Efficient Processing of MRFs for Unconstrained-Pose Face Recognition

Shervin Rahimzadeh Arashloo
Department of electrical engineering
Faculty of engineering
Urmia university
West Azarbaijan, Urmia, Iran
Sh.Rahimzadeh@urmia.ac.ir

Josef Kittler
CVSSP, department of electronic engineering
Faculty of engineering and physical sciences
University of Surrey
Guildford, Surrey, UK
J.Kittler@surrey.ac.uk

Abstract

The paper addresses the problem of pose-invariant recognition of faces via an MRF matching model. Unlike previous costly matching approaches, the proposed algorithm employs effective techniques to reduce the MRF inference time. To this end, processing is done in a parallel fashion on a GPU employing a dual decomposition framework. The optimisation is further accelerated taking a multi-resolution approach based on the Renormalisation Group Theory (RGT) along with efficient methods for message passing and the incremental subgradient approach. For the graph construction, Daisy features are used as node attributes exhibiting high cross-pose invariance, while high discriminatory capability in the classification stage is obtained via multi-scale LBP histograms. The experimental evaluation of the method is performed via extensive tests on the databases of XM2VTS, FERET and LFW in verification, identification and the unseen pair-matching paradigms. The proposed approach achieves state-of-the-art performance in pose-invariant recognition of faces and performs as well or better than the existing methods in the unconstrained settings of the challenging LFW database using a single feature for classification.

1. Introduction

Despite the significant progress in face recognition technology, recognition under unconstrained settings remains a challenge. The use of face recognition technology is currently hampered by several factors present in real-life images such as pose, illumination and expression changes, occlusion, low resolution *etc.* Among the numerous solutions proposed to deal with these problems, Markov random field modelling has been shown to be a promising approach, especially for handling large face pose variations [3, 4, 55, 36].

In this paper we focus on using MRFs for addressing one of the major problems in recognition of faces [58] *i.e.* pose variation. Pose-invariance is achieved via dense matching of images while illumination invariant representations adopted minimises the unwanted effects of illumination changes which is in contrast to some other approaches which use images of different frequency bands [35]. In order to minimise the adverse effects of the background and unavailability of frontal gallery images on the recognition performance in unconstrained settings, we symmetrise the process of matching two images, *i.e.* first match the first image to the second and then exchange the roles of the images. The procedure is repeated for the horizontally mirrored versions of both images and the final similarity score is taken as the minimum of the distances thus obtained. The proposed approach is inspired by the methods presented in [3, 4], which were shown to have appealing characteristics. Namely, they 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.

Unfortunately, the optimisation and inference over a graphical model still remains both challenging and computationally demanding [29, 54, 27, 13]. 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. The key contribution of the current work is to reduce the processing time of inference in MRF image matching. In this respect, we focus on parallel processing algorithms and show how an optimisation problem for image matching can be reformulated to be solved via the well-known dual decomposition approach [29] which in turn facilitates porting the problem onto a graphical processing unit for efficient implementation. Further speed ups are achieved by adopting a multi-resolution approach based on the renormalisation group theory (RGT) [20]. In addition, distance transform technique [18] is employed for efficient message passing. Last but not least, an incremental subgradient method [9] is used so that the more computationally demanding local updates in the decomposition framework are performed less frequently. While some of the aforementioned techniques have previously been used independently, this is the first time they have been integrated for joint use in a common framework. We show that their consolidation under a common umbrella results in massive speed-up gains. Note that this is also the first time these techniques have been jointly used for face MRF matching.

The current work also supersedes [3, 4] 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.* multi-resolution LBP histograms descriptor is used for classification. This contrasts with many other algorithms combining a number 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 [4, 3] 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 with an emphasis on pose-invariant methods. 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. Efficient message computation, multi-resolution analysis using RGT, the incremental subgradient method and GPU processing are discussed in Section 4. In Section 5, our classifier employing multi-resolution LBPs is introduced. The evaluation of the method in terms of processing time and a comparison to the state-of-the-art face recognition methods on three databases are discussed in section 6. In Section 7, conclusions are drawn.

2. Related work

The various approaches to unconstrained pose-invariant recognition of faces can coarsely be categorized into four major groups based on general concepts of classification. These are multi-view systems, generative methods, discriminative approaches and graph-based algorithms.

The first attempts of generalizing across different poses are represented by the multi-view systems which are direct extensions of the methods operating on the frontal images, storing multiple templates for different poses for each individual. The works in [37, 45] are examples of the methods falling in this category. The second group of methods takes a different approach, *i.e.* it attempts to reconstruct a novel image in a desired pose. The reconstructed image is then either directly or indirectly used for classification. As examples of the methods in this category, one may list [40, 21, 16, 49, 19]. The next group is constituted by discriminative approaches, which do not require input faces to be remapped into a reference pose. In fact, features used are projected into a pose-invariant space in which recognition is performed based on similarity in this space. Similar to the generative methods, for training, images corresponding to different poses or lighting conditions are employed to construct the model [26, 25]. One common drawback of the three mentioned classes of algorithms is that they need multiple images either to construct the model or to be stored as templates. In addition, the generative methods cannot recover atypical features that do not exist in the training set while the multi-view systems may require pose estimation and accurate alignment for example using optical flow. Part-based methods constitute the last category in our classification of pose-invariant methods for face recognition [12]. A successful subset of part-based approaches is the Graph-based methods. In this framework [3, 4, 55, 53, 1], different parts of an object are allowed to be considered independently of other non-neighboring 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 one gallery image per class. The approach proposed here uses a graph-based representation for dense and efficient pixel-wise matching of faces. Once correspondences are established between images, multi-resolution texture features are used for classification taking into account the accurate pixel-wise alignment of face images.

3. Image matching

Two distinct stages exist almost in all object recognition systems: finding correspondences between different parts of objects and then evaluating a similarity criterion for classification. For face recognition this concept has been motivated by the fact that unlike controlled frontal pose conditions, in which only two fiducial points (usually eye coordinates) are sufficient for alignment, for unconstrained faces rotated in-depth, a larger number of point correspondences are needed for effective alignment. Nevertheless, frontal-pose recognition of faces may also benefit from availability of such correspondence information.

3.1. Markov Random Fields for image matching

An MRF 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 between the nodes. The goal

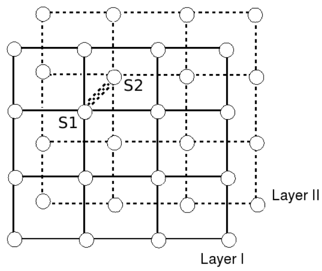

Figure 1. The structures of MRFs used with a sample inter-layer edge connecting the two layers.

is to assign each node a label from a predefined admissible label set $\mathbb{X} = \{1, 2, \ldots, L\}$ subject to contextual constraints such that the energy of the assignment is minimum. When the graphical structure only contains cliques of up to size two, the energy associated with the model is

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

where x is a labelling and θ parameterizes the energy. In the matching model adopted from [44, 3], 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 result in the coordinates of the corresponding block in the target image. The employed model consists of two layers for displacement modelling, one for horizontal and the other for vertical direction. The structure of the MRFs used is depicted in Fig. 1.

The severity of admissible local deformations is controlled by the smoothness prior [3, 4] on each layer of the model. Accordingly, for two neighbouring nodes s and t with states x_s and x_t respectively, the prior is set as

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

where ρ is a normalizing constant controlling the tradeoff between data fidelity and smoothness of the deformation.

Edge-based features are well known for their discriminatory, invariance and repeatability properties. Inspired by the ideas used in SIFT descriptor [30], many other features including geometric blur [8], GLOH histogram [32], SURF [7] and Daisy [50] *etc.* have been proposed and used for recognition. We use Daisy feature [50] for the construction of data term in our model.

There is freedom in choosing the number of radii and directions of samples to form the Daisy feature vector [47]. In this work, the points are sampled in one concentric circle and 8 directions in addition to the centre pixel. The feature vectors obtained at each pixel are normalized to unit length and compared using Euclidean distance. The data term is then defined 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 [13], dual decomposition [29] and TRW-S [27]. 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.

4. Solving the subproblems

In this work we choose each subproblem as a tree in an extreme case, *i.e.* an edge along with the two end nodes which makes our original problem decompose into a large number of subproblems. This choice is driven by the large number of streaming processors available in today's GPUs. 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. As a result, one needs to resort to more computationally efficient methods.

4.1. Efficient message passing

Two kinds of edges exist in the 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 each edge [52]. 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 [18] 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 [18] 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. As a result, one needs to perform an exhaustive search over these edges incurring a computational complexity of $O(L^2)$. Fortunately, in the dual decomposition framework, one may reduce the frequency of some updates via an incremental approach [9]. The idea is that in cases where a full subgradient update is costly, one hopes to make progress towards the optimum by making much cheaper incremental steps. We update the less computationally demanding updates of intra-layer edges at every iteration but update the inter-layer edges less frequently, *e.g.* once in every n iterations, where n is determined empirically to produce the best results.

4.3. Multi-resolution analysis using RGT

Processing the model only in the finest scale (pixel level) is computationally demanding and highly susceptible to noise and local minima. Multi-resolution analysis has been proposed to avoid such problems [20, 11, 28]. In this work we apply a multi-resolution scheme based on renormalisation group transform (RGT) [20, 11]. Usually one is only interested in preserving the maxima of the MRF probability distribution. In these cases, a potential-based coarsening technique, supercoupling transform [11], which is known to map the mode of the original distribution to the mode of the coarsened distribution is employed. For the transformation of the solution obtained in one level to a finer level, similar to [11, 4], we use a block-flat assumption. Using the block flat assumption it is shown [4] that the parameters of the energy should be chosen so that

$$\theta'_s(x'_s) = \sum_{i=1}^{4} \theta_{s_i}(x'_s) \tag{3}$$

and

$$\rho' = 2\rho \tag{4}$$

where θ' and ρ' are the parameters of the model in a coarse scale and θ and ρ are the parameters in the next finer level. We process the images in 4 scales, the finest being the pixel level.

4.4. Graphical Processing Units (GPUs)

Graphical processing hardware provides a massive number of processors which can operate in parallel together to speed up computationally intensive tasks [46, 14]. The most widely used GPUs are offered by the NVIDIA corporation and are released with the CUDA architecture. The CUDA architecture provides a hierarchy of threads grouped into block of threads and grid of blocks. Each thread operates in parallel on all computing resources in the GPU. As noted earlier, the dual decomposition method consists of two main stages: solving the sub-problems and updating the Lagrange multipliers controlling their interaction. Hence, if one solves the subproblems in parallel, large speed up gains can be achieved. In the CUDA architecture, a kernel is a group of instructions operated in parallel on independent data, *i.e.* each subproblem separately. In our implementation of the dual decomposition method, there are two kinds of kernels: one kernel is developed to solve for the inter-layer edges via an exhaustive search in an incremental fashion while the other solves for the intra-layer edges using efficient message passing technique. Once all the subproblems are solved, the next step is to update the Lagrange multipliers. These two stages are iterated in a loop.

5. Classification

In [4, 3], the energy of the match consisting of the data fidelity plus the geometric distortion is normalised and used as a similarity criterion for hypothesis selection. In this work we employ effective texture descriptors but do not use the shape information explicitly. This is due to the fact that a simple planar transformation (*e.g.* projective transformation) is inadequate to model the nonlinear complex structure of faces in poses deviating largely from frontal. Motivated by the work in [15, 49] we use multi-resolution LBPs for texture representation. The proposed method is different in having the ability to handle extreme poses in a discriminative way rather than generative. Once features are extracted from the images, a nearest neighbour classifier is used for decision making.

Our texture representation is based on the local binary pattern (LBP) [34]. The original LBP operator assigns a label to every pixel of an image by comparing its 3×3-neighbourhood with the value of the pixel under consideration and treating the results as a binary number. As long as the intensity order of the pixels in a neighbourhood is preserved, LBPs are known to be unaffected by monotonic gray scale changes.

One extension to this operator proposes to use neighbourhoods of different sizes in order to deal with textures at different resolutions. The second extension to the original operator is the notion of *uniform patterns*. Before applying the LBP operators, we normalize the face images using an effective photometric normalisation scheme [48]. For face description, the template images are partitioned into $n \times n$ 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 and concatenated to form a larger vector. A PCA transformation is applied to reduce the dimensionality of the feature vectors at each region. The resulting feature vectors are then compared and a match score is produced for each pair of regions using the cosine similarity metric. Taking a classifier fusion approach, the final similarity score of two faces is then defined as

$$Sim(I, I') = \sum_j \frac{d_j d'_j}{\|d_j\| \|d'_j\|} \tag{5}$$

where $Sim(I, I')$ stands for the similarity of the two images I and I'. d_j is the feature vector of region j in image I after PCA transformation and d'_j denotes the feature vector of its corresponding region in image I'.

Table 1. Speed up gains achieved via different techniques compared to the method in [3].

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

Table 2. Comparison of the performance of the proposed method to the state-of-the-art methods on the XM2VTS database.

Method	3D correc.[49]	face matching[3]	current work
EER	7.12	4.85	**4.27**

6. Experimental evaluation

For the evaluation of the proposed face matching method we first compare the proposed MRF matching method with those of [2, 3] in terms of computational time. We then compare the performance of the proposed face recognition method on different databases with other methods in verification, identification and the unseen pair matching paradigms.

6.1. Gains in running time

In order to compare the running times of different methods, we use the original source codes of [3] leaving the parameters unchanged. In this experiment, a template image of size 112×128 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 1 reports the effects of different techniques used in the proposed method. 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 [3]. 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 multi-resolution analysis accounts for $\sim 500\%$ efficiency. As noted earlier, the efficient message computation algorithm can be instrumental in reducing computational complexity of the algorithm. From the table, it is observed that the employed technique accelerates the method $\sim 1.4\times$. For the incremental subgradient approach, we update interlayer edges 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 techniques. It is observed that the proposed matching method is more than 200 times faster than the method of [3]. For the proposed method it takes almost 1.4 seconds to match a template image of size 112×128 to the target image compared to the method of [3] which it takes more than 5 minutes for the same task.

6.2. Verification Test on the XM2VTS database

In the XM2VTS rotation data set [31] 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:

$$FA = EI/I * 100\%, \qquad FR = EC/C * 100\% \tag{6}$$

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 2 reports the equal error rates obtained on the XM2VTS dataset using the proposed approach compared to some other methods. These include the method in [49], where the authors use a 3D morphable model for geometrically normalizing the rotated images and then use LBP histograms in the 2D geometrically normalized images. In [3] 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.

6.3. Identification test on the FERET database

Next, we evaluate our method on the rotation shots of the FERET database [38] *i.e.* the ***b*** series in an identification scenario. This part of the database consists of 200 subjects captured under 9 different yaw angles ranging from nearly $-60°$ to $+60°$. We use the ba image of each subject (almost frontal) as the gallery image and all the rest as test images. Table 3 reports the correct identification rates obtained on this data. The results of some other methods are also included for comparison. From the table, it can be concluded that our method outperforms all other approaches almost in all poses on this subset of the FERET database.

6.4. Unseen pair matching on the LFW database

Next we evaluate the performance of the proposed approach on the LFW dataset [23] which contains 13,233 images of 5,749 subjects collected from the news articles on the web and as a result includes real world variations in facial images such as pose, illumination, expression, occlusion, low resolution *etc.*. 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. There are three evaluation protocols on this database: the image unrestricted setting, the image restricted setting and the unsupervised setting. The most dif-

Table 3. Comparison of the performance of the proposed approach to the state-of-the-art methods on the FERET database.

Pose	bi	bh	bg	bf	be	bd	bc	bb
Horizontal deviation angle	-60°	-40°	-25°	-15°	$+15^\circ$	$+25^\circ$	$+40^\circ$	$+60^\circ$
PAN [19]	52.0	78.5	91.5	98.5	97.0	93.0	81.5	44.0
3D Morph. Model [10]	90.7	95.4	96.4	97.4	99.5	96.9	95.4	**94.8**
Prob. Stack Flow [5]	~ 43	~ 65	~ 89	~ 95	~ 93	~ 82	~ 57	~ 34
3D Pose Norm. [6]	na	90.5	98	98.5	97.5	97.0	91.5	na
current work	**92.0**	**98.5**	**99.5**	**100.0**	**99.5**	**99.0**	**99.5**	91.0

ficult 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 setting.

6.4.1 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 LFW-a version of the LFW database[57]. We crop face images closely to minimise the effects of background samples. In this experiment, the multi-resolution LBP histograms are computed from each region of the first image and compared against the corresponding region in the second image using χ^2 distance measure. As a result, our method is tested in a *training free* scenario. In order to minimise the effects of the background and unavailability of frontal gallery images on the recognition performance, we first match the first image to the second and then exchange the roles of first and second image and match them again. The procedure is repeated for the horizontally mirrored versions of both images. The final similarity measure is considered as the minimum of the distances thus obtained between a pair of images. Table 4 reports the results under this setting. The previous best results under this setting are 73.40% and 86.20% using LHS [43] and LQP [24] features respectively. We achieve a mean accuracy of 80.08% indicating a large 7.85% improvement over the previous second best result.

Table 4. Comparison of the performance of the proposed approach to the state-of-the-art methods on the LFW database in the unsupervised setting.

Method	$\mu \pm S_E$
SD-MATCHES, 125×125, aligned [17]	0.6410 ± 0.0062
H-SX-40, 81×150, aligned [17]	0.6945 ± 0.0048
GJD-BC-100, 122×225, aligned [17]	0.6847 ± 0.0065
LARK unsupervised, aligend [42]	0.7223 ± 0.0049
LHS, aligned [43]	0.7340 ± 0.0040
I-LQP, aligned [24]	0.8620 ± 0.0046
MRF-MLBP(current work), aligned	0.8008 ± 0.0013

Table 5. Comparison of the performance of the proposed approach to the state-of-the-art methods on the LFW database in the image restricted setting (strict LFW, no outside training data used).

Method	$\mu \pm S_E$
Eigenfaces, original [51]	0.6002 ± 0.0079
Nowak, original [33]	0.7245 ± 0.0040
Nowak, funneled [33]	0.7393 ± 0.0049
Hybrid descriptor-based, funneled [56]	0.7847 ± 0.0051
3×3 Multi-region Histograms(1024) [41]	0.7295 ± 0.0055
Pixels/MKL, funneled [39]	0.6822 ± 0.0041
V1-like/MKL, funneled [39]	0.7935 ± 0.0055
MRF-MLBP(current work), funneled	0.7908 ± 0.0014

6.4.2 Image restricted setting

Next, we evaluate the proposed method in the image restricted setting where the training data is provided as ”same” or ”not same” for each pair without providing subject identities. In this setting again the χ^2 distance measure is used as the dissimilarity measure. For this experiment, we use a second version of LFW images called funneled obtained using the aligning algorithm of [22]. 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 use mirrored images as well as exchanging the roles of the two images in a pair and use the minimum distance obtained as the dissimilarity measure. Table 5 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 other single feature-based methods. 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.

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 [3, 4] 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 were employed to obtain further efficiency gains of the proposed approach. The combination of these techniques was shown to result in a more than two hundred orders of magnitude speed up as compared to the baseline methods. In order to increase the efficacy of the approach, multi-resolution Daisy features were used to achieve invariance against deformations and lighting changes. Finally, for classification multi-resolution 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, also demonstrated an impressive face matching accuracy of the proposed approach.

8. Acknowledgement

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

The work was partially supported from Iranian national elite foundation (NEF) research grant.

References

[1] S. Arashloo, J. Kittler, and W. Christmas. Facial feature localization using graph matching with higher order statistical shape priors and global optimization. In *Biometrics: Theory Applications and Systems (BTAS), 2010 Fourth IEEE International Conference on*, pages 1–8, 2010. 2

[2] S. R. Arashloo and J. Kittler. *Pose-invariant face matching using mrf energy minimization framework*, page 5669. Springer, 2009. 5

[3] S. R. Arashloo and J. Kittler. Energy normalization for pose-invariant face recognition based on mrf model image matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(6):1274–1280, 2011. 1, 2, 3, 4, 5, 7

[4] S. R. Arashloo, J. Kittler, and W. J. Christmas. Pose-invariant face recognition by matching on multi-resolution mrfs linked by supercoupling transform. *Comput. Vis. Image Underst.*, 115(7):1073–1083, July 2011. 1, 2, 3, 4, 7

[5] A. Ashraf, S. Lucey, and T. Chen. Learning patch correspondences for improved viewpoint invariant face recognition. In *CVPR 2008*, pages 1–8, June 2008. 6

[6] A. Asthana, T. K. Marks, M. J. Jones, K. H. Tieu, and M. Rohith. Fully automatic pose-invariant face recognition via 3d pose normalization. *Computer Vision, IEEE International Conference on*, 0:937–944, 2011. 6

[7] H. Bay, T. Tuytelaars, and L. V. Gool. Surf: Speeded up robust features. In *In ECCV*, pages 404–417, 2006. 3

[8] A. C. Berg and J. Malik. Geometric blur for template matching. In *CVPR 2001*, pages 607–614, 2001. 3

[9] D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific, 2nd edition, Sept. 1999. 2, 3

[10] V. Blanz and T. Vetter. Face recognition based on fitting a 3d morphable model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25:2003, 2003. 6

[11] M. Bober, M. Petrou, and J. Kittler. Nonlinear motion estimation using the supercoupling approach. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20(5):550–555, 1998. 4

[12] K. Bonnen, B. Klare, and A. K. Jain. Component-based representation in automated face recognition. *IEEE Transactions on Information Forensics and Security*, 8(1):239–253, 2013. 2

[13] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(11):1222–1239, Nov. 2001. 1, 3

[14] A. Brunton, C. Shu, and G. Roth. Belief propagation on the gpu for stereo vision. In *Proceedings of the The 3rd Canadian Conference on Computer and Robot Vision*, CRV '06. IEEE Computer Society, 2006. 4

[15] C.-H. Chan. *Multi-scale Local Binary Pattern Histogram for Face Recognition*. PhD thesis, 2008. 4

[16] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. *PAMI, IEEE Transactions on*, 23(6):681–685, Jun 2001. 2

[17] J. R. del Solar, R. Verschae, and M. Correa. Recognition of faces in unconstrained environments: a comparative study. *EURASIP J. Adv. Signal Process*, pages 1–19, 2009. 6

[18] P. Felzenszwalb and D. Huttenlocher. Efficient belief propagation for early vision. I:261–268, 2004. 1, 3

[19] H. Gao, H. K. Ekenel, and R. Stiefelhagen. Pose normalization for local appearance-based face recognition. In *Proceedings of the Third International Conference on Advances in Biometrics*, ICB '09, pages 32–41, Berlin, Heidelberg, 2009. Springer-Verlag. 2, 6

[20] B. Gidas. A Renormalization Group Approach to Image Processing Problems. *IEEE Trans. Pattern Anal. Mach. Intell.*, 11(2):164–180, Feb. 1989. 1, 4

[21] D. Gonzalez-Jimenez and J. Alba-Castro. Toward pose-invariant face recognition through point distribution models and facial symmetry. *Information forensics and security*, 2(3):413–429, 2007. 2

[22] G. B. Huang, V. Jain, and E. Learned-Miller. Unsupervised Joint Alignment of Complex Images. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8, 2007. 6

[23] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, October 2007. 5

[24] S. U. Hussain, T. Napoleon, and F. Jurie. Face Recognition using Local Quantized Patterns. In *British Machive Vision Conference*, page 11 pages, Guildford, United Kingdom, 2012. 6

[25] T. Kanade and A. Yamada. Multi-subregion based probabilistic approach toward pose-invariant face recognition, 2003. ISCIRA, pages 954-959. 2

[26] T. Kim and J. Kittler. Locally linear discriminant analysis for multimodally distributed classes for face recognition with a single model image. *PAMI*, 27(3):318–327, 2005. 2

[27] V. Kolmogorov. Convergent Tree-Reweighted Message Passing for Energy Minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(10):1568–1583, Oct. 2006. 1, 3

[28] N. Komodakis. Towards More Efficient and Effective LP-Based Algorithms for MRF Optimization. In *ECCV 2010*, volume 6312 of *Lecture Notes in Computer Science*, chapter 38, pages 520–534. Springer Berlin / Heidelberg, 2010. 4

[29] N. Komodakis, N. Paragios, and G. Tziritas. MRF energy minimization and beyond via dual decomposition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33:531–552, 2011. 1, 3

[30] D. G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110, Nov. 2004. 3

[31] K. Messer, J. Matas, J. Kittler, and K. Jonsson. Xm2vtsdb: The extended m2vts database. In *In Second International Conference on Audio and Video-based Biometric Person Authentication*, pages 72–77, 1999. 5

[32] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 27(10):1615–1630, 2005. 3

[33] E. Nowak. Learning visual similarity measures for comparing never seen objects. In *Proc. IEEE CVPR*, 2007. 6

[34] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *PAMI, IEEE Transactions on*, 24(7):971–987, Jul 2002. 4

[35] N. Osia and T. Bourlai. Holistic and partial face recognition in the mwir band using manual and automatic detection of face-based feature, 2012. IEEE International Conference on Technologies for Homeland Security, , Waltham, MA, U.S.A. 1

[36] B.-G. Park, K.-M. Lee, and S.-U. Lee. Face recognition using face-ARG matching. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(12):1982–1988, 2005. 1

[37] A. Pentland, B. Moghaddam, and T. Starner. View-based and modular eigenfaces for face recognition, 1994. CVPR, pages 84-91. 2

[38] J. P. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The FERET Evaluation Methodology for Face-Recognition Algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, 2000. 5

[39] N. Pinto, J. J. DiCarlo, and D. D. Cox. How far can you get with a modern face recognition test set using only simple features? In *IEEE Computer Vision and Pattern Recognition*, 2009. 6

[40] S. Prince, J. Warrel, J. Elder, and F. Felisberti. Appearance based face recognition and light fields. *PAMI*, 26(4):449–465, 2004. 2

[41] C. Sanderson and B. C. Lovell. Multi-Region Probabilistic Histograms for Robust and Scalable Identity Inference. In M. Tistarelli and M. S. Nixon, editors, *ICB*, volume 5558 of *Lecture Notes in Computer Science*, pages 199–208. Springer, 2009. 6

[42] H. J. Seo and P. Milanfar. Face verification using the lark representation. *IEEE Transactions on Information Forensics and Security*, 6(4):1275–1286, 2011. 6

[43] G. Sharma, S. Ul Hussain, and F. Jurie. Local Higher-Order Statistics (LHS) for Texture Categorization and Facial Analysis. In *ECCV - European Conference on Computer Vision*, Florence, Italie, Aug. 2012. 6

[44] A. Shekhovtsov, I. Kovtun, and V. Hlavac. Efficient MRF deformation model for non-rigid image matching. *Computer Vision and Image Understanding*, 112:91–99, 2008. 3

[45] R. Singh, M. Vatsa, A. Ross, and A. Noore. A mosaicing scheme for pose-invariant face recognition. *Systems, man and Cybernetics*, 37(5):1212–1225, 2007. 2

[46] P. Strandmark, F. Kahl, and T. Schoenemann. Parallel and distributed vision algorithms using dual decomposition. *Comput. Vis. Image Underst.*, 115(12):1721–1732, Dec. 2011. 4

[47] T. Suzuki, Y. Amano, and T. Hashizume. Picking the best daisy. *SICE Annual Conference 2010 Proceedings of*, pages 2960–2964, 2010. 3

[48] X. Tan and B. Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. In *AMFG*, pages 168–182, 2007. 4

[49] J. Tena, R. Smith, M. Hamouz, J. Kittler, A. Hilton, and J. Illingworth. 2d face pose normalisation using a 3d morphable model. In *International Conference on Video and Signal Based Surveillance*, pages 1–6, September 2007. 2, 4, 5

[50] E. Tola, V. Lepetit, and P. Fua. Daisy: An efficient dense descriptor applied to wide baseline stereo. *IEEE TRANS. PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, 32(5), 2010. 3

[51] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 586–591. IEEE Comput. Sco. Press, 1991. 6

[52] M. Wainwright, T. Jaakkola, and A. Willsky. Map estimation via agreement on trees: message-passing and linear programming. *Information Theory, IEEE Transactions on*, 51(11):3697–3717, Nov. 2005. 3

[53] R. Wang, Z. Lei, M. Ao, and S. Li. Bayesian face recognition based on markov random field modeling, 2009. ICB, pages 42-51. 2

[54] T. Werner. Revisiting the linear programming relaxation approach to gibbs energy minimization and weighted constraint satisfaction. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 32(8):1474–1488, August 2010. 1

[55] L. Wiskott, J. Fellous, N. Kuiger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *PAMI*, 19(7):775–779, 1997. 1, 2

[56] L. Wolf, T. Hassner, and Y. Taigman. Y.: Descriptor based methods in the wild. In *In: Faces in Real-Life Images Workshop in ECCV. (2008) (b) Similarity Scores based on Background Samples*. 6

[57] L. Wolf, T. Hassner, and Y. Taigman. Similarity Scores based on Background Samples. In *Asian Conference on Computer Vision (ACCV)*, 2009. 6

[58] W. Zhao, R. Chellappa, A. Rosenfeld, and P. J. Phillips. Face Recognition: A Literature Survey. *ACM Computing Surveys*, pages 399–458, 2003. 1