculated by averaging recognition rates with different sizes of cell/block.

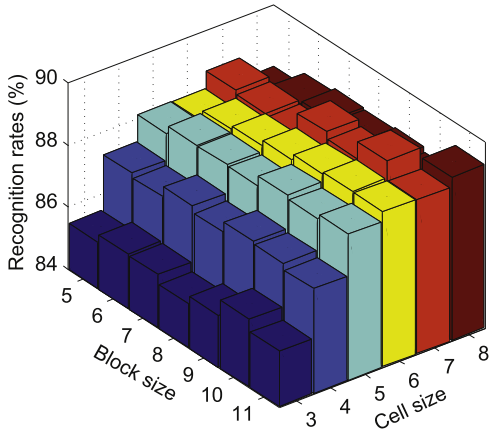

Fig. 7. Recognition rates as the cell and block sizes change.

best performance with only three unsigned bins. This should be noted as one advantage of the POEM descriptor since it contributes to the compactness of our representation. It is clear from Fig. 6(c) and (d) that using too many orientations significantly degrades the recognition rates on the age-related Dup1 and Dup2 sets. We believe this is explained by the hypothesis that increasing the number of orientation bins makes the POEM descriptor more sensitive to wrinkles appearing in faces over time.

3.3.2. Determining cell and block size

Average recognition rates of all four Probe sets are first calculated with different sizes of cells and blocks with three unsigned orientation bins. As can be seen from Fig. 7, using POEM descriptors built on 10×10 pixel blocks with a histogram of 7×7 pixel cells provides the best performance.

To verify the correctness of these parameters, we further calculate the average rates across cell sizes and across block sizes, meaning that these parameters are now considered independent. In this test, 10×10 pixel blocks and 7×7 pixel cells perform best. We have repeated this experiment using different numbers of orientation bins, and we obtain the same optimal parameters. It maybe worth highlighting that these optimal cell and block sizes are found by experiments on face images of 110×110 pixels where the distance between the two eye centres is around 50 pixels. When testing with images of around 90×90 pixels, similar optimal parameters are obtained.

In conclusion, the optimal POEM parameters for face recognition are: an unsigned representation with three direction bins built on 10×10 pixel blocks and 7×7 pixel cells.

3.4. A consideration of computational cost

In this section we compare the complexity of POEM with two of the most widely used descriptors for face recognition: LBP and Gabor wavelet based methods. Considering the pure one-LBP-operator method, POEM based face recognition requires a computational complexity which is three times higher (the calculation of LBP's integral gradient image is very fast when compared to the calculation of POEM codes and the construction of POEM-HS) but at the same time, there are remarkable improvements in recognition rates on the FERET database (+5%, +45%, +16.6% and +26.5% for the probe sets Fb, Fc, Dup1 and Dup2, respectively). And on the LFW set, our POEM method outperforms other variants of LBP, TPLBP and FPLBP. Full results on FERET can be found in Table 2 and on LFW in Table 3.

When we consider Gabor filter based descriptors, only the run time required for the convolution of the image with the family of Gabor kernels (eight orientations and five scales) is necessary. From Table 1 showing the feature extraction time at different image resolutions, we see that the computation of the whole POEM-HS description of a face is at least 20 times faster than that of just this first step of Gabor feature extraction. It maybe worth noting that throughout this paper, we report the results when dividing face image into 100 regions (PH-100) but PH-64 with

Please cite this article as: N.-S. Vu, et al., Face recognition using the POEM descriptor, Pattern Recognition (2012), doi:10.1016/j.patcog.2011.12.021

lower complexity still performs very well (it provides slightly worse results than PH-100). Although this is a rough comparison using unoptimised Matlab code, it clearly shows that our approach is considerably faster than the Gabor-based algorithms.

We do not calculate here the time required to extract SIFT descriptors (this is not much used for face recognition) and compare it to POEM, but as argued in [51], SIFT is about three times slower than centre-symmetric LBP (CSLBP), a variant of LBP. Thus it seems that POEM and SIFT have similar time complexity. However, for face recognition, POEM clearly outperforms SIFT, representing about 20% reduction in classification error on the LFW set.

Considering data storage requirements, for a single face, the size of a complete POEM-HS face description is *103 and 232 times* smaller than that of LGBP and HGPP, the best performing descriptors reported on the FERET dataset (LGBP calculates LBP on 40 convolved images and the size of patch histogram is 256). When compared to the "combined" method of Wolf et al. [47], the space complexity of POEM-HS is considerably smaller. For one patch, the size of POEM-HS is 59×3, whereas the size of the method in [47] is $59\times2+16+128$ (LBP and TPLBP are 59, and FPLBP is 16, and SIFT is 128) before we consider the size of the Gabor based descriptor which they also use. Thus, the use of POEM descriptors will accelerate both the feature extraction and classification phases.

4. Experiments and discussions

In this section we present recognition results on the FERET (Section 4.1) and LFW (Section 4.2) datasets and provide performance comparisons with the state-of-the-art descriptor based methods. We use three main classification techniques in conjunction with the POEM descriptor in order to determine whether a pair of face images represent the same person ("same") or different people ("diff"). The first of these is a simple threshold on histogram distance in feature space (POEM-HS); the second involves training a per-patch GMM for the *same* and *diff* cases; and the third involves modelling the relationship between patches using a multi-kernel support vector machine (MKL-SVM). We also present results combining the MKL-SVM and GMM techniques. For the FERET images we only present results using the first technique.

Table 1
Feature extraction time. The experiments are carried out on a computer with an Intel Core 2 Duo E6550 2.33 GHz (2 GB RAM). GC indicates the first step of Gabor-based representations, i.e., convolution with 40 Gabor kernels. PH-*p* stands for POEM-HS where *p* is the number of histograms per face image: PH-100, PH-64, and PH-16 indicate that faces are divided into 100, 64, and 16 regions, respectively. Note that these figures are calculated using unoptimised Matlab code, so these times are only suitable for rough comparisons of computational complexity.

Method	GC	PH-100	PH-64	GC	PH-64	GC	PH-16
Image size	115×115			130×180		60×60	
Time (ms)	450	22.1	14.2	840	24.2	145	3.1

4.1. Results based upon the FERET evaluation protocol

POEM-HS results in this section are based upon a simple nearest-neighbour classifier; this allows us to directly compare the performance of the POEM descriptor with others from the literature. We further employ the real-time retina filtering presented by Vu and Caplier [52] as a preprocessing step since this algorithm not only removes illumination variations but also enhances image edges, upon which our POEM is constructed. It is clear from Table 2 that the retina filter enhances the performance of POEM when used with high-quality images such as those found in FERET, especially for the probe set Fc.

We consider the FERET'97 results published in 2000 [53], results of the LBP [26], HGPP [21], LGBPHS [22], and more recent results in [31–33]. These results, to the best of our knowledge, are the state-of-the-art with respect to the FERET dataset.

As can be seen from Table 2, in comparison with the conventional LBP [26] and HOG (the performance of HOG for face recognition is reported in [31]), our POEM descriptor is much more robust to lighting, expression, and ageing, illustrated by significant improvements in recognition rates for all probe sets. While compared with LGBP [22] and HGPP [21], reported as being the best performing descriptors on the FERET database so far, POEM provides comparable performance for the probe sets Fb and Fc. When we consider the more challenging probe sets Dup1 and Dup2, POEM outperforms LGBP and is comparable to HGPP.

As mentioned in Section 2, the results of [33,32] are only suitable for very limited reference since they are obtained by using methods which are inappropriate for the kind of applications we aim to support. Notably, Zou et al. [33] use hand-labelled facial features to determine the regions to compare and also use large numbers of computationally expensive Gabor jets. In [32], Tan and Triggs fuse two feature sets, Gabor and LBP, but they also use a complex dimensionality reduction and classification phase, kernel PCA and DCV. Their method suffers from the disadvantage that adding a new individual to the gallery requires recalculating all existing coefficients: PCA coefficients of Gabor & LBP features, and the KDCV coefficients of the fused features.

4.2. Results on the LFW dataset

We present various classifiers in this section. The simplest, POEM-HS, is distance in histogram space, as used on the FERET database. That is, given two histogram sequences of POEM representing two face images, we use the chi-square distance

Table 2
Recognition rate comparisons with other state-of-the-art results tested with the FERET evaluation protocol.

Method	Fb	Fc	Dup1	Dup2	Note
LBP [26]	93.0	51.0	61.0	50.0	
LGBPHS [22]	94.0	97.0	68.0	53.0	
HGPP [21]	97.6	98.9	77.7	76.1	
HOG [31]	90.0	74.0	54.0	46.6	
POEM-HS	**97.6**	**96**	**77.8**	**76.5**	
Retina filter [52]+**POEM-HS**	**98.1**	**99**	**79.6**	**79.1**	
Results of [33]	*99.5*	*99.5*	*85*	*79.5*	Hand-labelling of facial features
Results of [32]	*98*	*98*	*90*	*85*	Gabor features and requires retraining

Please cite this article as: N.-S. Vu, et al., Face recognition using the POEM descriptor, Pattern Recognition (2012), doi:10.1016/j.patcog.2011.12.021

between histograms to measure the similarity between two images. Using the 10-fold training/testing splits provided with the LFW database, we calculate the distance between POEM-HS representations for each pair in the training set; determine the threshold which best separates the classes of "same" and "different" faces, and then apply this technique and threshold to the testing set. POEM-HS-flip involves flipping the second of the face images along the vertical axis, and comparing histogram distance to both original and flipped. The motivation behind this is to reduce the effect of pose variation—given the unconstrained nature of the LFW dataset (see Fig. 1), some images contain faces looking left and others contain faces looking right. The simple "flip" step enables us to reduce errors caused by this. We also use GMMs and MKL-SVMs as an additional classification step: the use of these systems is outlined in the following two subsections.

4.2.1. Modelling same and diff using per-patch GMMs

This classification methodology uses a pair of GMMs per-patch to model the distribution of histogram distances. For each patch, one GMM corresponds to the histogram distances between a pair of patches coming from the *same* individual and the other models those coming from *diff* case, giving us $2 \times n$ GMMs where n is the number of patches. Using the LFW View 1 training/testing split, we have determined that a GMM with 2 Gaussian components is sufficient to model the variation in histogram distances; and during performance analysis each of these GMMs is trained on 2700 pairs (for each of the same and diff) in the training split of View 2. Each pair of GMMs taken together give the probability of a particular similarity score sim_i between a pair of patches coming from the same person s and from a different person d.

Over G Gaussians, Eq. (4) details the determination of the probability that a pair of patches comes from images of the same person (ω is the weight of the Gaussian, σ is the standard deviation; μ is the mean; and sim_i is the similarity measure between the two histograms representing the ith patches).

$$P(sim_i|s) = \sum_{g=1}^{G} \omega_{gi}^{s} \bullet P^{s}(sim_i, \mu_{gi}, \sigma_{gi}) \tag{4}$$

We can calculate the probability $P(sim_i|d)$ that a set of patch pairs come from different faces in exactly the same fashion. For classification, using Bayes rule, we have

$$P(s|sim_i) = \frac{P(sim_i|s)P(s)}{P(sim_i|s)P(s) + P(sim_i|d)P(d)} \tag{5}$$

Eq. (5) (assuming that the class priors $P(s)$ and $P(d)$ are equal and 0.5) provides us with a means of calculating the probability that a particular pair of patches come from a pair of images of the same face. To move from patch-based similarity probabilities to a similarity measure between a pair of images I_1 and I_2 is a simple matter of summing over i:

$$P_{same}(I_1, I_2) = \frac{\sum_{i=1}^{n} P(s|sim_i)}{n} \tag{6}$$

Using a model such as the GMM outlined above allows us to capture the intuitive idea that for a pair of faces, we expect different patches to have different distributions of similarity scores for the inter-person and intra-person cases.

4.2.2. Using MKL-SVM classification with POEM-HS descriptor

The final classifier we consider is a support vector machine with multiple kernels (MKL-SVM) [54,55]. MKL-SVM is an

Table 3
Recognition results of (as far as we are aware) all published and forthcoming methods on LFW set, image-restricted training, View 2. Bold-face indicates methods introduced in this article, *italics* and the letter E: in the description column indicates extended methods. Note that the methods to date which outperform our method are nearly all either *extended* methods, or are methods which use multiple feature sets or large numbers of Gabor features, with the exception of the paper [43] using learned features.

Reference	Perf.	Std. Err.	Notes
1. Pixels and linear SVM [39]	0.5995	0.0064	
2. Eigenfaces, original [18]	0.6002	0.0079	
3. Gabor (Euclidean) [6]	0.6293	0.0047	
4. V1-like linear SVM [39]	0.6421	0.0069	
5. LBP (Hellinger) [6]	0.6782	0.0063	
6. V1-like+ linear SVM [39]	0.6808	0.0044	
7. Pixels/MKL [41]	0.6822	0.0041	
8. FPLBP (Euclidean) [6]	0.6865	0.0056	
9. TPLBP (Hellinger) [6]	0.6890	0.0040	
10. 3×3 Multi-region histograms [5]	0.7038	0.0048	
11. MERL [38]	0.7052	0.0060	
12. Combined lbp-gabor-tplbp-fplbp [6]	0.7062	0.0057	
13. Nowak, original [38]	0.7245	0.0040	
14. 3×3 Multi-Region Histograms (normalised) [5]	0.7295	0.0055	E: needs cohort of negative faces
15. Nowak, funneled [38]	0.7393	0.0049	
16. POEM-HS	0.7398	0.0062	
17. POEM-HS-flip	0.7542	0.0071	
18. POEM-HS-flip-GMM	0.7602	0.0059	
19. MERL+Nowak, funneled [38]	0.7618	0.0058	Combination of 11. and 15
20. One-shot learning on lpb-gabor-tplbp-fplbp [6]	0.7653	0.0054	E: Uses 1 shot learning
21. POEM-HS-flip-MKL-SVM	0.7748	0.0057	
22. POEM-HS-flip-GMM-MKL-SVM	0.7767	0.0054	
23. Hybrid descriptor-based, funneled [6]	0.7847	0.0051	E: Combination of 20. and 12
24. LDML, funneled [36]	0.7927	0.0060	Uses methods 5. 8. 9. and SIFT features
25. V1-like+/MKL [41]	0.7935	0.0055	Around 1000 Gabor filters+ad hoc features
26. Single LE+holistic [43]	0.8122	0.0053	Learned descriptors
27. Attribute classifiers [8]	0.8362	0.0158	E: big training set (65,000 hand labelled)
28. Hybrid, aligned [48]	0.8398	0.0035	E: 1 shot learning
29. Simile classifiers [8]	0.8414	0.0131	E: big training set (hundreds of images of 60 additional reference people)
30. Multiple LE+comp [43]	0.8445	0.0046	Learned descriptors
31. Attribute and Simile classifiers [8]	0.8529	0.0123	E: Combination of 27. and 29
32. Combined b/g samples based methods, aligned [48]	0.8683	0.0034	E: uses 1 shot/2 shot learning+ranked distances

Please cite this article as: N.-S. Vu, et al., Face recognition using the POEM descriptor, Pattern Recognition (2012), doi:10.1016/j.patcog.2011.12.021

extension to the standard SVM in which multiple kernels can be trained at the same time with different bandwidths, integrated using a linear combination of kernel weights which is also learned during training. In our experiments, we use 10 RBF kernels with different widths. The inputs to this SVM are the histogram distances between POEM descriptors for each patch.

In our experiments, the LFW grey images aligned automatically by Wolf et al. [47] are used and cropped to 100×116 pixels around their centre. Because of the poor quality of the images in the LFW dataset, retina filtering does not improve recognition results. With low quality images, the retina filter enhances image contours and removes illumination variations but also enhances image artifacts (such as those arising from compression). Thus we do not employ retina filtering (or any other additional preprocessing techniques) with the LFW dataset.

It is clear from Table 3 that the POEM-HS method on its own outperforms all other competing descriptors when considered singly: LBP, TPLBP, FPLBP, Gabor filters and SIFT. Even those methods which combine many descriptors ("*everything bar the kitchen sink*" methods) do not perform as well as the simple POEM-HS method. When compared with these descriptors, the POEM based method represents around 20% reduction in classification error. When combined with the GMM and MKL-SVM we achieve 77.67% accuracy, which for a fast and easy-to-compute method is very good indeed on such a challenging dataset.

Of the methods which outperform our techniques on the LFW dataset, the vast majority are what we call *extended* methods, requiring additional training sets or hand-labelling. Considering in more detail those methods which are not extended: Pinto et al.'s V1-like+/MKL method [41] uses over 1000 Gabor features alongside ad-hoc features, and is therefore unsuitable for any real-time applications; Guillaumin et al. use four different types of feature and a learned distance metric; and Cao et al. [43] use learned descriptors. We believe that the work of Cao et al. is of particular interest and learning descriptors which incorporate elements of the POEM framework is a direction for future work.

5. Conclusions

By applying the LBP operator on accumulated magnitudes across different directions, a novel robust descriptor for object representation named patterns of oriented edge magnitudes (POEM) with several advantages is presented. Studying the influence of some descriptor parameters, we find that unsigned three bin representations, built on 10×10 pixel blocks and 7×7 pixel cells, provide excellent performance for face recognition. Our experimental results obtained on the challenging LFW dataset show that the proposed method compares with the state-of-the-art for descriptor-based methods.

Additionally, we have shown that the computational cost of extraction of this new descriptor is considerably less than many of its competitors. Thus we argue that the POEM descriptor is the first to allow high performance real-time face recognition. Low complexity descriptors provide worse results; whilst representations based upon multiple feature types or upon very large numbers of Gabor features can achieve similar performance but are too slow for real-time systems.

Acknowledgements

We would like to thank the anonymous reviewers for their helpful comments.

References

[1] N.-S. Vu, A. Caplier, Face recognition with patterns of oriented edge magnitudes, in ECCV, 2010, Available online: ⟨http://www.springerlink.com/content/k510660437327600/⟩.
[2] H. Ling, S. Soatto, N. Ramanathan, D.W. Jacobs, A study of face recognition as people age, in: Proceedings of the International Conference on Computer Vision (ICCV), IEEE, October 2007, pp. 1–8.
[3] N.-S. Vu, A. Caplier, Patch-based similarity hmms for face recognition with a single reference image, in: ICPR, 2010, pp. 1204–1207.
[4] J.T. Chien, C.P. Liao, Maximum confidence hidden Markov modeling for face recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 30 (4) (2008) 606–616.
[5] C. Sanderson, B.C. Lovell, Multi-region probabilistic histograms for robust and scalable identity inference, in: M. Tistarelli, M.S. Nixon (Eds.), ICB, Lecture Notes in Computer Science, vol. 5558, Springer, 2009, pp. 199–208.
[6] L. Wolf, T. Hassner, Y. Taigman, Descriptor based methods in the wild, in: Faces in Real-Life Images Workshop at ECCV, 2008.
[7] R. Jenkins, A.M. Burton, 100 accuracy in automatic face recognition, Science 319 (5862) (2008) 435.
[8] N. Kumar, A.C. Berg, P.N. Belhumeur, S.K. Nayar, Attribute and simile classifiers for face verification, in: Proceedings of the International Conference on Computer Vision (ICCV), 2009.
[9] S. Shan, Y. Chang, W. Gao, B. Cao, P. Yang, Curse of mis-alignment in face recognition: problem and a novel mis-alignment learning solution. in: IEEE International Conference on Automatic Face and Gesture Recognition IEEE Computer Society, Los Alamitos, CA, USA, 2004, pp. 314+.
[10] G.B. Huang, V. Jain, E. Learned-Miller, Unsupervised joint alignment of complex images, in: Proceedings of the International Conference on Computer Vision (ICCV), 2007, pp. 1–8.
[11] Y. Zhou, W. Zhang, X. Tang, H. Shum, A Bayesian mixture model for multi-view face alignment, in: Proceedings of the Computer Vision and Pattern Recognition (CVPR), IEEE Computer Society, Washington, DC, USA, 2005, pp. 741–746.
[12] X. Zhang, S. Zhao, Y. Gao, On averaging face images for recognition under pose variations, in: Proceedings of the International Conference on Pattern Recognition (ICPR), 2009.
[13] S. Zhao, X. Zhang, Y. Gao, A comparative evaluation of average face on holistic and local face recognition approaches, in: Proceedings of the International Conference on Pattern Recognition (ICPR), 2009.
[14] P. Viola, M. Jones, Robust real-time face detection. in: Proceedings of the International Conference on Computer Vision (ICCV), vol. 2, 2001.
[15] K. Mikolajczyk, C. Schmid, A performance evaluation of local descriptors, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 27 (10) (2005) 1615–1630.
[16] W. Zhao, R. Chellappa, A. Rosenfeld, P. Phillips, Face recognition: a literature survey, ACM Computing Surveys 35 (4) (2003) 399–458.
[17] T. Zhang, Y.Y. Tang, B. Fang, Z. Shang, X. Liu, Face recognition under varying illumination using gradientfaces, IEEE Transactions on Image Processing (TIP) 18 (11) (2009) 2599–2606.
[18] M.A. Turk, A.P. Pentland, Face recognition using eigenfaces, in: Proceedings of the Computer Vision and Pattern Recognition CVPR, 1991, pp. 586–591.
[19] B. Heisele, P. Ho, J. Wu, T. Poggio, Face recognition: component-based versus global approaches, Computer Vision and Image Understanding 91 (1–2) (2003) 6–21.
[20] C. Liu, H. Wechsler, Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition, IEEE Transactions on Image Processing (TIP) 11 (4) (2002) 467–476.
[21] B. Zhang, S. Shan, X. Chen, W. Gao, Histogram of gabor phase patterns (hgpp): a novel object representation approach for face recognition, IEEE Transactions on Image Processing (TIP) 16 (1) (2007) 57–68.
[22] W. Zhang, S. Shan, W. Gao, X. Chen, H. Zhang, Local gabor binary pattern histogram sequence (LGBPHS): a novel non-statistical model for face representation and recognition. in: Proceedings of the International Conference on Computer Vision (ICCV), vol. 1, 2005, pp. 786–791.
[23] W.-P. Choi, S.-H. Tse, K.-W. Wong, K.-M. Lam, Simplified gabor wavelets for human face recognition, Pattern Recognition 41 (3) (2008) 1186–1199.
[24] T. Ojala, M. Pietikäinen, T. Mäenpä"a, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 24 (7) (2002) 971–987.
[25] Y.D. Mu, S.C. Yan, Y. Liu, T. Huang, B.F. Zhou, Discriminative local binary patterns for human detection in personal album, in: Proceedings of the IEEE Computer Vision and Pattern Recognition, 2008, pp. 1–8.
[26] T. Ahonen, A. Hadid, M. Pietikainen, Face recognition with local binary patterns, in: Proceedings of the European Conference on Computer Vision (ECCV), 2004, pp. 469–481.
[27] P.F. Felzenszwalb, D.A. McAllester, D. Ramanan, A discriminatively trained, in: Proceedings of the Computer Vision and Pattern Recognition (CVPR), 2008.
[28] D.G. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision 60 (2) (2004) 91–110.
[29] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection. in: Proceedings of the Computer Vision and Pattern Recognition (CVPR), vol. 1, IEEE Computer Society, Washington, DC, USA, 2005, pp. 886–893.

Please cite this article as: N.-S. Vu, et al., Face recognition using the POEM descriptor, Pattern Recognition (2012), doi:10.1016/j.patcog.2011.12.021

[30] M. Bicego, A. Lagorio, E. Grosso, M. Tistarelli, On the use of sift features for face authentication, in: CVPR Workshop, 2006, Available online: 〈http://dx.doi.org/10.1109/CVPRW.2006.149〉.
[31] E. Meyers, L. Wolf, Using biologically inspired features for face processing, International Journal of Computer Vision 76 (1) (2008) 93–104.
[32] X. Tan, B. Triggs, Fusing gabor and lbp feature sets for kernel-based face recognition, in: Analysis and Modeling of Faces and Gestures, 2007, pp. 235–249.
[33] J. Zou, Q. Ji, G. Nagy, A comparative study of local matching approach for face recognition, IEEE Transactions on Image Processing (TIP) 16 (10) (2007) 2617–2628.
[34] M.J. Jones, P. Viola, Face Recognition Using Boosted Local Features, MERL, Technical Report, 2003.
[35] G.B. Huang, M. Ramesh, T. Berg, E. Learned-Miller, Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments, Technical Report, University of Massachusetts, Amherst, October 2007, Available online: 〈http://vis-www.cs.umass.edu/papers/lfw.pdf〉.
[36] M. Guillaumin, J. Verbeek, C. Schmid, Is that you? Metric learning approaches for face identification, in: Proceedings of the International Conference on Computer Vision (ICCV), 2009.
[37] E. Nowak, F. Jurie, Learning visual similarity measures for comparing never seen objects, in: Proceedings of the Computer Vision and Pattern Recognition (CVPR), 2007, pp. 1–8.
[38] G.B. Huang, M.J. Jones, E. Learned-Miller, LFW Results Using a Combined Nowak plus MERL Recognizer, in: Faces in Real-Life Images Workshop at ECCV, 2008.
[39] N. Pinto, J.J. di Carlo, D.D. Cox, Establishing good benchmarks and baselines for face recognition, in: Faces in Real Life Images workshop at ECCV08, 2008.
[40] N. Pinto, D.D. Cox, J.J. DiCarlo, Why is real-world visual object recognition hard? PLoS Computational Biology 4 (1) (2008) e27.
[41] N. Pinto, J.J. DiCarlo, D.D. Cox, How far can you get with a modern face recognition test set using only simple features?, in: Proceedings of the Computer Vision and Pattern Recognition (CVPR), 2009, pp. 2591–2598.
[42] E. Nowak, F. Jurie, B. Triggs, Sampling strategies for bag-of-features image classification. in: Proceedings of the European Conference on Computer Vision (ECCV), vol. 3954, Springer, Berlin, Heidelberg, 2006, pp. 490–503.
[43] Z. Cao, Q. Yin, X. Tang, J. Sun, Face recognition with learning-based descriptor, in: Proceedings of the Computer Vision and Pattern Recognition, 2010.
[44] S. Winder, G. Hua, M. Brown, Picking the best DAISY, in: Proceedings of the Computer Vision and Pattern Recognition (CVPR), 2009, pp. 178–185.
[45] S.A. Winder, M. Brown, Learning local image descriptors, in: Proceedings of the Computer Vision and Pattern Recognition (CVPR), 2007, pp. 1–8.
[46] M. Brown, G. Hua, S. Winder, Discriminative learning of local image descriptors, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 53 (1) (2010) 43–57.
[47] L. Wolf, T. Hassner, Y. Taigman, Similarity scores based on background samples, in: Asian Conference on Computer Vision (ACCV), 2009.
[48] Y. Taigman, L. Wolf, T. Hassner, Multiple one-shots for utilizing class label information, in: British Machine Vision Conference, 2009.
[49] Y. Ke, R. Sukthankar, PCA-SIFT: a more distinctive representation for local image descriptors. in: Proceedings of the Computer Vision and Pattern Recognition CVPR, vol. 2, 2004, pp. 506–513.
[50] P. Viola, M.J. Jones, Robust real-time face detection, International Journal of Computer Vision 57 (2) (2004) 137–154.
[51] M. Heikkila, M. Pietikainen, C. Schmid, Description of interest regions with local binary patterns, Pattern Recognition 42 (3) (2009) 425–436.
[52] N.-S. Vu, A. Caplier, Illumination-robust face recognition using retina modeling, in: Proceedings of the International Conference on Image Processing (ICIP), IEEE, 2009, pp. 3289–3292.
[53] J.P. Phillips, H. Moon, S.A. Rizvi, P.J. Rauss, The FERET evaluation methodology for face-recognition algorithms, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 22 (10) (2000) 1090–1104.
[54] S. Sonnenburg, G. Ratsch, C. Schafer, B. Scholkopf, Large scale multiple kernel learning, Journal of Machine Learning Research 7 (2006) 1531–1565.
[55] M. Kloft, U. Brefeld, S. Sonnenburg, P. Laskov, K.-R. Müller, A. Zien, Efficient and accurate LP-Norm multiple kernel learning. in: Advances in Neural Information Processing Systems, vol. 22, 2009, pp. 997–1005.

Ngoc-Son Vu is currently a post-doctoral researcher at Grenoble Institute of Technology, France from which he received his Ph.D. degree in image processing and computer vision in 2010. Previously, he received his M.Sc. degree in computer science from Institut de la Francophonie pour l'Informatique, Hanoi, Vietnam in 2007. His research interests include image processing, pattern recognition, and human face analysis.

Hannah M. Dee is a lecturer in computer science at Aberystwyth University in the UK. Before this, she carried out post-doctoral work at the Grenoble Institute of Technology (INPG), the University of Leeds, and Kingston University. She received her Ph.D. in computer vision from the University of Leeds in 2006. Her research areas are computer vision for understanding human behaviour, including behaviour modelling, scene understanding, tracking, and person identification. She is deputy chair of BCSWomen, the British Computer Society's group for women in technology, and is active in encouraging women and girls to consider careers in computer science.

Alice Caplier graduated from the École Nationale Supérieure des Ingénieurs Électriciens de Grenoble (ENSIEG) of Grenoble INP, France, in 1991. She obtained her Master's degree in Signal, Image, Speech Processing and Telecommunications in 1992 and her Ph.D. from Grenoble INP in 1995. Since 1997 she has been teaching at the Phelma Engineering School and is professor at the GIPSA-lab in the Image and Signal Department in Grenoble. Her interest is on human motion analysis and interpretation and on facial biometry.

Please cite this article as: N.-S. Vu, et al., Face recognition using the POEM descriptor, Pattern Recognition (2012), doi:10.1016/j.patcog.2011.12.021