Matching Larger Image Areas for Unconstrained Face Identification

Gaston, J., Ji, M., & Crookes, D. (2018). Matching Larger Image Areas for Unconstrained Face Identification. *IEEE Transactions on Cybernetics*, 49(8), 3191-3202. <https://doi.org/10.1109/TCYB.2018.2846579>

Published in:
IEEE Transactions on Cybernetics

Document Version:
Peer reviewed version

Queen's University Belfast - Research Portal:
[Link to publication record in Queen's University Belfast Research Portal](#)

Publisher rights
© 2018 IEEE. This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

General rights
Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Matching Larger Image Areas for Unconstrained Face Identification

Jack Gaston, Ji Ming, *Member, IEEE*, Danny Crookes, *Senior Member, IEEE*

Abstract—Many approaches to unconstrained face identification exploit small patches which are unaffected by distortions outside of their locality. A larger area usually contains more discriminative information, but may be unidentifiable due to local appearance changes across its area, given limited training data. We propose a novel *block*-based approach, as a complement to existing *patch*-based approaches, to exploit the greater discriminative information in larger areas, while maintaining robustness to limited training data. A testing block contains several neighboring patches, each of a small size. We identify the matching training block by *jointly* estimating all the matching patches, as a means of reducing the uncertainty of each small matching patch with the addition of the neighboring patch information, without assuming additional training data. We further propose a multi-scale extension in which we carry out block-based matching at several block sizes, to combine complementary information across scales for further robustness. We have conducted face identification experiments using three datasets, the constrained GT dataset to validate the new approach, and two unconstrained datasets, LFW and UFI, to evaluate its potential for improving robustness. The results show that the new approach is able to significantly improve over existing patch-based face identification approaches, in the presence of uncontrolled pose, expression and lighting variations, using small training datasets. It is also shown that the new block-based scheme can be combined with existing approaches to further improve performance.

Index Terms—Block-based correlation, face identification, multi-patch blocks, multi-scale combination, patch-based image processing.

I. INTRODUCTION

Early approaches to face recognition like PCA [1] and LDA [2] for face recognition processed images holistically. However, for facial images with variable conditions, it was found that performance could be improved by dividing the image into smaller regions to process *independently* [3]. This (small) patch-based method has been adopted in most current approaches to unconstrained face recognition. A small patch isolates the local facial structure from variations outside its locality; independent patch matching allows similar local facial structures to be found in different images, even if they holistically differ. Thus, the patch-based method reduces the amount of training data required, compared to modeling larger facial areas which contain potentially greater variations in pose, expression and illumination. Popular patch-based approaches include sparse representation (SR), in which each testing patch is reconstructed *independently* of the other testing patches as a sparse linear combination of the training patches [4], [5].

More recently, a patch matching scheme based on normalized cross correlation (NCC), or Pearson's correlation coefficient, is described in [6]. For each testing patch, the NCC method identifies the closest matching training patch, *independently* of the other testing patches, based on maximum patch-sized correlation. To achieve robustness to unconstrained local pose or expression variations, the search for each matching training patch is conducted within a window around the testing patch's location, assuming variable alignment. The latest example of patch-based image recognition is deep convolutional neural nets (CNNs) (see, for example, [7]–[13]), in which each neuron (i.e., a feature extractor) is connected typically only to a small local region (receptive field) of the input image. This appears to be essential for reducing the number of parameters that must be trained [9], [14].

Patch-based image processing is also common for dealing with variable lighting conditions. Because illumination varies slowly in natural images, it can be treated as piecewise constant [15]–[18]. Therefore, globally variable lighting can be treated as constant lighting within small patches. For example, illumination normalization filters such as self quotient images (SQIs) [19] and Gradientfaces [20] are applied in small patches; local binary patterns (LBPs) [21], [22], and its variants (e.g., [23]–[25]), are also used to code high frequency variations which represent facial structure in local regions (e.g. patches), while discarding the low frequency lighting. The NCC metric mentioned above is also invariant to linear transforms in image amplitude, and therefore provides an illumination-invariant match score between two patches with constant lighting difference [6].

However, while the independent matching of small patches has the advantages of being easier to train and to normalize for lighting variation, small patches can be limited in their discriminative ability, which makes their matching prone to errors. Matching larger facial areas will improve discrimination, and hence the matching accuracy, but larger areas may be unidentifiable due to greater local appearance variations, given limited training data. In the past, several approaches have been taken to try to solve this problem. Mainly, these approaches include: 1) to include, or synthesize, diverse conditions or adaptive data in the training data to reduce the training and testing mismatch (see, for example, [4], [11]–[13], [26]–[31]), 2) to pre-align the images to alleviate the pose variation, by estimating the pose of an input image by fitting local patch-based features to 2-D or 3-D models (see, for example, [11]–[13], [32]–[34]), and 3) to use multi-scale approaches to combine complementary information at several patch-sizes to help improve discrimination (see, for example, [35]–[42]).

The authors are with the Institute of Electronics, Communications and Information Technology Queen's University Belfast, Belfast BT3 9DT, UK. (e-mail: jgaston06@qub.ac.uk; j.ming@qub.ac.uk; d.crookes@qub.ac.uk)

Specifically, multi-scale extensions have concatenated multi-scale LBPs, and derived sparse representations over several dictionaries of patches, at several patch-sizes. Multi-scale collaborative representations have also been proposed. More recent subspace-based approaches attempt to derive a locality-preserving or part-based representation of the image, which still preserves the global geometric structure of the data, for robustness to corruptions in the data [43]–[47].

In this paper, we propose a new approach, as a complement to existing approaches, towards solving the problem of low discriminative ability with small image patches. Specifically, we study the problem of enlarging the matching areas, from the usual small patches to some larger regions, as a means of increasing the discrimination for unconstrained face image comparison, *without* necessarily requiring extra training data. We call the larger matching regions *blocks*. A block is composed of a number of neighboring patches which are matched *jointly*, i.e., to capture the cross-patch dependence in a larger facial area for more discriminative matching. It might be expected that, to model the larger blocks, with potentially greater variations, we may require more training data than usually required for modeling smaller patches. In this paper, we show that this is not necessarily the case. We do not assume additional training data, but present a novel algorithm that forms the matching training blocks by combining suitable small training patches from the available training data, to overcome the data sparsity problem. Our new approach extends the existing approaches, mentioned above, and the largest matching area (LMA) approach [18] (by two of the present authors), by effectively breaking the patch-size barrier. That is, the new approach effectively allows blocks of any size, in theory up to whole images, to be matched holistically with limited training data. In this paper, we demonstrate our new approach by using a simple metric, NCC, for comparing images. We show that the new approach improves over conventional approaches under similar training conditions, based on three face image datasets, two being considered to contain unconstrained face images.

This paper is a substantial expansion of our preliminary study described in [48]. The expansions include an expanded theoretical study, including a new novel semantic constraint and an efficient implementation algorithm, and extensive experimental investigations across constrained, unconstrained and normalized image conditions to evaluate the block-based approach’s ability to improve identification accuracy by exploiting the greater discriminative information in larger image regions. A more detailed explanation and analysis of our algorithm is also given.

The remainder of the paper is organized as follows. Section II introduces the new block-based NCC approach to face image identification. This includes the formulation of block matching as a constrained maximization problem, and the extension to multi-scale block matching. Section III focuses on implementation and details a computationally efficient iterative algorithm to solve the constrained optimization problem for block matching. Section IV presents experiments which evaluate the new approach and compare it with existing approaches. Finally, conclusions are presented in Section V.

Fig. 1. Illustration of the proposed idea. (a) The selection of 4 matching training patches for a testing block of size 2 patches by 2 patches. The matching training patches are found *jointly*, each within a search window, indicated by dotted lines, to allow for certain local displacement to model unseen local variations, and also to maintain a semantic constraint between the matching patches. (b) A real example for matching a face block with 2×2 partially overlapped patches between two images with pose variation.

II. BLOCK-BASED FACE MATCHING

A. Block Correlation

Our idea for improving face matching accuracy through matching larger image areas holistically can be illustrated with Fig. 1, in which Fig. 1(a) illustrates our algorithm, and Fig. 1(b) shows a real example for face matching. Assume that we compare a testing image against a training image of a person, and that the testing image is divided into small patches located by the patch indexes $k_1, k_2, \dots$. For clarity, in Fig. 1(a) we assume that there is no overlap between adjacent patches. But in practice, to improve smoothness, adjacent patches are normally partially overlapped. Conventional approaches tend to process these patches independently. In our new approach, we aim to match a number of neighboring patches *jointly*, which we call *block matching*, where block size is defined by the number of patches in a block. For example, in Fig. 1(a) we show the match of a testing *block* of 2 patches by 2 patches, k_1, k_2, k_3 and k_4 . By considering a larger testing block area holistically we will normally gain greater discriminative ability, and hence reduce uncertainty of the individual matching training patches, compared to independent patch matching. However, because of the potential pose/expression variations between unconstrained images, large, perfectly matching blocks may not exist given limited training data. We assume that the unseen local variations can be modeled, approximately, by allowing for certain local displacement of the individual matching training patches. The displacement should maintain a similar geometric relationship to the corresponding testing patches, as a semantic constraint on valid facial structure, as Fig. 1(a) illustrates. Fig. 1(b) shows an example of the effect of the algorithm, for matching a face block with 2×2 partially overlapped patches between two images with pose variation. In Fig. 1(b), the left-hand side of the nose is partially obscured

in the testing image. Its visible texture covers a smaller region than in the training image, due to pose variation. The selected matching patches are therefore overlapping, recovering the applicable facial texture, accounting for pose changes. When multiple training images are available for a person, the different training patches forming a particular matching block can be selected from different training images, subject to the same geometric semantic constraint. Using this method, we effectively form the matching training blocks by exploiting the full range of available training data conditions, thereby reducing the requirement for extra training data for modeling the potentially greater variations between large blocks. In principle, the number of patches in a block can increase up to the whole testing image.

In this paper, we extend the NCC metric from identifying independent matching patches [6] to identifying matching *blocks*, i.e., to identifying all the matching patches *jointly* within each block, for unconstrained training and testing images. Let $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_K)$ represent a testing image divided into K small, partially overlapped patches, where $\mathbf{y}_k$ is the k th patch, with origin (top-left corner coordinates) (i_k, j_k) . For simplicity, we assume that all patches have the same size: $I \times J$ pixels. We apply the same representation to the training images such that the training image of a person, $\mathbf{X}$, has an expression $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_L)$, where $\mathbf{x}_l$ is the l th $I \times J$ -sized patch. The NCC metric for comparing a single testing patch $\mathbf{y}_k$ and a single training patch $\mathbf{x}_l$ can be expressed as:

$$R(\mathbf{y}_k, \mathbf{x}_l) = \frac{\sum_{i,j} [y_k(i,j) - \mu(\mathbf{y}_k)][x_l(i,j) - \mu(\mathbf{x}_l)]}{\sigma(\mathbf{y}_k)\sigma(\mathbf{x}_l)} \quad (1)$$

where $y_k(i,j)$ denotes a pixel at (i,j) inside patch $\mathbf{y}_k$, and the sum is taken over all pixels within the patch, with $\mu(\mathbf{y}_k)$ being the mean pixel value of patch $\mathbf{y}_k$, and $\sigma(\mathbf{y}_k)$ being the zero-mean Euclidean norm of patch $\mathbf{y}_k$, i.e., $\sigma^2(\mathbf{y}_k) = \sum_{i,j} [y_k(i,j) - \mu(\mathbf{y}_k)]^2$. The same definitions apply to the training patch $\mathbf{x}_l$ with mean pixel value $\mu(\mathbf{x}_l)$ and zero-mean Euclidean norm $\sigma(\mathbf{x}_l)$. In [6], face identification is performed by identifying a matching training patch for each testing patch $\mathbf{y}_k$, independently of the other testing patches, by maximizing $R(\mathbf{y}_k, \mathbf{x}_l)$ over the appropriate training patch candidates $\mathbf{x}_l$ for each person to be identified.

When the two matching patches $\mathbf{y}_k$ and $\mathbf{x}_l$ are small, the NCC $R(\mathbf{y}_k, \mathbf{x}_l)$ defined above may or may not offer much advantage over other types of distance or similarity metrics that have been used for image comparison. However, when $\mathbf{y}_k$ and $\mathbf{x}_l$ are very large, its advantage could become overwhelming. For example, in Appendix A we provide a theory which shows that, when comparing two very large patches, the NCC should become immune to any independent additive noise but depends only on the two patches being compared. This theory has motivated our research in this paper. We focus on scaling up the matching facial areas to improve robustness to variations or corruption etc., for face recognition.

Assuming that there is insufficient training data to include appropriate matching estimates of all possible very large patches, we consider block-based NCC, for matching blocks of neighbouring patches jointly as a means of scaling up the matching facial areas to improve robustness. Let $\mathbf{Y}_{k_1, k_Q} =$

$(\mathbf{y}_{k_1}, \mathbf{y}_{k_2}, \dots, \mathbf{y}_{k_Q})$ represent a testing block with Q patches from patch k_1 to k_Q (block $\mathbf{Y}_{1,K}$ contains all patches in the image and thus is equivalent to the whole image $\mathbf{Y}$), and let $\mathbf{X}_{l_1, l_Q} = (g_{l_1}\mathbf{x}_{l_1}, g_{l_2}\mathbf{x}_{l_2}, \dots, g_{l_Q}\mathbf{x}_{l_Q})$ represent a possible matching training block, formed over the training data of some person, where the individual matching training patch locations l_q ($q = 1, 2, \dots, Q$) may be a nonlinear function of the corresponding testing patch locations k_q to account for the potential nonlinear local displacement between the training and testing images. In $\mathbf{X}_{l_1, l_Q}$, g_{l_q} is a constant multiplicative gain factor for each local training patch, to compensate for the lighting difference between the matching training and testing patches. As discussed earlier, for each small patch, we can assume that the lighting difference, g_{l_q} , is a constant, but across the patches g_{l_q} can be variable to match the uneven lighting condition over the larger testing block. This is identical to a piece-wise constant lighting model [15]–[18]. Also, as mentioned earlier, when multiple training images are available for a person, the different training patches $\mathbf{x}_{l_q}$ can come from different training images of the person to simulate arbitrary unseen testing blocks. The NCC for comparing the two blocks can be written as:

$$\begin{aligned} R(\mathbf{Y}_{k_1, k_Q}, \mathbf{X}_{l_1, l_Q}) &= R(\mathbf{y}_{k_1} \mathbf{y}_{k_2} \dots \mathbf{y}_{k_Q}, g_{l_1}\mathbf{x}_{l_1} g_{l_2}\mathbf{x}_{l_2} \dots g_{l_Q}\mathbf{x}_{l_Q}) \\ &= \frac{\sum_{q=1}^Q \sum_{i,j} [y_{k_q}(i,j) - \mu(\mathbf{Y}_{k_1, k_Q})][g_{l_q}x_{l_q}(i,j) - \mu(\mathbf{X}_{l_1, l_Q})]}{\sigma(\mathbf{Y}_{k_1, k_Q})\sigma(\mathbf{X}_{l_1, l_Q})} \\ &= \frac{\sum_{q=1}^Q g_{l_q} \sum_{i,j} y_{k_q}(i,j) x_{l_q}(i,j) - T\mu(\mathbf{Y}_{k_1, k_Q})\mu(\mathbf{X}_{l_1, l_Q})}{\sigma(\mathbf{Y}_{k_1, k_Q})\sigma(\mathbf{X}_{l_1, l_Q})} \end{aligned} \quad (2)$$

where $T = Q \times I \times J$ is short-hand notation for the total number of pixels in each block, $\mu(\mathbf{X}_{l_1, l_Q})$ and $\sigma(\mathbf{X}_{l_1, l_Q})$ are the global mean pixel value and zero-mean Euclidean norm of training block $\mathbf{X}_{l_1, l_Q}$, respectively, i.e.,

$$\mu(\mathbf{X}_{l_1, l_Q}) = \frac{1}{T} \sum_{q=1}^Q g_{l_q} \sum_{i,j} x_{l_q}(i,j) \quad (3)$$

$$\sigma^2(\mathbf{X}_{l_1, l_Q}) = \sum_{q=1}^Q g_{l_q}^2 \sum_{i,j} x_{l_q}^2(i,j) - T\mu^2(\mathbf{X}_{l_1, l_Q}) \quad (4)$$

The above expressions (3) and (4) apply to $\mu(\mathbf{Y}_{k_1, k_Q})$ and $\sigma(\mathbf{Y}_{k_1, k_Q})$, the global mean pixel value and zero-mean Euclidean norm of testing block $\mathbf{Y}_{k_1, k_Q}$ (without the gain terms). It should be noted that the above representation of large testing image areas by using concatenations of small training patches is common in unconstrained image matching (e.g., the sparse representation, patch NCC, and CNN based approaches). One major difference between these previous approaches and the proposed new approach is that the previous approaches tend to process each of the small patches independently. Because of this, they have lacked the ability to capture the larger, cross-patch dependence of image for image matching. But in the *block*-based NCC $R(\mathbf{Y}_{k_1, k_Q}, \mathbf{X}_{l_1, l_Q})$ defined above, there is no assumption of independence between the patches in a block. We will use this new metric to perform block matching,

i.e., to identify the matching patches in each block *jointly* for unconstrained face matching.

B. Block Matching - A Constrained Optimization Problem

Given a testing block $\mathbf{Y}_{k_1, k_Q} = (\mathbf{y}_{k_1}, \mathbf{y}_{k_2}, \dots, \mathbf{y}_{k_Q})$, for each enrolled subject, we maximize the block-based NCC $R(\mathbf{Y}_{k_1, k_Q}, \mathbf{X}_{l_1, l_Q})$ to estimate a best matching training block, denoted by $\hat{\mathbf{X}}_{l_1, l_Q} = (\hat{g}_{l_1} \hat{\mathbf{x}}_{l_1}, \hat{g}_{l_2} \hat{\mathbf{x}}_{l_2}, \dots, \hat{g}_{l_Q} \hat{\mathbf{x}}_{l_Q})$, from all possible concatenations of the training patches. Then we combine the similarity scores of all the matching training blocks of the person to perform person identification. However, not all of the possible concatenations of the training patches constitute realistic facial blocks. To constrain the estimate to valid facial blocks, we formulate the estimation of the matching training blocks for each person as a constrained maximization problem. We give a general expression:

$$\begin{aligned} \hat{\mathbf{X}}_{l_1, l_Q} &= \arg \max_{\mathbf{X}_{l_1, l_Q}} RH(\mathbf{Y}_{k_1, k_Q}, \mathbf{X}_{l_1, l_Q}) \\ &= \arg \max_{\mathbf{X}_{l_1, l_Q}} R(\mathbf{Y}_{k_1, k_Q}, \mathbf{X}_{l_1, l_Q}) H^\lambda(\mathbf{X}_{l_1, l_Q} | \mathbf{Y}_{k_1, k_Q}) \end{aligned} \quad (5)$$

where $H(\mathbf{X}_{l_1, l_Q} | \mathbf{Y}_{k_1, k_Q})$ is a likelihood function that the training block $\mathbf{X}_{l_1, l_Q}$ is a valid facial block given the testing block $\mathbf{Y}_{k_1, k_Q}$, $\lambda \geq 0$ denotes a weight for the likelihood function, and $RH(\mathbf{Y}_{k_1, k_Q}, \mathbf{X}_{l_1, l_Q})$ is an abbreviation for the validity-constrained block-based NCC. Eq. (5) means that the optimal matching training block estimate should have maximum block-based NCC, subject to the constraint of being most likely to be a valid facial block.

The formulation of a semantic constraint between the patches to constitute large valid facial regions remains a challenging problem. In our experiments, as an example, we use a simple semantic constraint on facial structure to regularize the formation of the potential matching training blocks for unconstrained testing facial images. Specifically, we assume that the location of the matching training patch l_q corresponding to testing patch k_q can be expressed as a function of k_q as follows:

$$l_q = k_q + d + \delta_q \quad (6)$$

where d represents a common displacement for all patches in the block, which may be caused by some major misalignment which affects all the patches equally, and δ_q represents a local displacement for a specific patch, which may be caused, for example, by random local pose or expression changes. We assume that δ_q is bounded: $0 \leq \|\delta_q\| \leq \Delta_q$, where Δ_q is a pre-chosen upper limit on the small local displacement δ_q . That is, some displacement is allowed for each matching training patch to account for unseen local variations in pose, expression and alignment, but an upper limit is used as a semantic constraint, requiring that the matching training patches maintain a similar geometric relationship to the corresponding testing patches, to

form a valid matching block. We use a simple expression for the constraining likelihood function:

$$H(\mathbf{X}_{l_1, l_Q} | \mathbf{Y}_{k_1, k_Q}) = \prod_{q=1}^Q h[l_q - (k_q + d)] \quad (7)$$

where $h[l_q - (k_q + d)]$ could take the form of an indicator function as defined below:

$$h[l_q - (k_q + d)] = \begin{cases} 1 & \text{if } \|l_q - (k_q + d)\| \leq \Delta_q, \\ 0 & \text{Otherwise.} \end{cases} \quad (8)$$

Eq. (8) is the function which is used in the majority of our experiments for its simplicity. Later, in an attempt to more realistically model valid facial regions, we study a refined version of (8). For this particular constraining function, a $\lambda = 0$ in (5) means the semantic constraint is not applied, and a $\lambda = 1$ means the semantic constraint is applied, in search for the matching training block estimate.

C. Multi-Scale Block Combination

The block-based matching approach, (5), aims to maintain invariance to local variations of small patches, while estimating a classification score for larger, more discriminative regions. The score for each testing block is given by $RH(\mathbf{Y}_{k_1, k_Q}, \hat{\mathbf{X}}_{l_1, l_Q})$. While we can seek the matches of large blocks up to the whole image, it is found in our experiments that, given limited training data, the combination of small, multi-scale block matching usually generate more efficient and more robust face recognition. Multi-scale combination has been used in other patch-based matching approaches to combine the complementary information across different patch sizes [35]–[42]. We extend this method to our block-based matching. Specifically, at each testing patch location k_1 , we identify a number of matching training blocks for the testing blocks $\mathbf{Y}_{k_1, k_Q}$ developed at this location with variable sizes Q , using (5), and then calculate the overall matching score $S(k_1)$ between these blocks at location k_1 using the following expression

$$S(k_1) = \sum_Q RH(\mathbf{Y}_{k_1, k_Q}, \hat{\mathbf{X}}_{l_1, l_Q}) w_Q \quad (9)$$

where w_Q denotes the weight for block size Q , reflecting its importance in the combination. The overall score for a specific person is formed by summing the $S(k_1)$ calculated using the person's training data over all the testing block locations k_1 . In our experiments, we have found that this multi-scale approach produces more correct classifications than any of the constituent block-sizes do, indicating its ability to capture complementary information from the variable-sized blocks.

III. AN ITERATIVE SOLUTION

A. Iterative Procedure

We propose a computationally efficient iterative algorithm to solve the constrained maximization problem (5). For each given testing block $\mathbf{Y}_{k_1, k_Q} = (\mathbf{y}_{k_1}, \mathbf{y}_{k_2}, \dots, \mathbf{y}_{k_Q})$, it seeks a block-wide joint estimation of the matching training patches to form the optimal matching block $\hat{\mathbf{X}}_{l_1, l_Q} =$

$(\hat{g}_{l_1}\hat{\mathbf{x}}_{l_1}, \hat{g}_{l_2}\hat{\mathbf{x}}_{l_2}, \dots, \hat{g}_{l_Q}\hat{\mathbf{x}}_{l_Q})$. We start with an initial estimate $\hat{\mathbf{X}}_{l_1, l_Q}$ by separately estimating each training patch $\hat{\mathbf{x}}_{l_q}$ based on maximizing the patch-based NCC $R(\mathbf{y}_{k_q}, \mathbf{x}_{l_q})$ defined in (1) assuming a unit gain, $\hat{g}_{l_q}$. Then we update this initial estimate by alternately re-estimating each training patch, with gain, to maximize the appropriate block-based validity-constrained NCC $RH(\mathbf{Y}_{k_1, k_Q}, \hat{\mathbf{X}}_{l_1, l_Q})$; in re-estimating a specific training patch, the other training patches are fixed to their latest estimates. This alternate re-estimation process iterates until convergence is achieved. For example, consider re-estimating the training patch $\hat{g}_{l_q}\hat{\mathbf{x}}_{l_q}$ in the order from $q = 1$ to Q . In the n th iteration, to obtain a new estimate of the optimal q th training patch, denoted by $\hat{g}_{l_q}^n\hat{\mathbf{x}}_{l_q}^n$, we maximize the block-based validity-constrained NCC with respect to $g_{l_q}\mathbf{x}_{l_q}$, with the succeeding training patches $g_{l_m}\mathbf{x}_{l_m}$ ($m > q$) taken from the $(n - 1)$ th iteration, and the preceding training patches $g_{l_m}\mathbf{x}_{l_m}$ ($m < q$) taken from the n th iteration. Therefore, in the n th iteration while re-estimating the q th training patch, the optimal training block estimate can be expressed as

$$\begin{aligned} & \hat{\mathbf{X}}_{l_1, l_Q}^n(g_{l_q}\mathbf{x}_{l_q}) \\ &= (\hat{g}_{l_1}^n\hat{\mathbf{x}}_{l_1}^n, \dots, \hat{g}_{l_{q-1}}^n\hat{\mathbf{x}}_{l_{q-1}}^n, g_{l_q}\mathbf{x}_{l_q}, \hat{g}_{l_{q+1}}^{n-1}\hat{\mathbf{x}}_{l_{q+1}}^{n-1}, \dots, \hat{g}_{l_Q}^{n-1}\hat{\mathbf{x}}_{l_Q}^{n-1}) \end{aligned} \quad (10)$$

which is only a function of $g_{l_q}\mathbf{x}_{l_q}$, with the rest of the training patches fixed to their latest optimal estimates from the appropriate iterations. Thus, we can obtain a new estimate of $\hat{g}_{l_q}\hat{\mathbf{x}}_{l_q}$, and a corresponding new block estimate, by maximizing $RH[\mathbf{Y}_{k_1, k_Q}, \hat{\mathbf{X}}_{l_1, l_Q}^n(g_{l_q}\mathbf{x}_{l_q})]$ with respect to $g_{l_q}\mathbf{x}_{l_q}$, i.e.,

$$\begin{aligned} & \hat{\mathbf{X}}_{l_1, l_Q}^n(\hat{g}_{l_q}^n\hat{\mathbf{x}}_{l_q}^n) \\ &= \arg \max_{g_{l_q}\mathbf{x}_{l_q}} RH[\mathbf{Y}_{k_1, k_Q}, \hat{\mathbf{X}}_{l_1, l_Q}^n(g_{l_q}\mathbf{x}_{l_q})] \\ &= \arg \max_{g_{l_q}\mathbf{x}_{l_q}} R[\mathbf{Y}_{k_1, k_Q}, \hat{\mathbf{X}}_{l_1, l_Q}^n(g_{l_q}\mathbf{x}_{l_q})] h^\lambda[l_q - (k_q + d)] \\ & \quad q = 1, 2, \dots, Q; n = 1, 2, \dots \end{aligned} \quad (11)$$

where $\hat{g}_{l_q}^0\hat{\mathbf{x}}_{l_q}^0$ corresponds to the initial estimate. Eq. (11) represents an iterative algorithm to solve the constrained maximization problem (5). It manages to estimate the optimal training patches one patch at a time, subject to the constraints of all the other patches in the block, and hence can be calculated efficiently. During the selection of a matching training patch, both the patch's identity, $\mathbf{x}_{l_q}$, and its corresponding gain factor, g_{l_q} , are optimized. The best gain factor $\hat{g}_{l_q}$ can be decided by finding, for each patch candidate $\mathbf{x}_{l_q}$, an optimal value within a pre-defined range to maximize $RH[\mathbf{Y}_{k_1, k_Q}, \hat{\mathbf{X}}_{l_1, l_Q}^n(\hat{g}_{l_q}\mathbf{x}_{l_q})]$ for each patch candidate $\mathbf{x}_{l_q}$. The optimal gain factors across the block form an estimate of the potentially arbitrary lighting condition in the testing block, based on a piece-wise constant lighting model. Let $\hat{\mathbf{X}}_{l_1, l_Q}^n$ denote the matching block estimate at the end of the n th iteration after all the training patches have been updated. In our experiments, we have seen that the above iterative algorithm converges in terms of generating block estimates that always increase the valid-constrained $RH(\mathbf{Y}_{k_1, k_Q}, \hat{\mathbf{X}}_{l_1, l_Q}^n)$ with each iteration. A more detailed step-by-step implementation of the algorithm is presented in

the next section along with a discussion on speeding up the algorithm.

B. The Step-by-Step Algorithm

The following summarizes our multi-scale block-based NCC algorithm for face recognition.

Given a testing image, we divide it into small over-lapping patches $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_K)$ as normally done in other patch-based approaches. Assume that we are given a set of training images for each person to be identified. These are also divided into over-lapping patches, each training image taking a form $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_L)$. Given the testing image, for each person we perform:

1) *Initialization*. We first find an initial matching image $\hat{\mathbf{X}}_{l_1, l_K}^0 = (\hat{g}_{l_1}^0\hat{\mathbf{x}}_{l_1}^0, \hat{g}_{l_2}^0\hat{\mathbf{x}}_{l_2}^0, \dots, \hat{g}_{l_K}^0\hat{\mathbf{x}}_{l_K}^0)$ from the training data for the whole testing image $\mathbf{Y}$, in which each matching training patch $\hat{\mathbf{x}}_{l_k}^0$ ($k = 1, 2, \dots, K$) is selected by maximizing the patch-based NCC $R(\mathbf{y}_k, \mathbf{x}_l)$ over all the subject's training patches $\mathbf{x}_l$ that satisfy $||l - (k + d)|| \leq \Delta_k$, assuming $\hat{g}_l^0 = 1$.

2) *Multi-scale, iterative block matching*. At each testing block location k_1 , for each block size Q , we form a testing block $\mathbf{Y}_{k_1, k_Q}$ from patch k_1 to k_Q . Then, we estimate the optimal matching training block by iteratively re-estimating each of the matching training patches by solving (11). Specifically, for each iteration n ($n \geq 1$), for each initial training patch of the matching block $\hat{g}_{l_q}^{n-1}\hat{\mathbf{x}}_{l_q}^{n-1}$ ($q = 1, 2, \dots, Q$) to be re-estimated, for each possible training patch candidate $\mathbf{x}_{l_q}$, we solve (11) to obtain a new optimal estimate $\hat{g}_{l_q}^n\hat{\mathbf{x}}_{l_q}^n$, by using the above initial estimates based on independent patch matching to start the iteration. We stop iterating when there is no change in the matching block estimate $\hat{\mathbf{X}}_{l_1, l_Q}^n$ between successive iterations.

3) *Multi-scale combination*. Having derived convergent, block-based scores $RH(\mathbf{Y}_{k_1, k_Q}, \hat{\mathbf{X}}_{l_1, l_Q}^n)$ for each block size Q at each testing block location k_1 for each person, we carry out a weighted sum of the scores at the variable sizes at each location for each person, using (9), to derive an average score for the blocks at the location for each person.

4) *Classification*. The above location-specific matching scores are summed over all the locations, for each person, to form the overall score of that person for identification.

It should be noted that the patched-based NCC approach [6] is a special case of our approach, which just corresponds to Step 1) of the above algorithm. Step 2) of the algorithm, iterative block matching, takes most of the computation time. In our experiments, we accelerate the computation in Step 2) by only considering in Step 2) the most-likely training patch candidates $\mathbf{x}_{l_q}$ for each testing patch $\mathbf{y}_{k_q}$. The most likely training patch candidates for each testing patch are selected in Step 1). The selected candidates satisfy $R(\mathbf{y}_{k_q}, \mathbf{x}_{l_q}) \geq R_{min}$, subject to the appropriate geometric semantic constraint, where R_{min} is a threshold used to prune unlikely matching training patches because of their extremely low correlation values. In our experiments, we choose $R_{min} = 0.2$.

Fig. 2. Examples of a testing image (left) and seven training images (right) from the three datasets used in the experiments: (a) GT, (b) LFWCrop, and (c) UFI.

IV. EXPERIMENTAL EVALUATION

A. Experimental Settings

We focus mainly on assessing the efficacy of the new approach for face identification in unconstrained conditions with limited training data. But first, we look for an additional validation of the performance of the new approach, by conducting experiments on the Georgia Tech (GT) dataset [49] which contains constrained face images. The GT dataset [49], contains 50 subjects with 15 images each. These images contain controlled variations such as pose, illumination and expression. In our experiments, we placed seven images per subject in training, as in [27].

Then, we conducted experiments on two more datasets, both containing unconstrained face images: the cropped Labelled Faces in the Wild (LFWCrop) [50], and Unconstrained Facial Images (UFI) [51]. On LFWCrop, we used a subset of 86 subjects with 11-20 images each, previously used in [27]. LFWCrop contains pre-cropped images which still exhibit real-life conditions, including (small) misalignment. As in [27], we placed seven images per subject in training and used the rest as testing images. Note that our new system uses much less training data than usually used in the CNN-based systems (e.g., [11]–[13]). Our experiments are designed to demonstrate the feasibility of matching large image areas with limited training data to improve face identification accuracy. We compare our new system with other systems built on similar training conditions to demonstrate this.

The UFI set contains 605 subjects with a single testing image each, and an average of 7.1 training images. We used the cropped version for our experiments. Example images from the GT, LFWCrop and UFI datasets are shown in Fig. 2.

For all three datasets, the images were cropped automatically with the Viola-Jones face detector and resized the images to 64×64 pixels. Then, we divided each image into patches of 8×8 pixels with an overlap of 4 pixels for our experiments. We assumed that the facial features occupy similar image regions, and therefore that the common patch displacement $d = 0$ in (6), and that the local patch displacement δ_q in (6) for each matching training patch is subject to a 30×30

pixels search window surrounding the corresponding testing patch, i.e., $\Delta_q = 15$ pixels in (8) for all patch indexes q . We have also tested other values for Δ_q and this appears to be the best for the databases used in our experiments, described above. The optimal gain factor for each matching training patch, $\hat{g}_{l,q}$, is estimated by golden section search within the range $[0.1, 3]$. We have tested some larger ranges, for example, $[0.1, 5]$, and found some small improvement. We note that when there are more than one training image available for one subject, we can have two different methods to form the matching image or matching score for each testing image: 1) the matching training patches forming the matching image all come from one training image, which produces the highest matching score among all the training images, and 2) the matching training patches forming the matching image can come from different training images, based on optimal (e.g., maximum NCC) selection. The original patch-based NCC scheme effectively used the first method (see [6] for more details). In our experiments, we used the second method, which was found to produce better results.

For convenience of presentation, we call our new single block-based NCC algorithm [i.e., (5) or (11)] BNCC, and call our new multi-scale block-based NCC algorithm [i.e., (9)] M-BNCC, as against the conventional patch-based algorithm NCC. M-BNCC weights matching scores of blocks of variable sizes to form the final score. Unless otherwise indicated, on the GT and LFWCrop datasets, we weighted matching scores of blocks from 1×1 patch up to 3×3 patches; on the UFI dataset, to account for the slightly increased training data for each subject, we went to a larger scale and weighted matching scores of blocks from 1×1 patch up to 4×4 patches. In our experiments, smaller block scales generally receive higher weights.

B. Experimental Results on the GT Dataset

We first present identification results on the GT dataset. We compare our M-BNCC approach to a number of recent approaches to unconstrained face recognition using the same training/testing subsets, particularly local patch-based and multi-scale techniques. Specifically, we compare M-BNCC to the original patch-based NCC [6] and to a recent extension to the SR (sparse representation) framework which synthesizes additional gallery images under different conditions to aid unconstrained classification [27]. To include a benchmark comparison of our approach with those which combine complementary information across scales, we consider the multi-scale approaches HMLBP [38] and MPCRC [41], which have been shown to outperform their single-scale counterparts. We also compare our approach to the classical PCA [1] and LDA [2] approaches, which may not represent the most state-of-the-art, but their strengths and limitations are well understood, making them useful in evaluating performance improvement. We obtained the results for these approaches by running the code provided by the authors, except for the patch-based NCC, which may be viewed as a special case of our M-BNCC approach (see Section III-B).

Table I presents identification accuracy by the six approaches discussed above, and our M-BNCC approach. On

TABLE I
IDENTIFICATION RATE (%) ON THE GT, COMPARING THE NEW M-BNCC
APPROACH WITH EXISTING APPROACHES.

PCA	LDA	SR	NCC	HMLBP	MPCRC	M-BNCC
43.50	75.00	75.25	77.00	74.75	66.50	91.25

TABLE II
IDENTIFICATION RATE (%) ON THE LFWCROP USED IN THIS PAPER, AND
ON OTHER DIFFERENT LFW SUBSETS CITED FROM PUBLICATIONS, FOR
THE NEW M-BNCC APPROACH COMPARED WITH EXISTING APPROACHES.

LFW subset	PCA	LDA	SR	NCC	HMLBP	MPCRC	M-BNCC
LFWCrop	7.13	14.57	28.37	48.37	36.28	15.04	65.27
Others			34.36	54.00	47.00	49.00	

the GT dataset the existing patch-based approaches struggle to overcome the classical LDA approach, as their benefits are more apparent in unconstrained conditions. In suitably constrained conditions such as the GT dataset, holistic approaches such as LDA can perform well. Our approach achieved substantial improvement over the best performing existing approach, NCC. This improvement is obtained in three ways. First, we exploit larger, more discriminative areas by jointly optimizing matching patches for blocks. Secondly, we combine the block-based scores at several block-sizes to make use of the complementary information at different scales. Finally, we select matching patches from any of the enrolled training images for a subject thereby exploiting the range of conditions in the available training data to find a robust match.

On the GT dataset higher recognition rates have been presented in the literature. For example, the authors in [52] reported improved identification accuracy over the classical LDA, 96.60%, by using Fisher scores for LDA-based classification, and by preprocessing the images to remove in-plane rotation. The authors in [53] reported identification accuracy of 98% by applying illumination normalization for edge-based face recognition and by using fourteen training images for training and one image for testing for each subject. The authors in [54] described an SR-based approach that uses an adaptive dictionary learned from the testing data to iteratively improve performance, reporting 96% identification accuracy after two iterations. Though our results are not directly comparable with these previous results, we have shown that the new M-BNCC approach, based on joint patch processing, is able to significantly improve over existing patch-based approaches and achieve good accuracy for constrained images without any preprocessing operations. In the following sections we assess the efficacy of M-BNCC for face identification in unconstrained conditions.

C. Experimental Results on the LFWCrop Dataset

On the LFWCrop dataset we compare our new M-BNCC approach to the same approaches as used in the GT experiments above. Table II shows the identification accuracy for each approach. As shown in the table, M-BNCC comfortably outperformed the existing techniques. The classical PCA and

LDA approaches failed to handle the large variations in unconstrained scenarios.

In previous work, the SR [27], patch-based NCC [6], HMLBP [38] and MPCRC [41] approaches have been tested on different subsets of the LFW database. These previously published results, cited directly from the respective papers, are included in Table II for information. As described in [27], the SR approach achieved identification accuracy of 34.36% on the same sets of images as we used in the LFWCrop dataset, but manually cropped and resized from the original LFW images, rather than taken directly from LFWCrop. Therefore the performance drop in Table II may be caused by the uncontrolled misalignment in the LFWCrop images. As reported in [41], MPCRC was able to achieve accuracy of 49% on a different LFW subset, with 158 subjects, each with five training images and two testing images, but used pre-aligned images rather than the LFWCrop images with uncontrolled misalignment. As reported in [6], on a subset of 50 LFW subjects, each with five training images and three testing images, and after applying the pose and illumination normalization processes from [55], the original patch-based NCC scheme and the HMLBP approach achieved identification rates of 54% and 47% respectively. However, on a larger dataset than this, without any normalization process, we are able to produce higher identification rates using the M-BNCC approach, as shown in Table II. In Section IV-E we will show that our block-based approach can also improve identification accuracy after the use of normalization processes by exploiting the greater discriminative information in larger image blocks.

Some other techniques have presented higher identification accuracy on various subsets of the LFW database. For example, the authors in [30] reported an identification rate of 83% when considering five subjects (but the rate was degraded to 58% when considering 50 subjects), based on a discriminative similarity measure learned with over 1000 negative example images. Exploiting weakly labeled data from the Internet as additional training data, considering the 50 LFW subjects with most training images, the authors in [31] reported identification accuracy of 81.8%. Using a subset of 12 LFW subjects with over 40 training images each, the authors in [24] reported accuracy over 90% by using local vector pattern (LVP) descriptors. However, the most recent breakthrough on the LFW database was achieved by the CNN-based systems (see, for example, [11]–[13]). The CNN-based OpenFace approach presented identification results on an LFW subset [56]. Considering 100 subjects, with 18 training images and two testing images each, and with the images aligned by pre-processing before classification, they reported identification accuracy of about 90%. The DeepID3 approach [12] used an external dataset of 200 thousand examples to train both the feature extraction CNN, and a joint Bayesian model for classification. This approach presented impressive identification results of 96% on the full LFW dataset. It is noted that the CNN-based systems make use of more training data than used in this paper, some obtained from various external sources including the Internet, to gain improved robustness for modeling small local image patches (receptive fields), and pre-align the images for training and recognition. The significance of this present

TABLE III

IDENTIFICATION RATE (%) ON THE UFI, COMPARING THE NEW M-BNCC APPROACH WITH BASELINE APPROACHES DESCRIBED IN [51], AND WITH PATCH-BASED NCC.

LBP	LDP	FS-LBP	POEM	NCC	M-BNCC
55.04	50.25	63.31	67.11	59.50	74.55

research is to demonstrate the feasibility of matching larger areas between unconstrained face images, with limited training data, to improve robustness, and achieving this without having to apply the usual pre-alignment or normalization process. The techniques described in this paper may well be combined with existing approaches for further improved performance. In Section IV-E we will show an example by presenting results on pose and illumination normalized data to evaluate the approach's ability to capture the greater discriminative information in a larger area following pre-processing to reduce image variations.

D. Experimental Results on the UFI Dataset

The more recent, and unconstrained, UFI dataset, is intended to be used for the benchmarking of face *identification* methods [51]. On this dataset, we compare our new approach with the baseline approaches presented in [51], as well as with the patch-based NCC [6]. The baseline UFI results include the traditional LBP approach [57], and the more recent local derivative patterns (LDPs), which extracts LBPs in high order derivative space [23], face-specific LBP (FS-LBP) [58], where facial feature points are identified to extract LBPs from, rather than extracting LBPs in a regular grid, and the multi-resolution POEM descriptor [39] which achieved the best baseline performance. These results are cited from [51] and presented in Table III, along with results from the patch-based NCC approach and with the new M-BNCC approach. To the best of our knowledge, M-BNCC has superior performance to the best published results on this unconstrained dataset based on the given training-testing data subdivision (with an average of 7.1 training images for each of the 605 subjects).

In the above, we have shown that our block-based M-BNCC approach to exploiting larger, multi-scale image regions for discrimination improves on conventional patch-based and multi-scale approaches. Improvement is achieved using only a small set of training images in both constrained and unconstrained conditions. For the latter, M-BNCC has shown robustness to local pose or lighting variations without pre-processing such as pose or illumination normalization. In the following, we further discuss the impact of each component of the M-BNCC algorithm on improving identification accuracy, by using examples from our experiments. We will also examine the ability of the M-BNCC algorithm to improve identification accuracy on normalized images with reduced variations.

E. Algorithmic Evaluation

In this section we analyze the effect of different characteristics of our algorithm on identification accuracy. We start with

TABLE IV

EFFECT OF LIGHTING COMPENSATION BY THE NEW APPROACH, SHOWING IDENTIFICATION ACCURACY (%) ON LFWCROP WITH VARIABLE-SIZED BLOCKS OF PATCHES, WITHOUT PATCH GAIN OPTIMIZATION (I.E., $g_{l_q} = 1$), AND WITH OPTIMIZED PATCH GAINS $\hat{g}_{l_q}$. THE SINGLE-PATCH BLOCK (I.E., THE NCC) RESULT IS INCLUDED FOR COMPARISON.

Block size	$g_{l_q} = 1$	Optimized $\hat{g}_{l_q}$
1×1	48.37	n/a
2×2	56.74	63.87
3×3	52.40	60.62
4×4	47.13	59.22

TABLE V

EFFECT OF JOINT PATCH ESTIMATION BY THE NEW APPROACH, SHOWING IDENTIFICATION ACCURACY (%) ON LFWCROP WITH VARIABLE-SIZED BLOCKS OF PATCHES, BASED ON INDEPENDENTLY AND JOINTLY ESTIMATED PATCHES, RESPECTIVELY. THE SINGLE-PATCH BLOCK (I.E., THE NCC) RESULT IS INCLUDED FOR COMPARISON.

Block size	Independent	Joint
1×1	48.37	n/a
2×2	24.96	63.87
3×3	25.27	60.62
4×4	25.89	59.22

our approach to improving robustness to illumination variation across the large blocks. Our patch-based gain factor estimates, $\hat{g}_{l_q}$, attenuate the effect of local illumination variation on the estimation of matching training blocks. We show this in Table IV, where we present the identification accuracy obtained by our single block-based BNCC algorithm at three block-sizes, with and without optimized local gain factors (the single-patch block case is included for comparison). We have found that optimizing the patch-based gain factors improves the quality of the matching training block estimates, and in turn improves the identification accuracy. As the block-size increases, the benefit of the gain optimization becomes more pronounced, because the larger regions are more likely to contain greater variations in lighting conditions. By assuming constant gain differences between small matching patches, and by jointly optimizing the gain factors of matching patches over entire blocks, we are able to effectively model non-uniform lighting changes between larger image blocks for accurate block matching.

Next, we show the effect of our *joint* patch estimation algorithm, for estimating the matching training blocks, on improving identification accuracy. We compare this algorithm with block matching based on independently estimated patches. In our iterative block matching algorithm described in Section III-B, these independently estimated blocks are used as the initial estimates for iterative block re-estimation. Table V shows the comparison results obtained on the LFWCROP dataset, considering three different block sizes. Independent patch matching may suffer poor accuracy due to low discriminative ability of the small patches, and due to local variations in unconstrained images. The proposed joint patch estimation improves robustness to local patch variations to find the matching blocks, improving the discriminative ability and hence improving the face identification accuracy.

Further, we evaluate the importance of the semantic con-

TABLE VI

EFFECT OF THE SEMANTIC CONSTRAINT ON THE SEARCH FOR THE MATCHING TRAINING PATCHES FOR THE NEW APPROACH, SHOWING IDENTIFICATION ACCURACY (%) ON LFWCROP WITH VARIABLE-SIZED BLOCKS OF PATCHES, SUBJECT TO THE STRICTEST CONSTRAINT ASSUMING PERFECT PATCH ALIGNMENT ($\Delta_q = 0$), WITHOUT ANY CONSTRAINT ($\lambda = 0$), AND SUBJECT TO THE PROPOSED SEMANTIC CONSTRAINT (8) WITH $\Delta_q = 15$ AND $\lambda = 1$. THE SINGLE-PATCH BLOCK (I.E., THE NCC) RESULT IS INCLUDED FOR COMPARISON.

Block size	$\Delta_q = 0$	$\lambda = 0$	Constraint (8)
1×1	24.50	42.79	48.37
2×2	30.85	53.18	63.87
3×3	28.37	50.39	60.62
4×4	27.13	47.75	59.22
Full image	22.17	52.56	50.54

straint, (8), on the location of selected matching training patches to form matching training blocks. We consider three different settings concerning the application of the constraint: 1) $\Delta_q = 0$ for all q , that is, only the training patches at the exact same location as the testing patches are searched for matching patches; this setting effectively assumes perfectly aligned training and testing images; 2) the exponential weight $\lambda = 0$ [see (5)], that is, the matching patches can come from anywhere in the training images or there is no constraint on their selection; 3) the constraint shown in (8) is applied with $\lambda = 1$ and $\Delta_q = 15$, which is the condition used to generate the above experimental results. Table VI shows the comparison results obtained on the LFWCROP dataset, considering different block sizes up to full image. We can see that our proposed search strategy subject to the semantic constraint (8) achieved the best accuracy when the matching blocks are small. When the semantic constraint is too strict (i.e., $\Delta_q = 0$ or assuming perfect patch alignment between training and testing images), we obtained the poorest accuracy for matching the actually unconstrained images without any alignment or normalization preprocessing. However, when the semantic constraint is too loose or removed (as demonstrated in Table VI with $\lambda = 0$), we found that the selected training patches may not form a valid facial region when concatenated into a block, as they may be drawn from distant locations, and hence different facial features, in the training images. This led to matching errors. Table VI also indicates that as the size of the matching block increases, the matching accuracy decreases. This can be caused by limited training data, an inaccurate semantic constraint, or both, for modeling very large facial variations. When the matching blocks become large, forming an accurate semantic constraint on the small constituent patches remains a challenging problem. This is indicated by the poorer accuracy for the new approach compared to unconstrained matching when using full-image blocks. In the following we discuss two of our attempts to address this problem.

First, we tried to refine the semantic constraint model. We have noted that (8), the constraint on the location of a potential matching patch, is applied to each patch in a block independently of the other patches in the block. In an attempt to more closely relate the matching patches in a block, we revised (8). Specifically, we assume that the location of each matching training patch $\mathbf{x}_{l_q}$, with the top-left corner

TABLE VII

EFFECT OF THE REVISED CONSTRAINT, AND POSE AND ILLUMINATION NORMALIZATION, FOR THE NEW APPROACH, SHOWING IDENTIFICATION ACCURACY (%) ON LFWCROP.

Block size	Constraint (8)	Constraint (12)	Normalization + Constraint (8)
2×2	63.87	64.03	69.46
3×3	60.62	60.16	72.40
4×4	59.22	57.21	72.87
Multi-scale	65.27	65.89	74.57

Fig. 3. An example image from the LFWCROP dataset: (a) before normalization, and (b) after the pose and illumination normalisation.

coordinates denoted by (i_{l_q}, j_{l_q}) , is subject to the constraint of not only the location of the testing patch [i.e., (8)], but also the locations of its immediate neighboring matching training patches, in terms of forming a valid facial block given the testing block. The additional constraint can be expressed as: $\max i_{l_q}^{\text{lower}} \leq i_{l_q} \leq \min i_{l_q}^{\text{upper}}$, and $\max j_{l_q}^{\text{left}} \leq j_{l_q} \leq \min j_{l_q}^{\text{right}}$, where $i_{l_q}^{\text{lower}}$, $i_{l_q}^{\text{upper}}$, $j_{l_q}^{\text{left}}$ and $j_{l_q}^{\text{right}}$ represent the appropriate coordinates of the matching training patches in the block, which are located in the immediate lower row, upper row, left column and right column of $\mathbf{x}_{l_q}$, respectively, if applicable. Thus, we have a revised constraint indicator function:

$$h[l_q - (k_q + d)] = \begin{cases} 1 & \text{if } ||l_q - (k_q + d)|| \leq \Delta_q, \\ & \max i_{l_q}^{\text{lower}} \leq i_{l_q} \leq \min i_{l_q}^{\text{upper}}, \\ & \max j_{l_q}^{\text{left}} \leq j_{l_q} \leq \min j_{l_q}^{\text{right}}. \\ 0 & \text{Otherwise.} \end{cases} \quad (12)$$

New experiments were conducted by using (12) in place of (8), for the LFWCROP dataset. The results for the new constraint are shown in the third column in Table VII, alongside the results based on the old constraint (8) for comparison. We see that the new constraint did not bring any significant improvement in terms of identification accuracy. One possible avenue for future work is to use a confidence score from a face detector as a measure of validity for a reconstructed face, or facial region.

Second, we addressed this problem by trying to reduce the variations across the images with pre-processing, thereby reducing the complexity of the models of their semantic constraint for matching. This also serves as an example of combining the new approach with existing approaches for further improved performance. Specifically, we applied both pose and illumination normalization to the training and testing images. We used the pose-corrected images from [59], which used a commercial 3D face modeler to render frontal faces. We then applied SQI illumination normalization [19]. An example normalized image is shown in Fig.3. For the normalized

images, we show that larger matching blocks can be more effectively estimated, producing superior identification results. Table VII shows, in the last column, the identification accuracy by the new approach on the pose and illumination normalized LFWCrop images, based on the constraint (8). We can see that considerably better results are achieved when using normalized data. We can also see that, on the normalized dataset, the identification accuracy increases when the block-size increases from 2×2 to 4×4 . This is not the case on the un-normalized LFWCrop images. The reduced variations in the normalized images enable more accurate estimates of larger matching blocks, with the relatively simpler semantic constraint (8), and without increased training data.

In Table VII we can also see that the multi-scale block combination improves identification accuracy over each of the constituent block-sizes. In our experiments the weights for combining the NCC scores at different block-sizes were manually tuned for each dataset. In [41] weights are learned for multiple scales from the training set, by l_1 -regularized optimization, deriving weights which retain as many correct classifications as possible across the scales. Our future work will include identifying the weights optimally.

Based on our results on the LFWCrop dataset shown above, with unconstrained images with very limited training data, and without pre-processing, we see that we may only be able to significantly improve the identification accuracy, over the conventional single-patch based matching, by using blocks up to 2×2 patches. Further increasing the size of the blocks added little improvement, due to the increasing image variations, and due to the lack of effective semantic constraints to perform accurate matching. However, Table VII indicates that, when the image variations are reduced with appropriate pre-processing, further, reliable performance gain can be obtained by using larger-sized blocks (in this example, up to 4×4 patches). We also assume that increasing the amount of training data could help cover a wider range of variations, as for the CNNs, and hence find larger matching blocks. In theory, our approach should be operational with a single unconstrained training image per subject (this is not feasible for the CNNs). But we found poor accuracy on the LFWCrop dataset with a single training image (24.03%) as compared to seven training images (59.22%), assuming a block size of 4×4 patches

Finally, we show the convergence of the iterative algorithm used to implement block matching, described in Section III-B. Fig. 4 summarizes the average number of iterations used in our experiments to reach convergent matching block estimates, and the initial and convergent validity-constrained block-based NCC scores (i.e., RH defined in (11)). The results are obtained on the LFWCrop dataset, for estimating four different sized matching blocks. All the estimates indicate an increasing RH value with each iteration. In our experiments, we have not seen a single case in which the iteration decreased the appropriate RH values.

We conduct a semantically constrained search of potential matching patches at various block-sizes. The computational complexity of our approach is dependent on the number of training images and the size of the search window. The search space for a testing patch could be reduced by clustering the

Fig. 4. Convergence of the iterative block estimation algorithm in the experiments, showing the average number of iterations used to reach convergent estimates, and the increase in the validity-constrained block-based NCC scores (RH) after convergence, obtained on the LFWCrop dataset for four different block sizes.

training image patches, subject to both structural and semantic similarity, into subsets of similar patches. The search for a matching training patch, for a testing patch, could then be carried out on a subset of patches which are similar to the testing patch.

V. CONCLUSIONS

Most current methods for unconstrained face recognition are patch based. The method described in this paper, based on blocks of patches, is a complement to existing methods. This research aims to improve the robustness to uncontrolled pose, expression and lighting variations by enlarging the matching areas without assuming additional training data. We extended the NCC metric, previously used in patch-based matching, to our block-based matching to demonstrate this feasibility. This paper described a realization of the block-based image matching for practical use. The core part of this realization is the solution of a constrained maximization problem, for jointly estimating the matching patches forming each matching block from the available training data with computational efficiency. We showed that the new algorithm is not only capable of dealing with pose or expression differences, but also capable of compensating for uneven lighting differences across the blocks. Finally, we presented a multi-scale extension, to further exploit the block-based matching at variable block sizes to improve the performance at larger scales. On three datasets we have shown the benefit of our contributions, and improved identification accuracy significantly over existing patch-based methods for unconstrained face identification with limited training data. We suggest that our approach to improving robustness for patch-based image processing can also be exploited in future to improve the quality of features and classifications produced by CNNs. More research is required to identify a suitable way to combine this idea with an appropriate CNN architecture. In future, we would also like to develop an algorithm for learning weights to combine scales more effectively, rather than manually tuning.

APPENDIX A: NCC AND ROBUSTNESS

Consider the NCC $R(\mathbf{y}_k, \mathbf{x}_l)$ between a testing patch $\mathbf{y}_k$ and a training patch $\mathbf{x}_l$. Assume that the testing patch $\mathbf{y}_k$ is noisy and can be expressed as $\mathbf{y}_k = \mathbf{x}'_k + \mathbf{n}_k$, where $\mathbf{x}'_k$ represents the underlying clean patch to be matched, and $\mathbf{n}_k$ represents the noisy patch. Then we can decompose $R(\mathbf{y}_k, \mathbf{x}_l)$, defined in (1), into two terms

$$\begin{aligned} R(\mathbf{y}_k, \mathbf{x}_l) &= R(\mathbf{x}'_k + \mathbf{n}_k, \mathbf{x}_l) \\ &= \frac{\sigma(\mathbf{x}'_k)}{\sigma(\mathbf{y}_k)} R(\mathbf{x}'_k, \mathbf{x}_l) + \frac{\sigma(\mathbf{n}_k)}{\sigma(\mathbf{y}_k)} R(\mathbf{n}_k, \mathbf{x}_l) \end{aligned} \quad (13)$$

where $\mu(\mathbf{x}'_k)$, $\mu(\mathbf{n}_k)$ (implied) and $\sigma(\mathbf{x}'_k)$, $\sigma(\mathbf{n}_k)$ represent the mean pixel value and zero-mean Euclidean norm of patch $\mathbf{x}'_k$ and $\mathbf{n}_k$, respectively. In (13), the first term contains the NCC we seek, for matching the underlying image patch $\mathbf{x}'_k$ against the training image patch $\mathbf{x}_l$. This NCC is weighted by $\sigma(\mathbf{x}'_k)/\sigma(\mathbf{y}_k)$ which is a constant for all the training image patches, subject only to the signal-to-noise ratio (SNR) in the observation. The second term contains the NCC between the noisy patch and the training image patch, i.e., the NCC introduced by noise, weighted by $\sigma(\mathbf{n}_k)/\sigma(\mathbf{y}_k)$ which is again independent of the training image patch, subject only to the SNR in the observation. For large patch sizes and noise independent of the training images, it follows that the second term in (13) tends to zero:

$$\begin{aligned} R(\mathbf{n}_k, \mathbf{x}_l) &= \frac{\sum_{i,j} [n_k(i,j) - \mu(\mathbf{n}_k)][x_l(i,j) - \mu(\mathbf{x}_l)]}{\sigma(\mathbf{n}_k)\sigma(\mathbf{x}_l)} \\ &\propto E\{[n_k(i,j) - \mu(\mathbf{n}_k)][x_l(i,j) - \mu(\mathbf{x}_l)]\} \quad (14) \\ &= E[n_k(i,j) - \mu(\mathbf{n}_k)]E[x_l(i,j) - \mu(\mathbf{x}_l)] = 0 \quad (15) \end{aligned}$$

where $n_k(i,j)$ denotes the noise value at pixel location (i,j) inside patch $\mathbf{y}_k$. Eq. (14) is based on the assumption that as the patch size (i.e., the number of pixels used in the calculation) becomes large, the sample average converges to the ensemble average (here we assume ergodicity for both the image and noise processes [60]), and (15) [from (14)] is based on the assumption that the training images and noise are statistically independent. Thus, with (13)–(15), for large patches and independent noise we may have

$$R(\mathbf{y}_k, \mathbf{x}_l) \propto R(\mathbf{x}'_k, \mathbf{x}_l) \quad (16)$$

That is, the maximum NCC (i.e., the matching accuracy) could become independent of the noise but depends only on the two image patches being compared, one being the underlying clean patch to be recognized and the other a potential matching estimate from the training data.

REFERENCES

- [1] M. Turk and A. Pentland, "Eigenfaces for recognition," *Cognitive Neuroscience*, vol. 3, no.1, pp. 71–86, 1991.
- [2] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no.7, pp. 711–720, 1997.
- [3] K. Obha and K. Ikeuchi, "Detectability, uniqueness, and reliability of eigen windows for stable verification of partially occluded objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 9, pp. 1043 – 1047, 1997.
- [4] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.
- [5] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, and Y. Ma, "Towards a practical face recognition system: Robust alignment and illumination by sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 2, pp. 372–386, 2012.
- [6] M. D. Marsico, M. Nappi, D. Riccio, and H. Wechsler, "Robust face recognition for uncontrolled pose and illumination changes," *IEEE Transactions on Systems, Man and Cybernetics: Systems*, vol. 43, no. 1, pp. 149–163, 2013.
- [7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1106–1114.
- [8] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1–9.
- [9] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *International Conference on Learning Representations (ICLR)*, 2015, pp. 1701–1708.
- [10] A. Coates and A. Y. Ng, "Selecting receptive fields in deep networks," in *International Conference on Neural Information Processing Systems*, 2011, pp. 2528–2536.
- [11] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 1701–1708.
- [12] Y. Sun, D. Liang, X. Wang, and X. Tang, "Deepid3: Face recognition with very deep neural networks," *CoRR*, 2015.
- [13] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Web-scale training for face identification," in *Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 2746–2754.
- [14] H. Hu, Y. Yang, D. Yi, J. Kittler, W. Christmas, S. Z. Li, and T. Hospedales, "When face recognition meets with deep learning: An evaluation of convolutional neural networks for face recognition," in *IEEE International Conference on Computer Vision Workshop (ICCVW)*, 2015.
- [15] E. H. Land and J. J. McCann, "Lightness and retinex theory," *Journal of Optical Society of America*, vol. 61, no. 1, pp. 1–11, 1971.
- [16] A. A. Shafie, F. Hafiz, and Y. M. Mustafah, "Face recognition using illumination-invariant local patches," *International Conference on Intelligent and Advanced Systems*, pp. 1–6, 2014.
- [17] N. McLaughlin, J. Ming, and D. Crookes, "Illumination invariant facial recognition using a piecewise-constant lighting model," *2012 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1537–1540, 2012.
- [18] —, "Largest matching areas for illumination and occlusion robust face recognition," *IEEE Transactions on Cybernetics*, vol. 47, no. 3, pp. 796–808, 2017.
- [19] H. Wang, S. Z. Li, Y. Wang, and J. Zhang, "Self quotient image for face recognition," *International Conference on Image processing, 2004.*, vol. pp. 1397–1400, 2004.
- [20] T. Zhang, Y. Y. Tang, B. Fang, Z. Shang, and X. Liu, "Face recognition under varying illumination using gradientfaces," *IEEE Transactions on Image Processing*, vol. 18, no. 11, pp. 2699–2606, 2009.
- [21] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.
- [22] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006.
- [23] B. Zhang, Y. Gao, S. Zhao, and J. Liu, "Local derivative pattern versus local binary pattern: Face recognition with high-order local pattern descriptor," *IEEE Transactions on Image Processing*, vol. 19, no. 2, pp. 533–544, 2010.
- [24] K.-C. Fan and T.-H. Hung, "A novel local pattern descriptor - local vector pattern in high-order derivative space for face recognition," *IEEE Transactions on Image Processing*, vol. 23, no. 7, pp. 2877–2891, 2014.
- [25] C. Ding, J. Choi, D. Tao, and L. S. Davis, "Multi-directional multi-level dual-cross patterns for robust face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 3, pp. 518–531, 2016.
- [26] A. S. Georgiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Analysis and Machine intelligence*, vol. 23, no. 6, pp. 643–660, 2001.

- [27] Y. Xu, X. Fang, X. Li, J. Yang, J. You, H. Liu, and S. Teng, "Data uncertainty in face recognition," *IEEE Transactions on Cybernetics*, vol. 44, no. 10, pp. 1950–1961, 2014.
- [28] Z.-Q. Zhao, Y.-M. Cheung, H. Hu, and X. Wu, "Expanding dictionary for robust face recognition: pixel is not necessary while sparsity is," *IET Computer Vision*, vol. 9, no. 5, pp. 648–654, 2015.
- [29] Y. Chen, T. Do, and T. Tran, "Robust face recognition using locally adaptive sparse representation," *IEEE International Conference on Image Processing*, pp. 1657–1660, 2010.
- [30] L. Wolf, T. Hassner, and Y. Taigman, "Effective unconstrained face recognition by combining multiple descriptors and learned background statistics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 10, pp. 1978–1990, 2010.
- [31] D. Rim, M. K. Hasan, F. Puech, and C. J. Pal, "Learning from weakly labelled faces and video in the wild," *Pattern Recognition*, vol. 48, no. 3, pp. 759–771, 2015.
- [32] D. Yi, Z. Lei, and S. Li, "Towards pose robust face recognition," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3539–3545, 2013.
- [33] S. A. Asthana, T. K. Marks, M. J. Jones, K. H. Tieu, and M. Rohith, "Fully automatic pose-invariant face recognition via 3D pose normalization," *Proceedings of the International Conference on Computer Vision*, pp. 937–944, 2011.
- [34] S. Romdhani, V. Blanz, and T. Vetter, "Face identification by fitting a 3D morphable model using linear shape and texture error functions," *European Conference on Computer Vision*, pp. 3–19, 2002.
- [35] C.-H. Chan, J. Kittler, and K. Merron, "Multi-scale local binary pattern histograms for face recognition," *Advances in Biometrics: International Conference*, pp. 809–818, 2007.
- [36] O. Nikisins, "Weighted multi-scale local binary pattern histograms for face recognition," *International Conference on Applied Mathematics and Computational Methods*, pp. 76–81, 2013.
- [37] Q.-C. Tao, Z.-M. Liu, G. Bebis, and M. Hussain, "Face recognition using a novel image representation scheme and multi-scale local features," *International Journal of Biometrics*, vol. 7, no. 3, pp. 191–212, 2015.
- [38] Z. Guo, L. Zhang, D. Zhang, and X. Mou, "Hierarchical multiscale LBP for face and palmprint recognition," *International Conference on Image Processing*, pp. 4521–4524, 2010.
- [39] N.-S. Vu, H. M. Dee, and A. Caplier, "Face recognition using the POEM descriptor," *Pattern Recognition*, vol. 45, no. 7, pp. 2478–2488, 2012.
- [40] M. X. Nguyen, Q. M. Le, V. Pham, T. Tran, and B. H. Le, "Multi-scale sparse representation for robust face recognition," *Third International Conference on Knowledge and Systems Engineering (KSE)*, pp. 195–199, 2011.
- [41] P. Zhu, L. Zhang, Q. Hu, and S. C. K. Shiu, "Multi-scale patch based collaborative representation for face recognition with margin distribution optimization," *Computer Vision - ECCV 2012*, vol. 7572, pp. 822–835, 2012.
- [42] G. Gao, J. Yang, X. Jing, P. Huang, J. Hua, and D. Yue, "Robust face recognition via multi-scale patch-based matrix regression," *PLoS One*, vol. 11, no. 8, 2016.
- [43] Y. Lu, Z. Lai, Y. Xu, X. Li, D. Zhang, and C. Yuan, "Low-rank preserving projections," in *IEEE Transactions on Cybernetics*, vol. 46, no. 1, 2016, pp. 1900–1913.
- [44] —, "Nonnegative discriminant matrix factorization," in *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 7, 2017, pp. 1392–1405.
- [45] Y. Lu, C. Yuan, Z. Lai, X. Li, W. K. Wong, and D. Zhang, "Nuclear norm-based 2dlpp for image classification," in *IEEE Transactions on Multimedia*, vol. 19, no. 11, 2017, pp. 2391–2403.
- [46] Y. Lu, Z. Lai, Y. Xu, J. You, X. Li, and C. Yuan, "Projective robust nonnegative factorization," in *Information Sciences*, vol. 364–365, 2016, pp. 16–32.
- [47] Y. Lu, Z. Lai, Z. Fan, J. Cui, and Q. Zhu, "Manifold discriminant regression learning for image classification," in *Neurocomputing*, vol. 166, 2015, pp. 475–486.
- [48] J. Gaston, J. Ming, and D. Crookes, "Unconstrained face identification with multi-scale block-based correlation," in *Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2017, pp. 1477–1481.
- [49] A. V. Nefian, "Georgia tech face dataset," http://anefian.com/research/face_reco.htm, [Online; Accessed 3-May-2016].
- [50] C. Anderson, "LFWcrop face dataset," <http://conradsanderson.id.au/lfwcrop/>, [Online; accessed 9-February-2016].
- [51] L. Lenc and P. Král, "Unconstrained Facial Images: Database for face recognition under real-world conditions," in *14th Mexican International Conference on Artificial Intelligence*, 2015, pp. 25–31.
- [52] L. Chen, H. Man, and A. V. Nefian, "Face recognition based on multi-class mapping of fisher scores," *Pattern Recognition*, vol. 38, no. 6, pp. 799–811, 2005.
- [53] C. C. Chude-Olisah, G. Sulong, U. A. K. Chude-Okonkwo, and S. Z. M. Hashim, "Illumination normalization for edge-based face recognition using the fusion of rgb normalization and gamma correction," *2013 IEEE International Conference on Signal and Image Processing Applications (ICSIPA)*, pp. 412–416, 2013.
- [54] K. Estabridis, "Face recognition and learning via adaptive dictionaries," *2012 IEEE Conference on Technologies for Homeland Security*, pp. 280–285, 2012.
- [55] B. C. Becker and E. G. Ortiz, "Evaluation of face recognition techniques for application to facebook," in *8th IEEE International Conference on Automatic Face & Gesture Recognition*, 2009.
- [56] B. Amos, B. Ludwiczuk, and M. Satyanarayanan, "Openface: A general-purpose face recognition library with mobile applications," CMU-CS-16-118, CMU School of Computer Science, Tech. Rep., 2016.
- [57] T. Ojala, M. Pietikainen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.
- [58] L. Lenc and P. Král, "Automatically detected feature positions for lbp based face recognition," *Proceedings of the International Conference on Artificial Intelligence Applications and Innovations*, pp. 246–255, 2014.
- [59] L. Best-Rowden, H. Han, C. Otto, B. F. Klare, and A. K. Jain, "Unconstrained face recognition: Identifying a person of interest from a media collection," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 12, pp. 2144–2157, 2014.
- [60] B. Gnedenko, *The Theory of Probability* (translated from the Russian by B.D. Seckler). Chelsea, N.Y., 1962.

Jack Gaston received the B.Sc. degree in Computer Science from Queen's University Belfast in 2013. In 2017 he received the Ph.D. degree, also from Queen's University, at The Institute Of Electronics, Communications And Information Technology. Since then he has been developing software for scientific imaging devices, including image restoration algorithms for applications such as denoising and super-resolution, at Andor Technology, Belfast. His research interests include image restoration and recognition, particularly biometric identification.

Ji Ming received the B.Sc. degree from Sichuan University, Chengdu, China, in 1982, the M.Phil. degree from Changsha Institute of Technology, Changsha, China, in 1985, and the Ph.D. degree from Beijing Institute of Technology, Beijing, China, in 1988, all in electronic engineering.

He was Associate Professor with the Department of Electronic Engineering, Changsha Institute of Technology, from 1990 to 1993. Since 1993, he has been with the Queen's University Belfast, Belfast, U.K., where he is currently a Professor in the School of Electronics, Electrical Engineering and Computer Science. From 2005 to 2006, he was a Visiting Scientist at the Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge. His research interests include speech and language processing, image processing, signal processing and pattern recognition.

Danny Crookes was appointed to the Chair of Computer Engineering in 1993 at Queen's University Belfast, Belfast, U.K., and was Head of Computer Science from 1993-2002. He is currently Director of Research for Speech, Image and Vision Systems at the Institute of Electronics, Communications and Information Technology, Queen's University Belfast. His current research interests include the use of novel architectures (especially GPUs) for high performance speech and image processing. Professor

Crookes is currently involved in projects in automatic shoeprint recognition, speech separation and enhancement, and processing of 4D confocal microscopy imagery. Professor Crookes has over 200 scientific papers in journals and international conferences.