Pattern Recognition Letters 48 (2014) 49–59

Contents lists available at ScienceDirect

Pattern Recognition Letters

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

Fast pose invariant face recognition using super coupled multiresolution Markov Random Fields on a GPU ☆

Shervin Rahimzadeh Arashloo [a,*], Josef Kittler [b]

[a] Department of Electrical Engineering, Faculty of Engineering, Urmia University, Urmia, West Azarbaijan, Iran
[b] CVSSP, University of Surrey, Guildford, Surrey, UK

ARTICLE INFO

Article history:
Available online 14 June 2014

Keywords:
Multiresolution MRFs
Renormalisation group transform
Super coupling transform
Image matching
Unconstrained-pose face recognition

ABSTRACT

We discuss the problem of pose invariant face recognition using a Markov Random Field (MRF) model. MRF image to image matching has been shown to be very promising in earlier studies (Arashloo and Kittler, 2011) [4]. Its demanding computational complexity has been addressed in Arashloo et al. (2011) [6] by means of multiresolution MRFs linked by the super coupling transform advocated by Petrou et al. (1998) [37, 11]. In this paper, we benefit from the daisy descriptor for face image representation in image matching. Most importantly, we design an innovative GPU implementation of the proposed multiresolution MRF matching process. The significant speed up achieved (factor of 25) has multiple benefits: It makes the MRF approach a practical proposition. It facilitates extensive empirical optimisation and evaluation studies. The latter conducted on benchmarking databases, including the challenging labelled faces in the wild (LFW) database show the outstanding potential of the proposed method, which consistently achieves state-of-the-art performance in standard benchmarking tests. The experimental studies also show that the super coupled multiresolution MRFs deliver a computational speed up by a factor of 5 over and above the speed up achieved using the GPU implementation.

1. Introduction

Although performing well in controlled conditions, current face recognition systems are seriously challenged by a number of factors including unconstrained pose and face expression and varying illumination. From the face image representation point of view, recognition methods can be roughly divided into global and local feature based approaches. As global features are not robust to local distortions, local feature based methods have been favoured in the recent developments. However, local regions from which features are extracted should correspond to the same regions across different images to be effective in recognition. Hence, image matching and alignment have become an integral part of the face recognition pipeline. Among the numerous solutions proposed to matching, the Markov Random Field methodology is known to be a promising approach, especially for handling large non-rigid transformations. Unfortunately, the optimisation and inference over a spatial graphical model is computationally demanding. Many algorithms based on MRFs are not commercially viable because of their computational complexity. The prohibitive computation time is also acutely felt in evaluating such algorithms on large databases.

Recently, the computational complexity has been somewhat alleviated by the development of efficient optimisation algorithms [27,57,26,12]. In particular, the idea of formulating the image matching problem on two interacting Markov Random Fields, focusing on disparities in the horizontal and vertical directions respectively, has been proven to be very powerful. This idea was successfully adapted for unconstrained pose invariant face recognition in [4,6,3,5]. In this work, pose-invariance is achieved via dense matching of images while illumination invariant representations adopted minimises the unwanted effects of illumination changes. Minimising the adverse effects of the background and unavailability of frontal gallery images on the recognition performance in unconstrained settings is achieved via a symmetrical matching process, *i.e.* the first image is matched to the second and then the roles of the two images to be compared are exchanged. The procedure is then repeated for the horizontally mirrored versions of both images and the final score is taken as the minimum of the distances thus obtained. For optimisation, distance transform technique [16] is employed for efficient message passing in the context of dual decomposition framework [27]. In addition, an incremental subgradient method [10] is used so that the more computationally demanding local updates in the decomposition

☆ This paper has been recommended for acceptance by Edwin Hancock.

* Corresponding author. Tel.: +98 9144416200.
E-mail addresses: Sh.rahimzadeh@hotmail.co.uk (S. Rahimzadeh Arashloo), J.kittler@surrey.ac.uk (J. Kittler).

http://dx.doi.org/10.1016/j.patrec.2014.05.017
0167-8655/

framework are performed less frequently. This MRF approach to face recognition, advocated in [4], has been shown to have appealing characteristics. It can cope with moderate translation, in and out of plane rotation, scaling and perspective effects without the need for non-frontal images in training. Furthermore, no strict assumption is made about the pose of the subject prior to matching.

Further speed up gains in optimisation are achieved by adopting a multiresolution approach [6] inspired by the work of Petrou et al. [37,11], who developed the methodology for linking the MRFs at multiple resolution levels to ensure a consistent model. Their method of coarsening the image is based on the renormalisation group theory (RGT) originally advocated by Gidas [18]. The RGT method provides a principled way of processing information of a Markov Random Field at multiple scales. For the multiresolution analysis of a given lattice problem, the method constructs coarser and coarser grids of an original lattice in different scales and then associates an energy function to each scale in a way that it is consistent with the energy of the original problem. During processing, a coarse to fine multiscale approach is pursued starting from the coarsest scale and moving to the finest scale. The RGT method provides certain benefits in a multiscale analysis. A major advantage of the approach is the accelerated optimisation. In a coarse scale, not only the number of sites are reduced but their admissible states are fewer compared to the original problem. Hence, a faster convergence is expected. Moving to a next finer scale, only those configurations of sites which are consistent with the previous coarser scale are considered. That is, the optimisation is now performed in a subspace of the finer scale, reducing the computational cost. In addition to accelerating convergence, it has been found that a multiresolution analysis based on RGT is instrumental in preventing the optimisation algorithm to get stuck in local minima. However, the RGT is known to be difficult in practice. To avoid the difficulties with RGT, Petrou et al. proposed an alternative transformation which does not have the full properties of the RGT but which preserves the global minimum of the cost function.

The key contribution of the current work is to reduce the processing time of inference in MRF image matching even further. In this respect, we focus on parallel processing algorithms and show how the optimisation problem for image matching can be reformulated to be solved on a GPU. We show how the dual decomposition approach to face image matching advocated in [4,6] can be ported onto a graphical processing unit for efficient implementation.

The current work also supersedes [4,6] in terms of texture modelling by employing more descriptive and distinctive features both for dense matching and recognition. Once the correspondence has been established, a *single* feature, *i.e.* multiscale LBP histogram descriptor [2] is used for classification. This contrasts with many other algorithms combining a plethora of different features to achieve an acceptable level of performance. The combination of all the modifications proposed in the current work results in a significant improvement over the best performing graph-based pose-invariant methods of face recognition [6,4] and other unconstrained face recognition methods, not only in terms of efficiency and computational cost but recognition performance confirmed by various extensive tests performed on different databases.

The rest of the paper is organized as follows. In Section 2, we review the literature on unconstrained face recognition. Section 3 introduces MRFs, leading to the formulation of image matching in their context, and the discussion of the role of the dual decomposition framework in the process of inference. An efficient message passing computation, a multiresolution analysis using RGT, the incremental subgradient method and GPU processing are discussed in Section 4. In Section 5, our classifier employing multiresolution LBPs is introduced. The evaluation of the method in terms of processing time and recognition accuracy, including a comparison to the state-of-the-art face recognition methods on different databases are discussed in Section 6. In Section 7, conclusions are drawn.

2. Related work

Most of the earliest attempts at face recognition exploit global features extracted via subspace approaches. Examples include Eigenface and Fisher-face methods [8,52]. However, presently the majority of the best performing methods widely make use of local features for characterising the face images. As an example, in [13], the authors use vector quantized local pixels to extract discriminative information. The proposed approach encodes the microstructures of the face via a learning-based encoding scheme. To handle the large pose variation in real-life scenarios, the authors proposed a pose-adaptive matching method that uses pose-specific classifiers to deal with different pose combinations of the matching face pair using features extracted from different components of faces. While [39] uses spatially localized Gabor filters in a multilayer approach using a multiple kernel learning technique for verification, [2,49] use histogram of local pattern features (such as LBP, LTP *etc.*) extracted locally from face images for recognition. In [36], the authors propose to use histogram of local binary pattern features extracted from orientation images for the single sample face recognition task. A recent approach to improve performance under difficult settings is to combine multiple features extracted locally such as the works in [13,30,62], wherein the combination is performed in a wide range of schemes from combination at the decision level to multiple kernel learning. Some recent methods based on similarity metric learning which adopt metric-learning approaches are presented in [22,34]. In [22], the authors use common local image representations such as LBP and SIFT descriptors which are combined through a metric learning approach for face recognition in the wild. Similarly, in [34], local image representations such as LBP and SIFT are extracted and used in a distance metric learning approach called pairwise constrained component analysis (PCCA). In [60,61], the authors propose a two-level classifier, training a small number of one-shot and two-shot classifiers for each test pair employing one or both test images as positive samples and an additional set of negative samples. Different variations of LBP descriptors such as patch based LBP are proposed and shown to surpass the standard LBP representation in a local image representation framework based on histograms. The authors in [30,29] also make use of this two-level classifier, employing a set of attribute (race, gender, hair colour, *etc.*) classifiers in the first classifier of the cascade. The representations employed are features extracted from different regions of face locally. The authors in [1] use a validation set of face pairs to choose the most effective local features from among a large set and then feed these to an SVM for verification. Recently, a blur tolerant image descriptor called Local Phase Quantization (LPQ) operator is introduced by Rahtu *et al.* [41]. LPQ has been shown to perform better than the Local Binary Pattern (LBP) operator in face recognition and texture classification. In [63], global and local Gabor phase pattern histograms are proposed for face recognition. In [15], a kernel discriminant analysis fusion approach is proposed to combine multiscale LBP and LPQ regional histograms for face recognition. The method is reported to achieve good performance in challenging conditions on a number of different databases.

It is generally known that good alignment is important to achieve high performance in face recognition with uncontrolled images [21,54]. A method which is often applied is the funnelling method of [23] which is based on the congealing method of [31] to deal with real world images. These methods estimate transformations which minimise image differences. Graph-based methods

constitute a major category in face recognition. In this framework [4,6,58,55], different parts of an object are allowed to be considered independently of other non-neighbouring parts which is useful for dealing with geometrical distortions and also handling occlusions and cluttered background. Furthermore, graph-based methods require a minimum number of training images and good performance can be achieved even by using a single gallery image per class. In the well known elastic bunch graph matching approach [59], a system for recognizing human faces from single images using a dynamic link architecture is proposed. Fiducial points on a face are characterised using Gabor wavelets and linked by dynamic links to form face graphs to be matched for recognition. Drawing on this pioneering work other approaches such the one in [55] is proposed. In [55], a Bayesian method for face recognition based on Markov Random Fields (MRF) modelling is proposed. Gabor wavelet coefficients are used as the base features while relationships between Gabor features at different pixel locations are used to provide higher order contextual constraints. The posterior probability of matching configuration is derived based on MRF modelling. Local search and discriminate analysis are used to evaluate local matches, and a contextual constraint is applied to evaluate mutual matches between local matches. In [6,4] an efficient dense image matching MRF model is exploited to estimate dense correspondences between a pair of images for pose invariant recognition of faces. Once images are matched, local LBP histograms are used for recognition. The approach proposed here uses a graph-based representation for dense and efficient pixel-wise matching of faces. Once correspondences are established between images, multiresolution texture features are used for classification taking into account the pixel-wise alignment of face images.

3. Image matching

For pose-invariant face recognition dense image matching has been motivated by the fact that unlike frontal pose, in which only two fiducial points (usually eye coordinates) are sufficient for alignment, for faces possibly rotated in-depth, a larger number of point correspondences are needed for effective recognition. Nevertheless, even frontal pose face recognition may benefit from such correspondence information to cope with changes in expression.

3.1. Markov Random Fields for image matching

A Markov Random Field is composed of a set of nodes $\mathcal{V}$ and a set of edges/hyper-edges $\mathcal{E}$. The nodes correspond to individual primitives of the object while the edges/hyper-edges encode the conditional dependencies/neighbourhood system of the nodes. The goal is to assign each node a label from a predefined admissible label set $\mathbb{X} = \{1, 2, \ldots, L\}$ subject to contextual constraints in a way that the energy of the assignment is minimum. In Markov models, contextual information is conveyed by small groups of nodes, the so called cliques, which are defined by the neighbourhood system. When the maximum cardinality of the cliques in the graph is two, the energy associated with the model can be expressed as

$$E(x;\theta) = \sum_{s\in\mathcal{V}} \theta_s(x_s) + \sum_{(s,t)\in\mathcal{E}} \theta_{st}(x_s, x_t) \tag{1}$$

where x_s denotes the labelling of node s and θ parametrises the energy. In our matching model which we have adopted from [45,4], nodes correspond to individual blocks of the image, while the labels are 2D displacement vectors such that when added to the coordinates of a block in the template image results in the coordinates of the corresponding block in the target image. The matching model consists of two layers of displacement models, one for the horizontal and the other for the vertical direction, Fig. 1.

3.1.1. Smoothness prior

The severity of admissible local deformations is controlled by the smoothness prior [4,6] on each layer of the model. Accordingly, for two neighbouring nodes s and t in the same layer with states x_s and x_t, respectively, the prior is set as

$$\theta_{st}(x_s, x_t) = \rho(x_s - x_t)^2 \tag{2}$$

where ρ is a normalising constant controlling the trade off between data fidelity and smoothness of the deformation.

3.1.2. Data term

Edge-based features are well known for their discriminatory, invariance and repeatability properties. Inspired by the ideas used in the SIFT descriptor [32], many other features including geometric blur [9], GLOH histogram [35], SURF [7] and Daisy [51], *etc.* have been proposed and used for recognition. We use the Daisy feature [51] for the construction of the data term in our model.

Constructing the Daisy feature vector at each pixel location entails computing the oriented gradient maps at several quantized directions. In this work, the oriented gradient maps are computed in eight directions. The oriented edge maps are filtered with various Gaussian kernels in the next step and then sampled around each pixel in different radii and directions to form a feature vector. The number of radii and directions for the sampling scheme is decided by the user. In general, employing a larger number of radii and directions provide a richer representation [48], while choosing a fewer samples is instrumental in reducing the computational cost. In this work, the feature vectors extracted at every pixel location are of size 72, *i.e.* we sample the points in one concentric circle and 8 directions in addition to the central pixel. The feature vectors obtained are then normalised to a unit length and compared using Euclidean distance. The data term for the model is then computed as sum of the Daisy distances inside a block. Constructing the Daisy feature vector at each pixel location entails computing the oriented gradient maps at several quantized directions. In this work we compute the oriented gradient maps in eight directions. The maps are filtered with various Gaussian kernels in order to produce convolved orientation maps. The last stage is to sample the image around each pixel in different radii and directions to form a feature vector. The number of radii and directions for the sampling scheme is decided by the user. In general, employing a larger number of radii and directions provide a richer representation [48], while choosing a fewer samples is instrumental in reducing the computational cost. In this work, the feature vectors extracted at every pixel location are of size 72, *i.e.* we sample the points in one concentric circle and 8 directions in addition to the central pixel. The feature vectors obtained are then normalised to a unit length and compared using Euclidean distance. Finally, the data term is computed as sum of the Daisy distances inside a block.

3.2. Dual decomposition for MAP inference

Some of the well known algorithms for MRF optimisation are graph-cuts [12], dual decomposition [27], TRW-S [26] and Max-sum diffusion [56]. Dual decomposition is chosen in this work for its perfect adaptability to parallel processing. The general idea of the dual decomposition for MRF optimisation is as follows. Given a large problem, one decomposes it into solvable smaller and more manageable subproblems and then extracts a solution to the original problem by combining the solutions obtained from the subproblems, Table 1. In this work we choose each subproblem as an edge along with the two end nodes. This decomposes our original problem into a large number of subproblems, Fig. 1. This choice is driven by the large number of streaming processors available in today's GPUs.

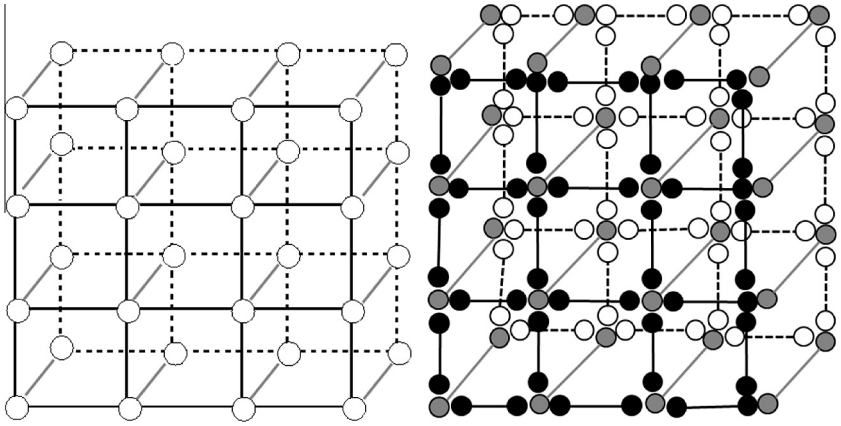

Fig. 1. Left: original two-layer graph, right: decomposition into edge-wise subproblems.

Table 1
Dual decomposition via subgradient updates [28].

1. Solve each slave MRF independently, *i.e.* compute $x^{slae} = \text{argmin}\, E(x, \theta^{Slae})$
2. Update parameters of each node in slave using subgradient updates [28]: $\theta^{slae} = \theta^{slae} + \delta_{iter}(x^{slae} - \frac{1}{N^{slaes}} \sum_{slaes} x^{slae})$
3. Repeat steps 1 and 2 till convergence.

4. Solving the subproblems

Once the original graph is decomposed into smaller subproblems, the dual decomposition approach starts by solving each subproblem independently. Although solving each subproblem may be performed via an exhaustive search, the complexity of such an algorithm is quadratically proportional to the number of admissible states of each node, making the inference inefficient. Consequently, it is essential to resort to more computationally efficient methods.

4.1. Efficient message passing

Two kinds of edges exist in our graph: the inter-layer and the intra-layer edges. Intra-layer edges encode a smoothness prior on the deformation field. One may employ max-product message passing algorithm for MAP inference over intra-layer each edge [53]. However, in a tree consisting of only one edge along with the two end nodes, the method fails to provide any computational advantage. This is due to the fact that direct computation of a message is of complexity $O(L^2)$. Fortunately, following the ideas proposed in [16], a message can be computed in linear time, *i.e* with a complexity of $O(L)$. We use the max-product algorithm and the distance transform [16] to infer the MAP state of each intra-layer edge.

4.2. Incremental subgradient updates

The distance transform method cannot be employed for inter-layer edges encoding the data term. Instead, one needs to perform an exhaustive search over these edges incurring a computational complexity of $O(L^2)$. However, in the dual decomposition framework, one may reduce the frequency of some updates via an incremental approach [10]. We update the less computationally demanding updates of intra-layer edges at every iteration but the inter-layer edges are updated less frequently, *e.g.* once in every n iterations, where n is determined empirically to achieve the best trade-off between accuracy and speed.

4.3. Multiresolution analysis

From a recognition point of view, the denser the correspondences between objects, the better the performance. However, there are certain drawbacks in increasing the resolution of the model. The computational complexity of the inference in graphical models may be considered as a major bottleneck of these approaches which is increased as the resolution of the model increases. In addition, increasing the number of variables makes the method more vulnerable to noise. Moreover, employing a high resolution increases the probability of the optimisation to get stuck in a local minimum as a result of increased dimensionality of the configuration space.

In order to reduce the computational burden while increasing accuracy and improving robustness to noise, one option is to link a number of sites to a single unit which may be achieved by grouping nodes together to produce bigger nodes and estimate a single label for each [45]. However, one drawback of this approach is the labelling error introduced by assuming that all nodes, constituting a bigger node, have similar labels which can adversely affect the recognition performance. The other option is to use a multiresolution approach. Multiresolution analysis has successfully been employed with the aim of avoiding the following problems: Although larger groups of nodes are used in coarser scales of the hierarchy, groups with smaller number of nodes are processed in finer scales. The idea supporting such an approach is that the coarser levels provide a rough estimate of the displacements and finer levels serve to fine-tune the result of a previous coarser level. Motivated by the successes of multiscale MRF models, in this work we employ a multiresolution method for multi-level matching of face images. The method starts with large blocks in a *coarse* configuration. As the method proceeds, successively smaller blocks and *finer* scale configurations are used for inference. The differences between the method adopted here which is based on the supercoupling transform [11,37,18] and some other heuristic methods [3] are as follows. In a heuristic approach, there is no explicit consistency between the energies being optimised at different levels whereas the consistency of the energy functionals in different scales in the applied scheme is maintained through the supercoupling transform [11,37]. In addition, in this work one achieves a further speed-up, as compared to a heuristic approach, by introducing a lumpiness into the configuration. This effectively translates into subsampling the displacements in the coarser levels, thus reducing the complexity of inference using the dual decomposition, *i.e.* the core optimisation method. Moving to the finer levels, in the proposed technique, only those fine level labels which are consistent with the solution in the previous coarser

resolution are considered which in turn serves to reduce the complexity of the whole optimisation.

4.4. RGT for multiresolution analysis

There are two main considerations in applying multiresolution techniques. The first addresses how to coarsen the data and how to transform the posterior distribution in a way that the solution to the problem remains consistent across all scales. The second consideration is how to propagate the labelling obtained at a coarse resolution to the next finer scale so that the optimum at the finer level is reached more efficiently and it is consistent with the coarser scale solution.

The Renormalisation Group Transform (RGT) [18,11] provides solutions to both problems in a principled way. The RGT algorithm consists of two main stages: *renormalisation* and *processing*. In the renormalisation step, finer and finer grids of nodes and a corresponding sequence of energy functionals are iteratively constructed. Suppose there is an original grid of size $2^N \times 2^N$. Then in the next level, a coarser lattice is obtained by grouping every 4 nodes together and identifying them as a single node. For defining the energy functional for each coarser level one needs to choose a probability function ($P(X' \mid \bar{X})$) measuring how likely a coarse configuration (X') is given a finer configuration $\bar{X}$:

$$e^{E(X')} = \sum_{\bar{X}} P(X' \mid \bar{X}) e^{E(\bar{X})} \tag{3}$$

where $E(.)$ denotes the energy of a configuration. In the *processing* stage, a multiscale coarse-to-fine optimisation is pursued. In other words, optimisation is performed in a coarse scale and then the next finer level is processed wherein only those configurations which are *constrained* by the obtained solution in the previous coarse scale are considered. Maximum A Posteriori search in a *subspace* configuration of the next finer level reduces the computational complexity of the optimisation. If the conditional probabilities $P(X' \mid \bar{X})$ are chosen as delta functions, then the procedure finds the global minimum of the energy [18].

In a multiresolution approach based on the renormalisation group transform, the whole structure of the probability distribution is preserved. In practice the RGT is known to be difficult in implementation. However, in most cases, preserving the full structure of the probability distribution is not required as only its maximum is sought (just as in MAP-MRF estimation). In such cases, a potential-based coarsening technique, referred to as super-coupling transform [11], which is known to be order preserving, is employed. As a result of the order preserving property, the mode of the original fine configuration is mapped onto the mode of the coarsened distribution. The multiresolution approach based on the supercoupling transform is meant to reduce the time required for the optimisation process by coarsening the configuration space by enabling long range jumps to guide the whole optimisation faster to the global minimum.

As noted earlier, an important issue in multiresolution analysis is the propagation of the solution from one level to the next finer scale. A common practice is to employ a block-flat assumption giving the same label to all nodes inside a block. In this way, the solution obtained at a coarser level serves as a starting point for the optimisation in the finer resolution. According to the supercoupling transform, under the block-flat assumption the value of the cost function when moving from one level of resolution to another should stay unchanged. It was shown in [11] that this transformation, at the zero temperature limit, is identically the same as RGT.

4.5. Transforming the posterior distribution

Derivation of the parameters for transforming the posterior distribution in the supercoupling optimisation framework is presented elsewhere [6] but is restated here for the text to be self contained. For conciseness, we will consider two levels of resolution, a coarse level and the next fine scale. It is assumed that images are of size $2^N \times 2^N$. The coarse lattice is constructed by replacing every four nodes in the finer scale and identifying them as a single node. As a result, each node in the coarse lattice (denoted by s) corresponds to four nodes in the finer lattice (denoted by s_1, s_2, s_3 and s_4), Fig. 2. The fine configuration that can be produced using the block-flat assumption from the coarse configuration X' is denoted by $\bar{X}$. The theory of super-coupling transform then requires that the parameters of the posterior distribution should be determined in a way that

$$E(\bar{X}) = E(X') \tag{4}$$

As a result, for each site in the coarse lattice and its four corresponding sites in the fine level the following equation must hold

$$\begin{aligned}
&\theta_{s_1}(\bar{x}_{s_1}) + \sum_{(s_1,u_1)\in\mathcal{E}_f} \theta_{s_1u_1}(\bar{x}_{s_1},\bar{x}_{u_1}) + \theta_{s_2}(\bar{x}_{s_2}) + \sum_{(s_2,u_2)\in\mathcal{E}_f} \theta_{s_2u_2}(\bar{x}_{s_2},\bar{x}_{u_2}) \\
&\quad + \theta_{s_3}(\bar{x}_{s_3}) + \sum_{(s_3,u_3)\in\mathcal{E}_f} \theta_{s_3u_3}(\bar{x}_{s_3},\bar{x}_{u_3}) + \theta_{s_4}(\bar{x}_{s_4}) \\
&\quad + \sum_{(s_4,u_4)\in\mathcal{E}_f} \theta_{s_4u_4}(\bar{x}_{s_4},\bar{x}_{u_4}) \\
&\quad = \theta'_s(x'_s) + \sum_{(s,u)\in\mathcal{E}_c} \theta'_{su}(x'_s,x'_u)
\end{aligned} \tag{5}$$

where $\mathcal{E}_f$ and $\mathcal{E}_c$ represent the edge sets in the fine and coarse scales, respectively. Paying attention to the relative positions of the sites illustrated in Fig. 2, we have

$$\begin{aligned}
&\bar{x}_{s_1} = \bar{x}_{s_2} = \bar{x}_{s_3} = \bar{x}_{s_4} = x'_s, \\
&\bar{x}_{s_1t} = x'_{st}, \bar{x}_{s_1r} = x'_s, \bar{x}_{s_1b} = x'_s, \bar{x}_{s_1l} = x'_{sl}, \\
&\bar{x}_{s_2t} = x'_{st}, \bar{x}_{s_2r} = x'_{sr}, \bar{x}_{s_2b} = x'_s, \bar{x}_{s_2l} = x'_s, \\
&\bar{x}_{s_3t} = x'_s, \bar{x}_{s_3r} = x'_{sr}, \bar{x}_{s_3b} = x'_{sb}, \bar{x}_{s_3l} = x'_s, \\
&\bar{x}_{s_4t} = x'_s, \bar{x}_{s_4r} = x'_s, \bar{x}_{s_4b} = x'_{sb}, \bar{x}_{s_4l} = x'_{sl}
\end{aligned} \tag{6}$$

where the subscripts *t, b, l, r* are used to denote the top, bottom, left or the right neighbour of a site in an immediate four-connected neighbourhood system. Considering the data term separately in Eq. (5) we have

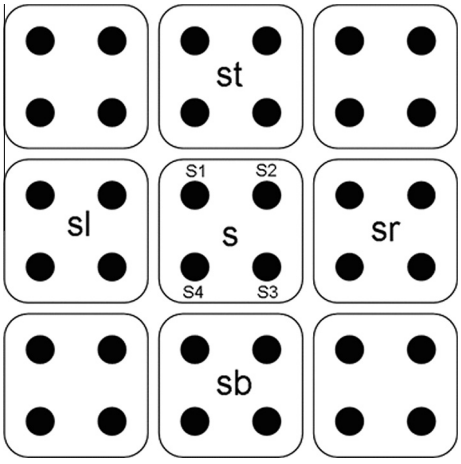

Fig. 2. Geometry of sites in the coarse and fine lattice under consideration.

$$\theta'_s(x'_s) = \theta_{s_1}(x'_s) + \theta_{s_2}(x'_s) + \theta_{s_3}(x'_s) + \theta_{s_4}(x'_s) = \sum_{i=1}^{4} \theta_{s_i}(x'_s). \tag{7}$$

Hence the data term associated with a block in the coarse level is determined as the sum of its four corresponding nodes in the next finer scale.

Similarly the pairwise potentials in Eq. (5) must satisfy

$$\begin{aligned}
&\theta_{s_1,s_1t}(\bar{x}_{s_1}, \bar{x}_{s_1t}) + \theta_{s_1,s_1r}(\bar{x}_{s_1}, \bar{x}_{s_1r}) + \theta_{s_1,s_1b}(\bar{x}_{s_1}, \bar{x}_{s_1b}) + \theta_{s_1,s_1l}(\bar{x}_{s_1}, \bar{x}_{s_1l}) \\
&\quad + \theta_{s_2,s_2t}(\bar{x}_{s_2}, \bar{x}_{s_2t}) + \theta_{s_2,s_2r}(\bar{x}_{s_2}, \bar{x}_{s_2r})\theta_{s_2,s_2b}(\bar{x}_{s_2}, \bar{x}_{s_2b}) \\
&\quad + \theta_{s_2,s_2l}(\bar{x}_{s_2}, \bar{x}_{s_2l}) + \theta_{s_3,s_3t}(\bar{x}_{s_3}, \bar{x}_{s_3t}) + \theta_{s_3,s_3r}(\bar{x}_{s_3}, \bar{x}_{s_3r}) \\
&\quad + \theta_{s_3,s_3b}(\bar{x}_{s_3}, \bar{x}_{s_3b}) + \theta_{s_3,s_3l}(\bar{x}_{s_3}, \bar{x}_{s_3l}) + \theta_{s_4,s_4t}(\bar{x}_{s_4}, \bar{x}_{s_4t}) \\
&\quad + \theta_{s_4,s_4r}(\bar{x}_{s_4}, \bar{x}_{s_4r}) + \theta_{s_4,s_4b}(\bar{x}_{s_4}, \bar{x}_{s_4b}) + \theta_{s_4,s_4l}(\bar{x}_{s_4}, \bar{x}_{s_4l}) \\
&\quad = \theta'_{st}(x'_s, x'_t) + \theta'_{sr}(x'_s, x'_r) + \theta'_{sb}(x'_s, x'_b) + \theta'_{sl}(x'_s, x'_l).
\end{aligned} \tag{8}$$

using (6) in (8) one gets

$$\begin{aligned}
&\theta_{s_1,s_1t}(x'_s, x'_t) + \theta_{s_1,s_1l}(x'_s, x'_l) + \theta_{s_2,s_2t}(x'_s, x'_t) + \theta_{s_2,s_2r}(x'_s, x'_r) + \theta_{s_3,s_3r}(x'_s, x'_r) \\
&\quad + \theta_{s_3,s_3b}(x'_s, x'_b) + \theta_{s_4,s_4b}(x'_s, x'_b) + \theta_{s_4,s_4l}(x'_s, x'_l) \\
&\quad = \theta'_{st}(x'_s, x'_t) + \theta'_{sr}(x'_s, x'_r) + \theta'_{sb}(x'_s, x'_b) + \theta'_{sl}(x'_s, x'_l).
\end{aligned}$$

and hence

$$\begin{aligned}
\theta'_{st}(x'_s, x'_t) &= \theta_{s_1,s_1t}(x'_s, x'_t) + \theta_{s_2,s_2t}(x'_s, x'_t) \\
\theta'_{sr}(x'_s, x'_r) &= \theta_{s_2,s_2r}(x'_s, x'_r) + \theta_{s_3,s_3r}(x'_s, x'_r) \\
\theta'_{sb}(x'_s, x'_b) &= \theta_{s_3,s_3b}(x'_s, x'_b) + \theta_{s_4,s_4b}(x'_s, x'_b) \\
\theta'_{sl}(x'_s, x'_l) &= \theta_{s_1,s_1l}(x'_s, x'_l) + \theta_{s_4,s_4l}(x'_s, x'_l)
\end{aligned} \tag{9}$$

By adopting the quadratic pairwise potential, we have:

$$q'(x'_s - x'_t)^2 = q(x'_s - x'_t)^2 + q(x'_s - x'_t)^2 \tag{10}$$

implying $q' = 2q$, which indicates that the model prescribes a stronger interaction between sites in the higher levels of the hierarchy. This is intuitive as in coarser resolutions the sites represent larger groups of pixels, which would require stronger interaction.

4.6. Graphical processing units (GPUs)

In recent years, parallel processing methods have attracted a lot of interest due to the technological advancements in designing new parallel computing devices. The quest to parallel processing has been intensified further with the advent of general easy to use programming interfaces. The many-core GPUs can be used as a large number of small core numeric computing engines. However, in general GPUs would not do well on tasks CPUs are optimised to perform. For this reason, the CUDA (Compute Unified Device Architecture) programming model was introduced by NVIDIA to support joint CPU/GPU execution of an application.

4.6.1. Hardware model

At the hardware level, a CUDA enabled GPU is organized into an array of highly threaded streaming multiprocessors (SMs) each containing a number of streaming processors (SPs). Each multiprocessor has a high speed shared memory visible to all its processing elements. Also, the GPU has a number of registers, texture, and constant memory caches in addition to the global DRAM memory. All SPs in SMs execute the same instruction at the same time on their own data. Communication between MPs is only through the DRAM which has a higher latency compared to other types of memory available on the chip.

4.6.2. Programming model

From a CUDA programmer point of view, the computing system is composed of a host, *i.e.* a traditional central processing unit (CPU), and one or more devices, which are parallel processors fitted with a large number of execution units. In many applications, program sections include a rich amount of data parallelism, paving the way to many operations to be executed on program data structures simultaneously. As all threads perform the same instructions, CUDA programming model is an example of the widely known single-program, multiple-data (SPMD) parallel programming style. A *kernel* function in CUDA, determines the code to be executed by all threads during a parallel computation stage. The kernels usually generate a large number of threads to utilise data parallelism. When a kernel is launched, or invoked, it is executed as grid of parallel threads.

4.6.3. Porting the optimisation onto the GPU

As noted earlier, an essential prerequisite for a problem to benefit from the processing power of GPUs is to be composed of independent computationally intensive parts. Such parallelism exists in the dual decomposition framework. The dual decomposition method is an iterative scheme consisting of two core stages: solving the sub-problems independently and updating the Lagrange multipliers. As a result, by processing the subproblems and updating Lagrange multipliers in parallel, large speed up gains may be achieved. In addition to the core optimisation method, there are several other functions in the matching process which are essentially parallel processes. For clarity, in what follows the routines used in the main algorithm are provided in terms of pseudo-codes. The first algorithm presented is the main method implementing the multiresolution optimisation in four levels, starting from 8×8 blocks down to the pixel level.

Algorithm 1. The main matching pseudo-code

```
Data: template image; target image
Result: 2D displacements of matching
begin
    daisy_1 = Daisy(template);
    daisy_2 = Daisy(target);
    initialise X ;
    for i ← 3 to 0 do
        X = Level_i(daisy_1, daisy_2, X);
    end
    AppRes(template, X);
end
```

The **Daisy**(.) routine in Algorithm 1 computes the daisy feature vectors for all image pixels. The routine takes advantage of the processing power of the GPU by incorporating parallel subroutines for applying separable 2D filters on the image. These include the computation of image derivatives and applying Gaussian smoothing filters to compute the feature vector. Two additional subroutines to compute the gradient orientation maps and sampling are also implemented in parallel. The AppRes(,,) function applies the 2D deformation found (X) to the template image. The routine of $\textbf{Level}_i(.,.,.)$ performs the MRF optimisation in the ith level. The pseudo-code of $\textbf{Level}_i(.,.,.)$ is given in Algorithm 2. In this algorithm, the *step* parameter is the step size for updating dual variables in the dual decomposition approach. *NP* and *ILEP* stand for node potentials and inter-layer edge potentials, respectively. X is the solution vector, *i.e.* the 2D displacements initialised to zero at the coarsest resolution.

Algorithm 2. Pseudo-code of **Level**$_i$

Data: $daisy_1$; $daisy_2$;X
Result: 2D displacements of matching in the i^{th} level
begin
  initialise step;
  NP $\leftarrow 0$;
  ILEP $=$ **pots**$_i(daisy_1, daisy_2, X)$;
  for $iter \leftarrow 1$ **to** $iter\_final$ **do**
    $X =$ **sub_solve**$_i(NP, ILEP, X, iter)$;
    $NP =$ **update**$(NP, X, iter, step)$;
    update step;
  end
end

The **pots**$_i(.,.,.)$ routine in Algorithm 2 computes the data term in level i for all inter-layer edges simultaneously using the Euclidean distance on the Daisy feature vectors. The two other subroutines of **sub_solve**$_i(.,.,.,.)$ and **update**$(.,.,.,.)$ are massively parallel CUDA kernels detailed in Algorithms 3 and 4, respectively. In these algorithms, the term *forall* indicates a CUDA kernel invocation operating in parallel on all computing resources.

Algorithm 3. Pseudo-code of **sub_solve**_i

Data: ILEP,NP,X,iter
Result: X
begin
  //solve for intralayer(NP):
  forall the *intra_layer edges* **do**
    max-product message passing using distance transform;
    compute MAP state;
  end
  //solve for interlayer(NP,ILEP):
  if $(iter\%n == 0)$ **then**
    forall the *inter_layer edges* **do**
      exhaustive search and compute MAP state;
    end
  end
end

Algorithm 4. Pseudo-code of **update**_i

Data: NP,X,step
Result: X
begin
  forall the *intra_layer edges* **do**
    update the two end nodes' potentials;
  end
  forall the *inter_layer edges* **do**
    update the two end nodes' potentials;
  end
end

5. Classification

Fig. 3 presents an overview of our face recognition pipeline. The two images to be compared are shown in the leftmost column which are matched symmetrically. The symmetric dense matching entails matching 8 pairs of images which are generated by replacing the roles of the first and second image and making use of the mirrored versions of both images as well. Once correspondences are established between each pair of images, the multiscale local binary pattern (MLBP) histogram features are used to produce a distance score. Before applying the LBP operators, we normalise the face images using an effective photometric normalisation scheme [49]. For face description, the template images are partitioned into 64 non-overlapping rectangular regions and their corresponding regions in the target image are identified taking into account the registration information. Uniform LBP histograms are then extracted in 10 different resolutions from each region and their corresponding patches in the target images. These are concatenated to form a single vector. A PCA transformation is applied to reduce the dimensionality of the feature vectors for each region. The resulting feature vectors are then compared and a match score is produced for each pair of regions. Taking a classifier fusion approach, the final dissimilarity score of each pair out of the 8 pairs of images is defined as

$$Dis(I, I') = \sum_j \frac{-d_j d'_j}{\|d_j\| \|d'_j\|} \tag{11}$$

where $Dis(I, I')$ stands for the dissimilarity of the two images I and I'. d_j is the feature vector of region j in image I after PCA transformation. d'_j denotes the feature vector of the corresponding region in image I'. If no training is allowed, the χ^2 distance measure is used for comparison. After obtaining scores for all eight pairs, the minimum distance is considered as the final score between the two images being compared.

6. Experimental evaluation

In this section we first present some image matching results and next provide the results of evaluating the proposed approach both in terms of running time and recognition performance. Fig. 4 depicts intermediate results of multiresolution matching. The result of matching at each resolution is presented in addition to the final pixelwise correspondence (Level 0). It can be observed that the method provides very good results even in the presence of imperfections in imaging conditions such as pose, self-occlusion, non-uniform lighting conditions, *etc.*

Next, we evaluate the proposed face matching method in terms of computational time and compare it to that of [3,4].

6.1. Gains in running time

In order to compare the run time of different methods, we use the original source codes of [4] leaving the parameters unchanged. In this experiment, a template image of size 112×128 pixels is matched against a target image having the range of displacements set to 32 pixels in each direction. The GPU used in all experiments is an NVIDIA Geforce GTX 460 SE.

Table 2 reports the effects of different techniques used in the proposed approach. From the table it can be observed that the parallel computation on the GPU accelerates the matching process $\sim 24\times$ compared to the baseline method of [4]. This is achieved by exploiting the computational resources of the GPU which make it possible to process a large number of subproblems in parallel. Next, it is observed that the multiresolution analysis accounts for $\sim 500\%$ efficiency. In other words, the four-scale multiresolution

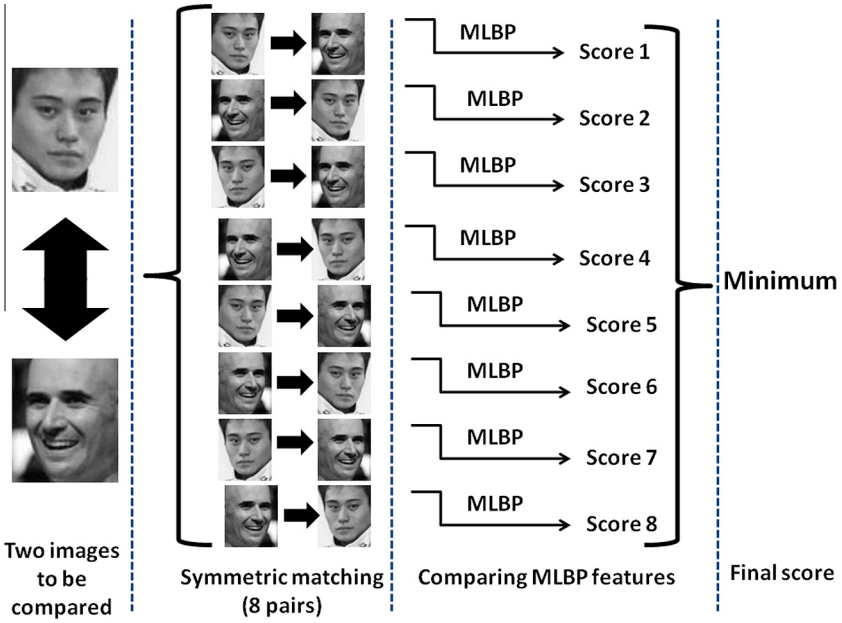

Fig. 3. Overview of our face recognition system.

Table 2
Speed up gains achieved via different techniques compared to the method in [4].

GPU	M.R. analysis	Eff. mess.	Inc. sub.	Overall
$\sim 24\times$	$\sim 5\times$	$\sim 1.4\times$	$\sim 1.3\times$	$\sim 218\times$

analysis based on the RGT makes the method faster 5 times over and above the speed up gain achieved using GPU. As noted earlier, the efficient message computation algorithm is instrumental in reducing the computational complexity of the algorithm. From the table, it is apparent that the employed technique accelerates the method $\sim 1.4\times$. For the incremental subgradient approach, we update inter-layer edges once in every two iterations. The incremental approach accounts for a $\sim 25\%$ reduction in running time making the algorithm $\sim 1.3\times$ faster. The final column of the table, reports the overall effects of the proposed technique. It is observed that the proposed matching method is more than 200 times faster than the method of [4]. For the proposed method it takes about 1.4 s to match a template image of size 112×128 to

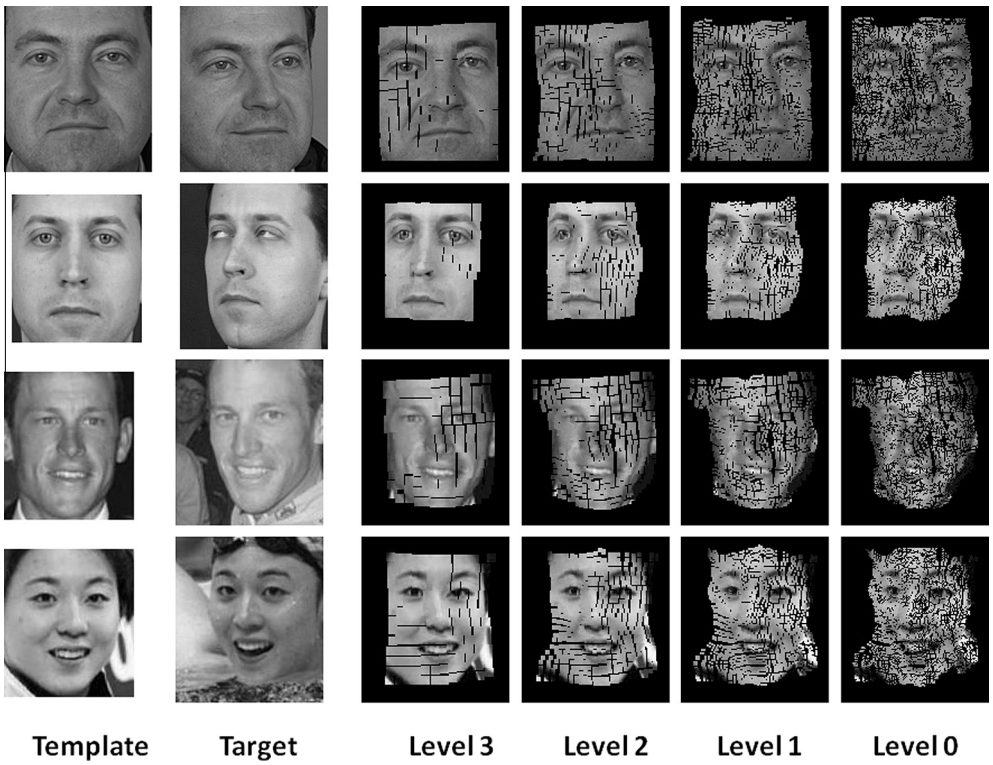

Fig. 4. From left to right in each row: template, target, deformed template in the forth level, third level, second level and the first (pixel) level.

the target image compared to the method of [4] which takes more than 5 min for the same task.

6.2. Unseen pair matching on the LFW database

Next, we evaluate the performance of the proposed approach in a task of face pair matching in challenging real-life situations. Recently, with the development of the LFW dataset [24] it has been possible to study the performance of face recognition methods in unconstrained settings. The LFW dataset includes real world variations in facial images such as pose, illumination, expression, occlusion, low resolution, *etc.* It contains 13,233 images of 5,749 subjects. The task is to determine whether a pair of images belongs to the same person or not. We evaluate the proposed approach on the "View 2" of the dataset consisting of 3,000 matched and 3,000 mismatched pairs divided into 10 sets. The evaluation is performed in a leave-one-out cross validation scheme. The aggregate performance of the method over ten folds is reported as the mean accuracy and the standard error on the mean. There are three evaluation settings on this database: the image unrestricted setting, the image restricted setting and the unsupervised setting. The most difficult one is the unsupervised setting where no training data is available. The two other settings allow the use of training data for the image pairs as "same" or "not same". The image unrestricted setting in addition provides the identity of the subjects in each pair. We evaluate the proposed approach on the two most difficult settings of the unsupervised and the image restricted.

6.2.1. Unsupervised setting

We first examine the effectiveness of the proposed methodology in the unsupervised setting of the LFW database where we do not use any training data. In order to compare the results with other methods under this setting, we use a version of the LFW database called LFW-a which is aligned using a commercial alignment software [61]. We crop face images closely to minimise the effects of background samples. In this experiment, the multiresolution LBP histograms are computed from each region of the first image and compared against the corresponding region in the second image using the χ^2 distance measure. As a result, our method is tested in a *training free* scenario. In contrast to many other databases, in the LFW dataset no *frontal* gallery image is available. In order to minimise the effects of the background and unavailability of frontal gallery images on the recognition performance, the previously described symmetric matching is employed. Table 3 reports the results under this setting. The previous two best results under this setting are 72.23% using LARK features [44] and 86.20% using I-LPQ method [25]. We achieve a mean accuracy of 80.08% indicating a large 7.85% improvement over the previous second best result ranking our approach second best.

6.2.2. Image restricted setting

Next, we evaluate the proposed method in the image restricted setting where the training data is provided as the "same" or "not the same" for each pair without providing subject identities. In this setting, the χ^2 distance measure is used as the dissimilarity measure. For this experiment, we use a second version of LFW images called funnelled. The images of this version are obtained using the aligning algorithm of [23], which does not require outside training data for alignment. We compare our results with the methods which use strictly LFW training data without making use of outside training data. Similar to the unsupervised setting, we apply the symmetric matching process to an image pair. Table 4 reports the performance of various approaches. Under this setting, we achieve an accuracy of $79.08 \pm .0014\%$ which ranks our method second best among all. The best performing method is the one in [39] with an accuracy of $79.35 \pm .0055\%$ only 0.27% better than what we achieve on average.

Table 3
Comparison of the performance of the proposed approach to the state-of-the-art methods on the LFW database in the unsupervised setting using a single feature.

Method	$\mu \pm S_E$
SD-MATCHES, 125×125, aligned [47]	0.6410 ± 0.0062
H-SX-40, 81×150, aligned [47]	0.6945 ± 0.0048
GJD-BC-100, 122×225, aligned [47]	0.6847 ± 0.0065
LARK unsupervised, aligned [44]	0.7223 ± 0.0049
I-LPQ, aligned [25]	0.8620 ± 0.0046
This work, aligned	0.8008 ± 0.0013

Table 4
Comparison of the performance of the proposed approach to the state-of-the-art methods on the LFW database in the image restricted setting using a single feature (strict LFW, no outside training data used).

Method	$\mu \pm S_E$
Eigenfaces, original [52]	0.6002 ± 0.0079
Nowak, original [38]	0.7245 ± 0.0040
Nowak, funnelled [38]	0.7393 ± 0.0049
Hybrid descriptor-based, funnelled [60]	0.7847 ± 0.0051
3×3 Multi-region Histograms (1024) [43]	0.7295 ± 0.0055
Pixels/MKL, funnelled [39]	0.6822 ± 0.0041
V1-like/MKL, funnelled [39]	0.7935 ± 0.0055
This work, funnelled	0.7908 ± 0.0014

6.3. Verification test on the XM2VTS database

In the XM2VTS rotation data set [33] the evaluation protocol is based on 295 subjects consisting of 200 clients, 25 evaluation imposters and 70 test imposters. Two error measures defined for a verification system are false acceptance and false rejection given below:

$$FA = EI/I * 100\%, \quad FR = EC/C * 100\% \tag{12}$$

where I is the number of imposter claims, EI the number of imposter acceptances, C the number of client claims and EC the number of client rejections. The performance of a verification system is often stated in *Equal Error Rate* (EER) in which the FA and FR are equal and the threshold for acceptance or rejection of a claimant is set using the true identities of test subjects. Table 5 reports the equal error rates obtained on the XM2VTS dataset using the proposed approach compared to some other methods. These include the method in [50], where the authors use a 3D morphable model for geometrically normalising the rotated images and then use LBP histograms in the 2D geometrically normalised images. In [4] the authors use a single resolution LBP histogram together with the shape information. From the table, it is observed that the proposed method outperforms both of these methods, despite the fact that we do not make use of shape information explicitly.

6.4. Identification test on the CMU-PIE database

In this test, we use images of the CMU-PIE database [46] captured under almost the same illumination conditions with neutral

Table 5
Comparison of the performance of the proposed method to the state-of-the-art methods on the XM2VTS database.

Method	3D correc. [50]	Face matching [4]	Current work
EER	7.12	4.85	**4.27**

The best results are indicated in bold.

Table 6
Comparison of the performance of the proposed approach to the state-of-the-art methods on the CMU-PIE database.

Pose	C02	C05	C07	C09	C11	C14	C22	C25	C29	C31	C34	C37
Horizontal deviation angle	−44°	−16°	0°	0°	32°	47°	−62°	−44°	17°	47°	66°	−31°
Vertical deviation angle	0°	0°	−13°	13°	0°	0°	1°	11°	0°	11°	1°	0°
Eigenlight-fields Complex [20]	58	94	89	94	88	70	38	56	57	56	47	89
PDM [19]	72	100	na	na	94	62	na	na	98	na	20	97
AA-LBP [64]	95	100	100	100	100	91	na	89	100	80	73	100
3D morphable model [42]	76	99	99	99	93	87	50	75	97	78	49	94
LLR [14]	na	98	98	98	89	na	na	na	100	na	na	82
Face matching [4]	95	98	98	100	89	91	79	95	91	88	83	100
CTRW-S [40]	98	98	100	100	100	98	84	97	98	94	90	**100**
CTSDP [17]	98	98	100	100	100	98	80	97	100	94	87	100
Current work	**100**	**100**	**100**	**100**	**100**	**100**	**92**	**100**	**100**	**100**	**95**	**100**

The best results are indicated in bold.

expression consisting of 884 images of 68 subjects viewed from 13 different angles. Frontal views of subjects (pose 27) are considered as gallery images while all the rest (12 different poses) are used as test images. The results are reported in Table 6. From the table it can be observed that the proposed technique outperforms all other approaches achieving 100% recognition rates in all poses but the two profile views. The best performing methods among other approaches are the ones in [17,40], with an average overall performance of 94.68% and 96.57%, respectively. The proposed method achieves a higher average recognition rate of 98.91% compared to both methods of [17,40]. Considering the performance of other approaches, our method achieves the lowest reported errors rates on the CMU-PIE dataset with less restrictive assumptions and minimal injection of prior information.

7. Conclusion

The unconstrained-pose face recognition problem was addressed using the framework of MRF dense image matching. We solved the challenging optimisation problem of MAP inference over the underlying MRFs formulated in [4,6] by exploiting the processing power of GPUs. A number of different techniques including multi-resolution analysis based on the RGT theory, efficient message passing using distance transform and the incremental subgradient approach are an integral part of the solution to obtain maximum efficiency gains of the proposed approach. The combination of these techniques was shown to result in a factor of more than two hundred speed up as compared to the baseline methods. In order to increase the efficacy of the approach, multiresolution Daisy features were used to achieve invariance against deformations and lighting changes. Finally, for classification, multiresolution LBP histograms were used to capture discriminative textural content of the images in different scales.

The experimental evaluation of the method, performed on different challenging databases in various scenarios, demonstrated an impressive face matching accuracy of the proposed approach.

Acknowledgement

The work was partially supported from the EPSRC research Grants reference EP/K014307/1, EP/H049665/1 and EP/F069421/1.

References

[1] Beyond simple features: a large-scale feature search approach to unconstrained face recognition, 2011.
[2] T. Ahonen, A. Hadid, M. Pietikainen, Face description with local binary patterns:application to face recognition, PAMI 28 (2006) 2037–2041.
[3] S.R. Arashloo, J. Kittler, Hierarchical image matching for pose-invariant face recognition, in: BMVC, British Machine Vision Association, 2009.
[4] S.R. Arashloo, J. Kittler, Energy normalization for pose-invariant face recognition based on mrf model image matching, IEEE Trans. Pattern Anal. Mach. Intell. 33 (2011) 1274–1280.
[5] S.R. Arashloo, J. Kittler, W.J. Christmas, Facial feature localization using graph matching with higher order statistical shape priors and global optimization, in: Proceedings of IEEE Fourth International Conference on Biometrics: Theory, Applications and Systems.
[6] S.R. Arashloo, J. Kittler, W.J. Christmas, Pose-invariant face recognition by matching on multi-resolution mrfs linked by supercoupling transform, Comput. Vis. Image Underst. 115 (2011) 1073–1083.
[7] H. Bay, T. Tuytelaars, L.V. Gool, Surf: speeded up robust features, in: ECCV, pp. 404–417.
[8] P. Belhumeur, J. Hespanha, D. Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (1997) 711–720.
[9] A.C. Berg, J. Malik, Geometric blur for template matching, in: CVPR 2001, pp. 607–614.
[10] D.P. Bertsekas, Nonlinear Programming, second ed., Athena Scientific, 1999.
[11] M. Bober, M. Petrou, J. Kittler, Nonlinear motion estimation using the supercoupling approach, IEEE Trans. Pattern Anal. Mach. Intell. 20 (1998) 550–555.
[12] Y. Boykov, O. Veksler, R. Zabih, Fast approximate energy minimization via graph cuts, IEEE Trans. Pattern Anal. Mach. Intell. 23 (2001) 1222–1239.
[13] Z. Cao, Q. Yin, X. Tang, J. Sun, Face recognition with learning-based descriptor, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010, pp. 2707–2714.
[14] X. Chai, S. Shan, X. Chen, W. Gao, Locally linear regression for pose-invariant face recognition, IEEE Trans. Image Process. 16 (2007) 1716–1725.
[15] C.H. Chan, M. Tahir, J. Kittler, M. Pietikainen, Multiscale local phase quantization for robust component-based face recognition using kernel fusion of multiple descriptors, IEEE Trans. Pattern Anal. Mach. Intell. 35 (2013) 1164–1177.
[16] P. Felzenszwalb, D. Huttenlocher, Efficient belief propagation for early vision I, in: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2004, pp. 261–268.
[17] T. Gass, L. Pishchulin, P. Dreuw, H. Ney, Warp that smile on your face: Optimal and smooth deformations for face recognition, in: IEEE International Conference on Automatic Face and Gesture Recognition, Santa Barbara, CA, USA, pp. 456–463.
[18] B. Gidas, A renormalization group approach to image processing problems, PAMI 11 (1989) 164–180.
[19] D. Gonzalez-Jimenez, J. Alba-Castro, Toward pose-invariant 2-d face recognition through point distribution models and facial symmetry, IEEE Trans. Inf. Forensics Secur. 2 (2007) 413–429.
[20] R. Gross, I. Matthews, S. Baker, Appearance-based face recognition and light-fields, IEEE Trans. PAMI 26 (2004) 449–465.
[21] L. Gu, T. Kanade, A generative shape regularization model for robust face alignment, in: ECCV08.
[22] M. Guillaumin, J. Verbeek, C. Schmid, Is that you? metric learning approaches for face identification, Kyoto, Japan.
[23] G.B. Huang, V. Jain, E. Learned-Miller, Unsupervised Joint Alignment of Complex Images, in: IEEE 11th International Conference on Computer Vision, 2007. ICCV, 2007 pp. 1–8.
[24] 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 07-49, University of Massachusetts, Amherst, 2007.
[25] S. ul Hussain, T. Napoleon, F. Jurie, Face recognition using local quantized patterns, in: BMVC2012, pp. 1–11.
[26] V. Kolmogorov, Convergent tree-reweighted message passing for energy minimization, IEEE Trans. Pattern Anal. Mach. Intell. 28 (2006) 1568–1583.
[27] N. Komodakis, N. Paragios, G. Tziritas, MRF energy minimization and beyond via dual decomposition, IEEE Trans. Pattern Anal. Mach. Intell. 33 (2011) 531–552.
[28] N. Komodakis, N. Paragios, G. Tziritas, Mrf energy minimization and beyond via dual decomposition, PAMI 33 (2011) 531–552.

[29] N. Kumar, A. Berg, P.N. Belhumeur, S. Nayar, Describable visual attributes for face verification and image search, IEEE Trans. Pattern Anal. Mach. Intell. 33 (2011) 1962–1977.
[30] N. Kumar, A.C. Berg, P.N. Belhumeur, S.K. Nayar, Attribute and simile classifiers for face verification, in: IEEE International Conference on Computer Vision (ICCV.
[31] E.G. Learned-Miller, Data driven image models through continuous joint alignment, PAMI 28 (2006) 2006.
[32] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vision 60 (2004) 91–110.
[33] K. Messer, J. Matas, J. Kittler, K. Jonsson, Xm2vtsdb: the extended m2vts database, in: Second International Conference on Audio and Video-based Biometric Person Authentication, pp. 72–77.
[34] A. Mignon, F. Jurie, PCCA: a new approach for distance learning from sparse pairwise constraints, in: IEEE Conference on Computer Vision and Pattern Recognition, France, pp. 2666–2672.
[35] K. Mikolajczyk, C. Schmid, A performance evaluation of local descriptors, IEEE Trans. Pattern Anal. Mach. Intell. 27 (2005) 1615–1630.
[36] H. Nguyen, L. Bai, L. Shen, Local gabor binary pattern whitened pca: a novel approach for face recognition from single image per person, in: M. Tistarelli, M. Nixon (Eds.), Advances in Biometrics, Lecture Notes in Computer Science, vol. 5558, Springer, Berlin Heidelberg, 2009, pp. 269–278.
[37] G. Nicholls, M. Petrou, On multiresolution image restoration, in: International Conference on Pattern Recognition ICPR94, pp. C:63–67.
[38] E. Nowak, Learning visual similarity measures for comparing never seen objects, in: Proceedings of IEEE CVPR.
[39] 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: IEEE Computer Vision and Pattern Recognition.
[40] L. Pishchulin, T. Gass, P. Dreuw, H. Ney, Image warping for face recognition: from local optimality towards global optimization, Pattern Recognit. (2011).
[41] E. Rahtu, J. Heikkilä, V. Ojansivu, T. Ahonen, Local phase quantization for blur-insensitive image analysis, Image Vision Comput. 30 (2012) 501–512.
[42] S. Romdhani, V. Blanz, T. Vetter, Face identification by fitting a 3d morphable model using linear shape and texture error functions, in: ECCV '02-Part IV, Springer-Verlag, London, UK, 2002, pp. 3–19.
[43] 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.
[44] H.J. Seo, P. Milanfar, Face verification using the lark representation, IEEE Trans. Inf. Forensics Secur. 6 (2011) 1275–1286.
[45] A. Shekhovtsov, I. Kovtun, V. Hlavac, Efficient MRF deformation model for non-rigid image matching, Comput. Vis. Image Underst. 112 (2008) 91–99.
[46] T. Sim, S. Baker, M. Bsat, The cmu pose, illumination, and expression (pie) database, in: AFGR, 2002, pp. 46–51.
[47] J.R. del Solar, R. Verschae, M. Correa, Recognition of faces in unconstrained environments: a comparative study, EURASIP J. Adv. Signal Process. (2009) 1–19.
[48] T. Suzuki, Y. Amano, T. Hashizume, Picking the best daisy, Proceedings of SICE Annual Conference 2010, 2010, pp. 2960–2964.
[49] X. Tan, B. Triggs, Enhanced local texture feature sets for face recognition under difficult lighting conditions, in: AMFG, pp. 168–182.
[50] J. Tena, R. Smith, M. Hamouz, J. Kittler, A. Hilton, J. Illingworth, 2D face pose normalisation using a 3d morphable model, in: International Conference on Video and Signal Based Surveillance, pp. 1–6.
[51] E. Tola, V. Lepetit, P. Fua, Daisy: an efficient dense descriptor applied to wide baseline stereo, IEEE Trans. Pattern Anal. Mach. Intell. 32 (2010).
[52] M.A. Turk, A.P. Pentland, Face recognition using eigenfaces, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Proceedings CVPR '91, IEEE Comput. Sco. Press, 1991, pp. 586–591.
[53] M. Wainwright, T. Jaakkola, A. Willsky, Map estimation via agreement on trees: message-passing and linear programming, IEEE Trans. Inf. Theory 51 (2005) 3697–3717.
[54] P. Wang, L. Cam Tran, Q. Ji, Improving face recognition by online image alignment, in: Proceedings of the 18th International Conference on Pattern Recognition – Volume 01, ICPR '06, IEEE Computer Society, Washington, DC, USA, 2006, pp. 311–314.
[55] R. Wang, Z. Lei, M. Ao, S. Li, Bayesian face recognition based on markov random field modeling, in: ICB, 2009, pp. 42–51.
[56] T. Werner, A linear programming approach to max-sum problem: a review, IEEE Trans. Pattern Anal. Mach. Intell. 29 (2007) 1165–1179.
[57] T. Werner, Revisiting the linear programming relaxation approach to gibbs energy minimization and weighted constraint satisfaction, IEEE Trans. Pattern Anal. Mach. Intell. 32 (2010) 1474–1488.
[58] L. Wiskott, J. Fellous, N. Kuiger, C. von der Malsburg, Face recognition by elastic bunch graph matching, PAMI 19 (1997) 775–779.
[59] L. Wiskott, J.M. Fellous, N. Krger, C. von der Malsburg, Face recognition by elastic bunch graph matching, IEEE Trans. Pattern Anal. Mach. Intell. 19 (1997) 775–779.
[60] L. Wolf, T. Hassner, Y. Taigman, Y.: Descriptor based methods in the wild, in: Faces in Real-Life Images Workshop in ECCV. (2008) (b) Similarity Scores based on Background Samples.
[61] L. Wolf, T. Hassner, Y. Taigman, Similarity scores based on background samples, in: Asian Conference on Computer Vision (ACCV).
[62] Y. Ying, P. Li, Distance metric learning with eigenvalue optimization, JMLR 13 (2012) 1–26.
[63] B. Zhang, S. Shan, X. Chen, W. Gao, Histogram of gabor phase patterns (hgpp): a novel object representation approach for face recognition, IEEE Trans. Image Process. 16 (2007) 57–68.
[64] X. Zhang, Y. Gao, M. Leung, Recognizing rotated faces from frontal and side views: an approach toward effective use of mugshot databases, IEEE Trans. Inf. Forensics Secur. 3 (2008) 684–697.