SPECIAL ISSUE

FPGA-based IP cores implementation for face recognition using dynamic partial reconfiguration

**Afandi Ahmad · Abbes Amira · Paul Nicholl ·
Benjamin Krill**

Abstract This paper presents a combination of novel feature vectors construction approach for face recognition using discrete wavelet transform (DWT) and field programmable gate array (FPGA)-based intellectual property (IP) core implementation of transform block in face recognition systems. Initially, four experiments have been conducted including the DWT feature selection and filter choice, features optimisation by coefficient selections and feature threshold. To examine the most suitable method of feature extraction, different wavelet quadrant and scales have been evaluated, and it is followed with an evaluation of different wavelet filter choices and their impact on recognition accuracy. In this study, an approach for face recognition based on coefficient selection for DWT is presented, and the significant of DWT coefficient threshold selection is also analysed. For the hardware implementation, two architectures for two-dimensional (2-D) Haar wavelet transform (HWT) IP core with transpose-based computation and dynamic partial reconfiguration (DPR) have been synthesised using VHDL and implemented on Xilinx Virtex-5 FPGAs. Experimental results and comparisons between different configurations using partial and non-partial reconfiguration processes and a detailed performance analysis of the area, power consumption and maximum frequency are also discussed in this paper.

A. Ahmad
Department of Computer Engineering,
Faculty of Electrical and Electronic Engineering,
Universiti Tun Hussein Onn Malaysia (UTHM),
Johor, Malaysia

A. Amira (✉) · B. Krill
Nanotechnology and Integrated Bio-Engineering Centre
(NIBEC), Faculty of Computing and Engineering,
University of Ulster (Jordanstown Campus),
Ulster, Northern Ireland
e-mail: a.amira@ulster.ac.uk; abbes.amira@qu.edu.qa

A. Amira
Department of Electrical Engineering, College of Engineering,
Qatar University, Doha, Qatar

P. Nicholl
School of Electronic, Electrical Engineering and Computer
Science, The Queens University, Belfast, Northern Ireland

1 Introduction

The use of biometric systems is growing rapidly. Face recognition technology has the potential to be a convenient, robust biometric, used for many applications [1]. Currently, recognition rates are adversely affected by variation in illumination, pose, gesture and other factors [2]. Much research is currently being undertaken in face recognition and has a large number of potential applications, such as port of entry logging, building access control, criminal identification and attendance logging [3]. On top of that, a number of commercial face recognition systems have been developed, including products from Cognitec [4], L-1 Identity Solutions [5], Geometrix [6], Technest [7] and Animetrics [8].

In this study, both software simulations and an implementation of intellectual property (IP) core for transform block in the face recognition systems are discussed. Initially, four experiments have been conducted including the discrete wavelet transform (DWT) feature selection and filter choice, features optimisation by coefficient selections and feature threshold. To examine the most suitable method of feature extraction, different wavelet quadrant and scales have been evaluated, and it is followed with an

evaluation of different wavelet filter choices and their impact on recognition accuracy. Results obtained from the software simulations are then being used to justify the hardware implementation. Throughout this study, AT&T database has been selected since it is a conveniently sized dataset and suitable for testing algorithms in development as well as the proposed IP core implementation.

In addition to the software simulation, another part of this study deals with the hardware implementation of transform block in the face recognition system using field programmable gate array (FPGA). With the ultimate aim to accelerate the process of transforming input images into the wavelet coefficients, FPGAs platform is selected for several reasons. First and foremost, it allows for truly parallel computations to take place in a circuit. Many modern general purpose processors (GPPs) and operating systems can emulate parallelism by switching tasks very rapidly. Having operations occur in a parallel fashion results in a much faster overall processing time. This is the case even though the clock speed of the FPGA is lower than the GPPs.

With the availability of advanced embedded resources on recent devices such as soft cores, dedicated logic and block multipliers, it is not surprising that there has been a considerable amount of research into the use of FPGAs to increase the performance of a wide range of computationally intensive applications [9–11]. One such application that could greatly benefit from the advantages offered by FPGAs is face recognition. The regular nature of the complex computations performed repeatedly within face recognition operations are well suited to a hardware-based implementation using FPGAs.

In this study, Xilinx FPGA devices with dynamic partial reconfiguration (DPR) technique have been selected to prototype the developed architectures. DPR is a technique that offers changing the configuration of a part of a circuit whilst the rest of it executes its task [12].

For the hardware implementation, this research aims at developing a novel implementation of two-dimensional (2-D) Haar wavelet transform (HWT) IP core for the transform block in face recognition systems. DPR technique is fully utilised as it capable to divide the designs into several sub-designs that fit into the available hardware resources and can be uploaded into the reconfigurable hardware when needed. Xilinx Virtex-5 is used to prototype the proposed architectures and an examination of the transform size influence on the area, power consumption and maximum frequency is also carried out. To further investigate the development of a complete system-on-a-chip (SoC) solution, the principal component analysis-discrete wavelet transform (PCA-DWT)-based face recognition system has been also deployed on the RC10 FPGA prototyping board equipped with the low-power Spartan 3l1500 Xilinx FPGA.

The rest of the paper is organised as follows. An overview of the algorithms and methodology are presented in Sect. 2.

Section 3 summarises the research analysis that have been obtained through four experiments: DWT feature selection and filter choice, optimising features by coefficient selection and feature threshold. Hardware implementation and results analysis for the proposed architectures are described in Sect. 4. Finally, concluding remarks are given in Sect. 5.

2 Algorithms and methodology

An overview of the algorithms and design methodology for the software simulation as well as the hardware implementation are presented in the following sections.

2.1 Discrete wavelet transform (DWT)

DWT can be implemented as a set of filter banks, comprising a high-pass and low-pass filter (also known as the scaling filter). At each stage, the output from the low-pass filter can be decomposed further, with the process continuing recursively in this manner. DWT can be mathematically expressed as shown in (1).

$$\text{DWT}_{x(n)} = \begin{cases} d_{j,k} = \sum x(n) h_j^*(n - 2^j k) \\ a_{j,k} = \sum x(n) g_j^*(n - 2^j k) \end{cases} \tag{1}$$

The coefficients $d_{j,k}$ refer to the detail components in signal $x(n)$ and correspond to the wavelet function, whereas $a_{j,k}$ refer to the approximation components in the signal. The functions $h(n)$ and $g(n)$ in the equation represent the coefficients of the high-pass and low-pass filters respectively, whilst parameters j and k refer to wavelet scale and translation factors. Figure 1 illustrates a three-level DWT decomposition.

The one-dimensional (1-D) DWT can be readily extended to two dimensions, allowing it to be employed for analysing images. The 2-D DWT exists in both standard and non-standard forms. In 2-D standard wavelet decomposition (SWD) [13], the image rows are fully decomposed, with the output then being fully decomposed column-wise. On the contrary, the rows are decomposed by one decomposition level followed by one decomposition level of the columns with the non-standard wavelet decomposition (NSWD) [13].

The decomposition continues by decomposing the low-resolution output from each step, row-wise followed by column-wise, until the image is fully decomposed. Figure 2 illustrates the effect of applying the non-standard wavelet transform to an image from the AT&T database of faces [14].

2.2 Principal component analysis (PCA)

PCA is a dimensional-reduction technique and it does not model relationships between neighbouring pixels in an

Fig. 1 A three-level wavelet decomposition system
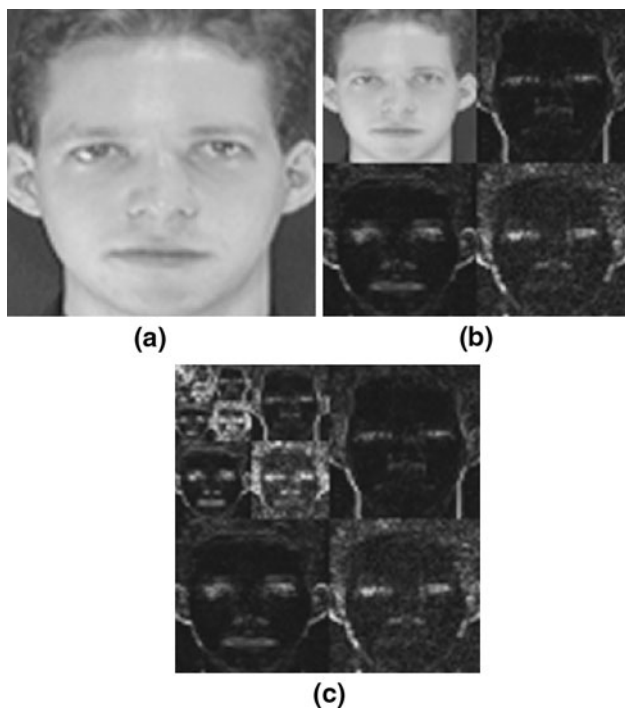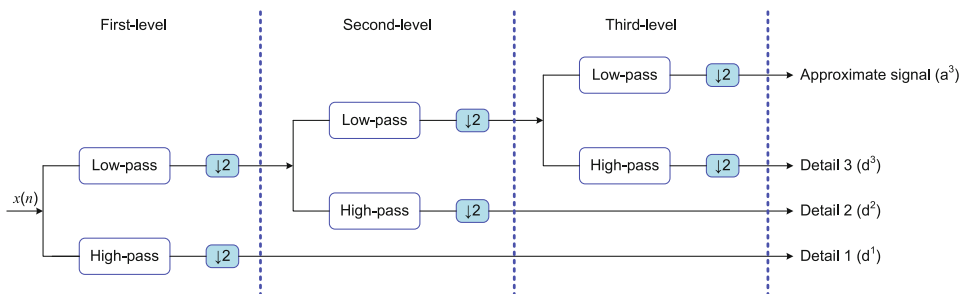
Fig. 2 Wavelet transform of image. **a** Original image. **b** One-level decomposition with Haar wavelet. **c** Complete decomposition with Haar wavelet

image and it analyses each one individually. Consequently, a face image x of dimension $i \times j$ is converted to a column vector of length N, where $N = ij$.

$$x = \begin{bmatrix} p_{1,1} \\ p_{1,2} \\ \vdots \\ p_{1,j} \\ p_{2,1} \\ \vdots \\ p_{i,j-1} \\ p_{i,j} \end{bmatrix} \quad (2)$$

$p_{s,t}$ corresponds to a pixel from the sth row and the tth column. A set of M face images $\{x_i\}$ may be represented as a matrix X of dimension $N \times M$, where,

$$X = [x_1 x_2 x_3 \ldots x_M] \quad (3)$$

The 'average' face is calculated and subtracted from each face in X, giving X',

$$X' = [(x_1 - \bar{x})(x_2 - \bar{x})(x_3 - \bar{x})\ldots(x_M - \bar{x})] \quad (4)$$

The principal components of this set are found by calculating the eigenvectors of the covariance matrix C, where,

$$C = \sum_{i=1}^{M} X'X'^T \quad (5)$$

The calculated eigenvectors are used as an orthogonal basis to represent the training set faces. In face recognition applications, the eigenvectors are known as eigenfaces, as their appearance is visually similar to faces when viewed in the form of a 2-D matrix. A selection of eigenfaces can be seen in Fig. 3.

2.3 HWT implementation using pipelined direct mapping

The HWT wavelet is simple and computationally cheap because it can be implemented by few integer additions, subtractions, and shift operations [15]. To justify the significance of hardware implementation, this wavelet is selected because of its simplistic nature, and mathematical features.

The mathematical features of the basis are as follows: the most simplistic wavelet basis, can be implemented using pairwise averaging and differencing, both unitary and orthogonal, and also it has compact support. Calculation for both processes are described in (6) and (7), where $i = 0 \ldots (\frac{N}{2} - 1)$.

$$H_i = \left(\frac{a_{2*i} + a_{2*i+1}}{2} \right) \quad (6)$$

$$H_{(\frac{N}{2}+i)} = (a_{2*i} - a_{2*i+1}) \quad (7)$$

From implementation point of view, the 1-D HWT flow diagram with N-inputs sample for pipelined direct mapping is shown in Fig. 4, with "Avg." and "Diff." refer for average and differencing processes, respectively.

Fig. 3 A selection of eigenfaces
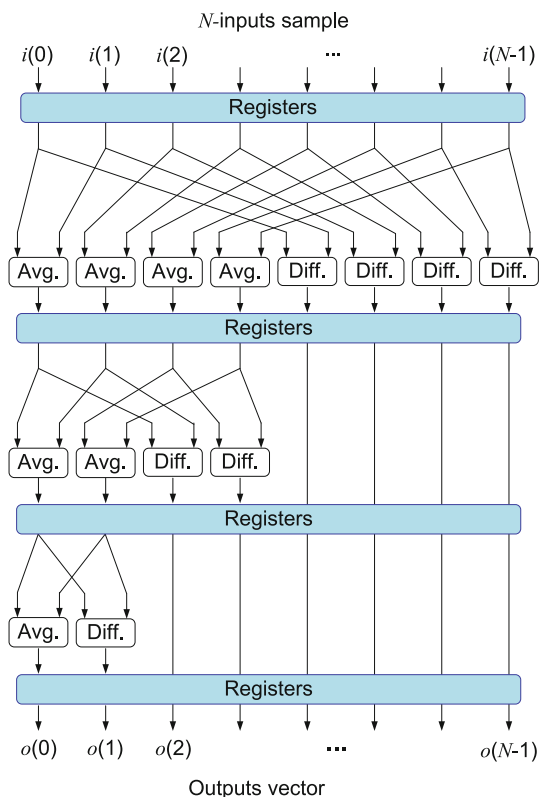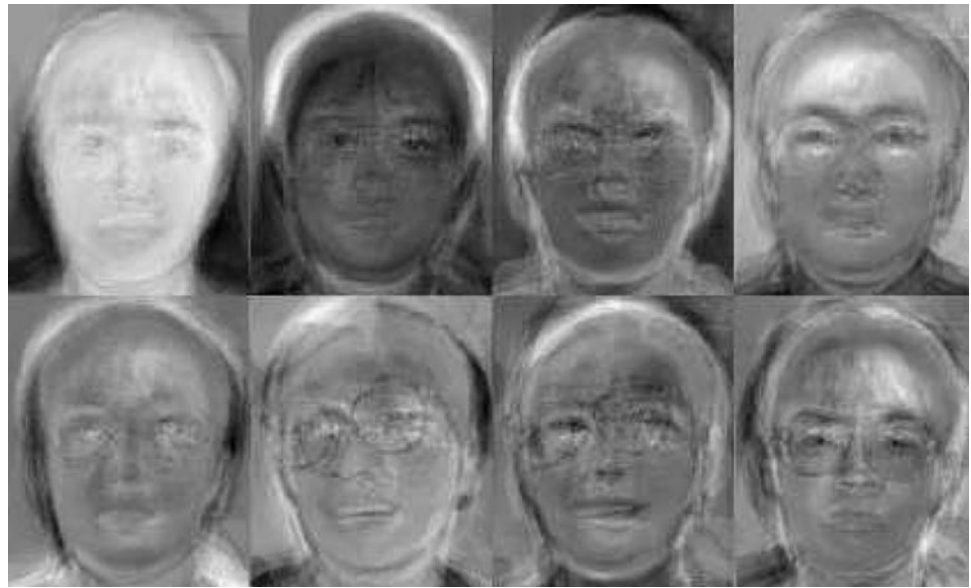
Fig. 4 1-D HWT flow diagram with *N*-inputs sample for direct mapped architecture [16]

3 Software simulation and results analysis

In the following sections, an explanation for each concept and the summary of results are presented.

3.1 DWT feature selection

3.1.1 Concepts

To assess whether DWT can enhance face recognition system performance, a study is performed which attempts to determine how to employ it for this purpose. A number of variables are assessed:

1. *Quadrant*—which DWT quadrant(s) should be used for feature extraction?
2. *Scale*—which scale(s) should be used for feature extraction? and
3. *Filter*—which wavelet filters produce the best results?

Each experiment is performed on coefficients taken from a specific wavelet scale and quadrant and conducted on the AT&T database.

Five randomly selected training images are used for each individual, with the remaining five being used as probe images. Other than minor rescaling, the images undergo no pre-processing with two filters are adopted, Haar and biorthogonal 4.4. PCA is then employed, hence reducing the feature set further.

The Haar wavelet has been chosen for its simplicity, whilst the biorthogonal 4.4 wavelet to represent a more sophisticated filter. Results are being evaluated for the first five scales, since the sixth scale are significantly lower, due to the reduced number of coefficients at this scale.

3.1.2 Summary of results

In brief, the results achieved for these experiments help to guide decisions regarding the experiments that are still to
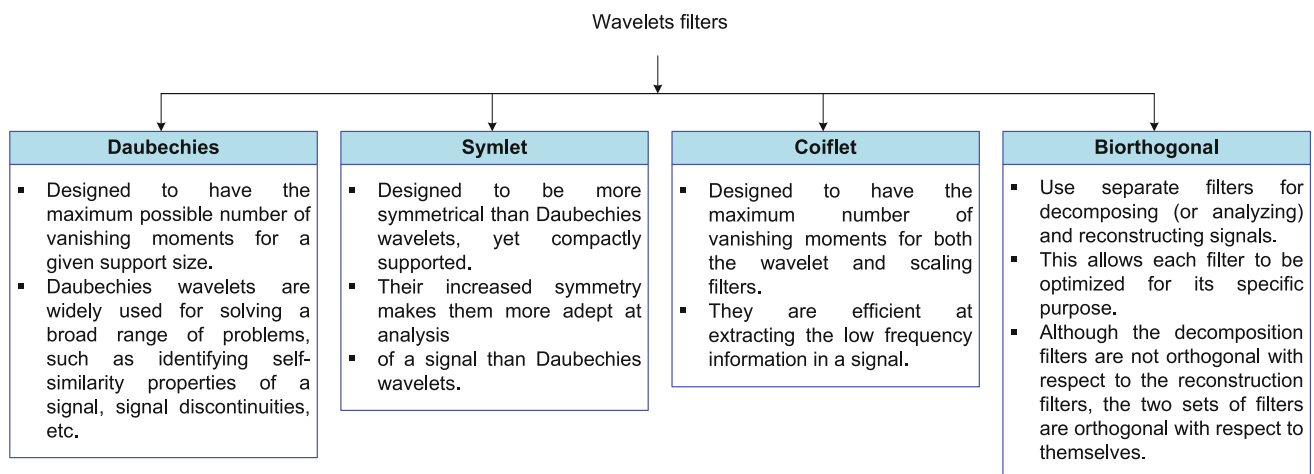
Fig. 5 Wavelets filters descriptions

be performed. When DWT coefficients are used for training a PCA-based recognition system, those from the LL quadrant appears to be much more discriminative in the process of face classification. As other quadrants isolate high-frequency features such as edges, small errors in alignment or facial expression between the images will significantly detract from accuracy.

Conversely, the LL quadrant benefits from the removal of high-frequency features, therefore, any quadrants other than LL need not be investigated further and the effect of scale in the LL quadrant is less clear. Although the third scale produced best results for both wavelet filters tested, there was less variation between results for different scales than there was for different quadrants. It would, therefore, be appropriate to investigate the effect of scale further in remaining experiments.

3.2 DWT filter choice

3.2.1 Concepts

A study is performed to determine whether the choice of wavelet filter has a significant effect on recognition accuracy. Various wavelet families exist, each providing a different compromise between compactness and smoothness. Most wavelets can be described as orthonormal, meaning that they have a unit magnitude and are orthogonal. With a unit magnitude, the convolution of a signal with a wavelet does not change the total energy of the signal. Orthogonality indicates that the inner product of the wavelet basis functions at different scales is zero. A signal can, therefore, be completely represented using a finite number of wavelet basis functions. The same wavelet filters are generally used for decomposition and reconstruction.

Five wavelets are tested from each of the wavelet families shown in Fig. 5. MATLAB is used for experimentation and the filters are provided by the MATLAB wavelet toolbox. As before, the AT&T database is used for experimentation, with five training images and five testing images used for each individual. Only the LL quadrant is used for feature extraction, at scales one to five.

3.2.2 Summary of results

Choice of wavelet family seems to have little effect on the maximum possible recognition rate—filters from the Daubechies and biorthogonal wavelet families matched up to 96.5% of faces correctly, whereas filters from the symlet and coiflet families recognised 97%. The choice of filter within a wavelet family seems to be more significant. For example, although the biorthogonal 5.5 wavelet matches up to 96.5% of faces correctly, the biorthogonal 3.3 wavelet only reaches 93%. The exact nature of the relationship between wavelet and recognition performance however is unclear.

The number of non-zero coefficients in a wavelet filter (known as support size) has a number of effects on the performance of the wavelet. Filters with a larger support size are more adept at analysing and representing complex features contained within the signal/image, however, they are more likely to be affected by artefacts at the edge of the image. Computational complexity of the wavelet transform is also increased when filters with larger support sizes are used.

Based on the results obtained to investigate the choice of wavelet filter with the effect on recognition accuracy, it is clear that DWT has the potential to significantly enhance recognition rates for PCA-based face recognition. For the AT&T database, maximum recognition rates increase from 93% for recognition in the spatial domain to 97% in the wavelet domain. There is not a substantial difference between recognition rates for the wavelet families tested, although coiflet filters produced slightly more consistent

Fig. 6 System overview
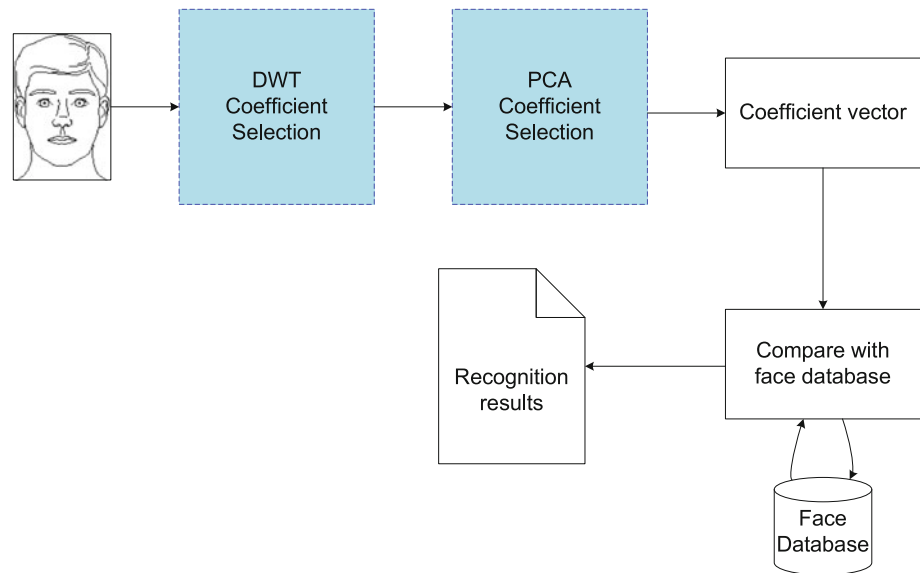
results. Across all the tested wavelet filters, there was no strong correlation between the support size of the low-pass filter and the results. Scale did appear to have an effect on results, with the second scale slightly outperforming the third and fourth scales. The first scale produced slightly lower results, with the fifth scale performing significantly worse.

3.3 Optimising features by coefficient selection

3.3.1 Concepts

The recognition approach is based on standard DWT/PCA face recognition as shown in Fig. 6. In the systems, face images firstly undergo DWT coefficient selection, followed by PCA coefficient selection. The output from this stage is a coefficient vector, which is compared with those of the gallery face images. Recognition results are returned as the identities of the most likely matches in the database.

The purpose of DWT coefficient selection is to select the most discriminative DWT coefficients. Each training image undergoes wavelet decomposition to a specified scale, with the low-pass coefficients being selected to form the image's observation vector. The distribution of these coefficient values is then examined to determine each coefficient's discriminative power. The inter- and intra-class standard deviations for each coefficient are calculated and the ratio of these two values is determined. This ratio indicates how tightly the coefficient's values are clustered within each class, compared to the spread within the complete training dataset. The selection of DWT coefficients is, therefore, based on the maximisation of the following criterion:

$$J = \frac{\sigma_{\text{inter}}(A_m)}{\sigma_{\text{intra}}(A_m)} \tag{8}$$

where $\sigma_{\text{inter}}(A_m)$ and $\sigma_{\text{intra}}(A_m)$ represent inter- and intra-subject standard deviation spanned by DWT coefficients in the feature space A_m, respectively. The DWT coefficients with the highest ratios are the most discriminative and chosen for recognition.

The approach adopted for this study is based on the inter- to intra-class standard deviation ratios. As with DWT coefficient selection, the ratios of inter- to intra-class standard deviations are calculated. Projection coefficients with the highest ratios indicate that the associated eigenvector is highly discriminative and may contribute to better recognition accuracy. This method eliminates the need to guess which eigenvectors represent mostly variation in image illumination. Once training is complete and the most discriminative eigenvectors have been selected, classification can be performed using a simple distance measure, such as Euclidean. The adoption of this approach brings together similar coefficient selection strategies for both stages of the feature vector selection—DWT coefficient selection and PCA eigenvector selection.

Experiments are performed, which determine the benefits of DWT coefficient selection and PCA eigenvector selection separately, as well as in a combined framework. As the technique is more suited to face data sets with little variation in pose/location, the AT&T database of faces is used for experimentation. The images contain variation in lighting, expression and facial details (e.g., glasses/no glasses). For the experiments described in this study, five images for each individual are used for system training, with the other five used for testing.

Table 1 Comparison of DWT coefficient selection recognition rates with those of standard DWT/PCA approach, along with percentages of DWT coefficients required to achieve maximum rate

Wavelet	Scale	Recognition rate (%)		Increase (%)	Coefficients required (%)
		Standard approach	Coefficient selection		
Haar	1	93	95	2	66
	2	94	95	1	50
	3	95	96	1	58
	4	93	95.5	2.5	68
Biorthogonal 4.4	1	94	96	2	69
	2	94.5	96	1.5	78
	3	94.5	96.5	2	83
	4	93	94	1	95
Coiflet 3	1	94	97	3	73
	2	95	97	2	85
	3	95	97	2	98
	4	96	96	0	95
Daubechies 10	1	94	95.5	1.5	66
	2	96.5	97.5	1	99
	3	94	96.5	2.5	75
	4	95.5	97	1.5	98
Symlet 10	1	95.5	96.5	1	99
	2	96.5	96.5	0	90
	3	95	95	0	90
	4	95.5	95.5	0	92
				Average increase (%):	1.37

3.3.2 Summary of results

Different wavelet filters and decomposition levels from one to four are investigated. Selection percentages from 1 to 100% are tested and PCA is used for classification. Where the selection percentage is 100%, this is equivalent to no coefficient selection being applied. Table 1 shows that DWT coefficient selection has increased maximum recognition rate in 16 out of the 20 cases tested.

As Table 2 shows, the approach described compares well with other techniques from the literature that have used this training set. It should be noted that although the AT&T database is relatively small, the technique could be extended to other face databases. However, the coefficient selection approach is particularly suited to data sets with little variation in pose and alignment, therefore, images would have to undergo a normalisation step prior to recognition. If this was performed, it is expected that results for other databases would be similar to those for the AT&T database.

3.4 Feature threshold

A study is performed to investigate ways of choosing the DWT coefficient selection threshold. Although the

Table 2 Comparative results on AT&T database

Method	Accuracy (%)	References
DCT/HMM	84	[17]
ICA	85	[18]
Weighted PCA	88	[19]
Gabor filters and rank correlation	91.5	[20]
2-D PHMM	94.5	[21]
NMF	96	[22]
TNPDP	96.5	[23]
LFA	97	[24]
DWT/PCA with coefficient selection	97.5	Proposed

recognition increases offered by DWT coefficient selection are significant, they are only achievable through a judicious choice of threshold. The maximum possible increases in accuracy offered by DWT coefficient selection can be seen in Table 1.

Increases in recognition accuracy range from 0 to 3%, with the average increase being 1.37%. However, the results presented are the best for each wavelet and scale, found after tests employing varying numbers of DWT coefficients. For coefficient selection to be viable, the number of DWT coefficients to use as features must be

chosen automatically. Two approaches are investigated for choosing this threshold.

3.4.1 Percentage midpoint average (PMA)

The first approach is referred to as PMA. PMA assumes that a number of tests runs have been carried out with appropriate wavelets and scales, and full accuracy data obtained. For each test set, the minimum percentage of DWT coefficients required to produce the maximum recognition accuracy is recorded. The highest percentage of DWT coefficients producing the same maximum accuracy is also noted. The average of these two figures is then calculated, as the percentage midpoint for the current test set. The average of the percentage midpoints for all the test runs is calculated, with this percentage being chosen as the selection threshold.

Tests are performed on the AT&T database to determine the effectiveness of this approach. The PMA value is calculated from recognition results obtained previously, and determined to be 81.36%. DWT coefficient selection results, using 81.36% of coefficients, are shown in Table 3. The results indicate that this approach is not effective, with recognition accuracy decreasing by an average of 0.025% from the results obtained without DWT coefficient selection. This is not unexpected, as the approach is not sophisticated. It assumes that the same percentage of coefficients should be chosen in each case, regardless of the choice of wavelet filter and scale or the individual characteristics of the data set, such as the amount of background (non-face) in the image.

3.4.2 Optimal ratio average (ORA)

The second approach is referred to as ORA. As with PMA, ORA assumes that a number of tests runs have been carried out with appropriate wavelets and scales, and full accuracy data obtained. As explained previously, DWT coefficient selection operates by calculating the ratios of inter- to intra-class standard deviations for each coefficient: this value is used to select the most discriminative coefficients. In ORA, the cut-off ratio that produces the highest recognition rate for each test run is recorded. The average of the cut-off ratios for all test runs is chosen as the selection threshold.

Tests are performed on the AT&T database to determine the effectiveness of this approach. The ratio threshold value is calculated from the DWT coefficient selection results obtained previously. Unlike with PMA, a different percentage of DWT coefficients may be chosen for each wavelet and scale, depending on how discriminative its coefficients are. Results are provided in Table 4 and indicate that the approach is effective, increasing recognition accuracy by an average of 0.6% over recognition without

Table 3 Maximum recognition rates using DWT coefficient selection with PMA threshold

Wavelet	Scale	Recognition rate (%)		Increase (%)
		All coefficients	PMA	
Haar	1	93	94	1
	2	94	94	0
	3	95	94.5	−0.5
	4	93	94.5	1.5
Biorthogonal 4.4	1	94	95	1
	2	94.5	95	0.5
	3	94.5	96.5	2
	4	93	92	−1
Coiflet 3	1	94	94.5	0.5
	2	95	97	2
	3	95	94	−1
	4	96	94	−2
Daubechies 10	1	94	94.5	0.5
	2	96.5	96	−0.5
	3	94	96.5	2.5
	4	95.5	93	−2.5
Symlet 10	1	95.5	95.5	0
	2	96.5	95	−1.5
	3	95	93.5	−1.5
	4	95.5	94	−1.5
		Average increase (%):		−0.025

DWT coefficient selection. However, this is <50% of the maximum possible increase of 1.37% that DWT coefficient selection could provide. Although ORA is more flexible than PMA in handling varying datasets, it is likely that an optimised system would utilise one specific wavelet and scale for both system training and identification. This would allow a more relevant threshold ratio to be chosen, which would increase recognition accuracy.

4 FPGA-based IP core implementation and results analysis

An overview of the IP core implementations including the proposed system applications and architectures, its FPGA implementation as well as the results analysis are discussed in the following sections.

4.1 Proposed system applications

Figure 7 illustrates an overview of the proposed system for both the trained and after the training phase. To accelerate the processes involved in face recognition system, two

Table 4 Maximum recognition rates using DWT coefficient selection with ORA threshold

Wavelet	Scale	Recognition rate (%)		Increase (%)
		All coefficients	ORA	
Haar	1	93	94.5	1.5
	2	94	94.5	0.5
	3	95	94.5	−0.5
	4	93	94	1
Biorthogonal 4.4	1	94	95.5	1.5
	2	94.5	95	0.5
	3	94.5	95.5	1
	4	93	93.5	0.5
Coiflet 3	1	94	95.5	1.5
	2	95	96.5	1.5
	3	95	96	1
	4	96	96	0
Daubechies 10	1	94	96	2
	2	96.5	96	−0.5
	3	94	96.5	2.5
	4	95.5	96	0.5
Symlet 10	1	95.5	94.5	−1
	2	96.5	95	−1.5
	3	95	95	0
	4	95.5	95.5	0
			Average increase (%):	0.6

FPGA-based IP core architectures of 2-D HWT have been proposed to transform an image to xth scale.

A high-level overview of the recognition approach adopted is given in Fig. 8a, whilst the generic proposed 2-D HWT architecture is illustrated in Fig. 8b. The whole chain to calculate the 2-D HWT gets an input as a 2-D image with $N \times N$ point, and outputs the coefficients of the $N \times N$ point. To simplify the hardware design, the 2-D HWT IP core is split into two one-dimensional (1-D) HWT calculation cascaded together with transpose modules in between. This is achieved by performing the first 1-D HWT along the rows (columns) of the array followed by 1-D HWT along the columns (rows) of the transformed array. Transposition module stores the transposed coefficients into memory with a fetch unit module that reads back the coefficients for the next calculation.

4.2 Proposed architectures

Both proposed architectures implementation on the FPGA are given in Fig. 9a, b. The implementation of 2-D HWT IP core without DPR defined the entire FPGA devices as
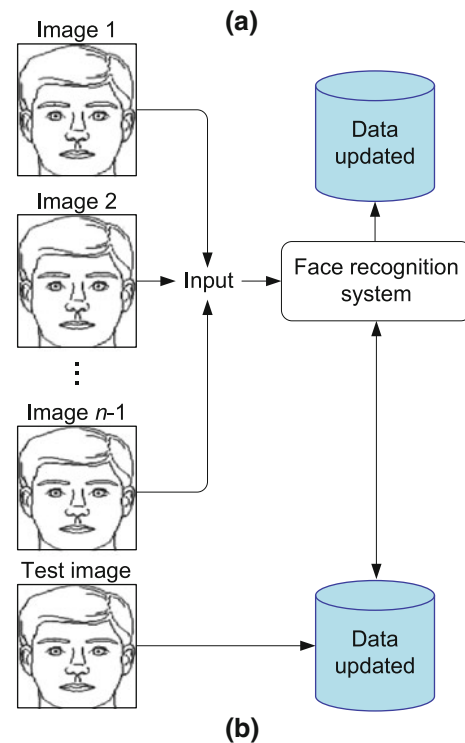
Fig. 7 Proposed system applications. **a** Trained phase. **b** After the training phase

one module. On the other hand, the implementation with DPR method and its framework consists of:

1. Two reconfigurable areas—for the 1-D HWT IP core and transposition module; and
2. A static area—for the data fetch unit and the memory controller (Wishbone compliant).

In both architectures, data fetch unit and HWT IP core are connected with a defined data bit width bus, a request line and back signal free. The fetch unit sends data and the request to the HWT core as long as the free signal is active. HWT and transposition module are connected with the defined data bit width bus and an enable signal. In each cycle and when the enable is active, the data will be transposed and written into the memory.

4.3 2-D HWT and transpose-based computation

The proposed 2-D HWT IP core implementation works as follows. The input to the first 1-D HWT is read row by row, the 1-D HWT is performed on each input vector as they are provided and the calculated values are sent to the transpose module, which calculated the memory addresses for the transposition and stores the data into memory.

The transpose acts as a memory forwarder and performs matrix transpose, since row vectors are provided by the 1-D
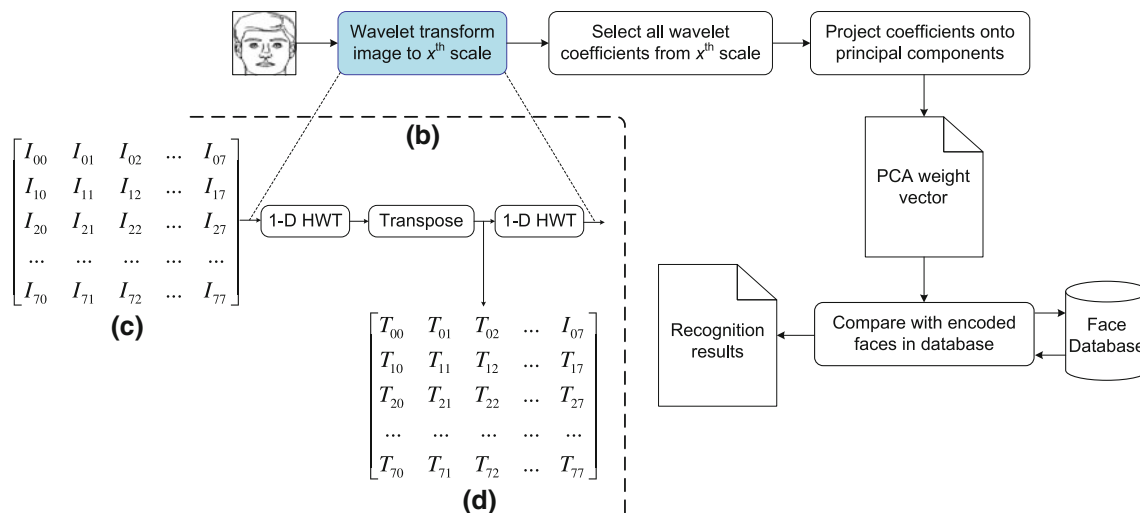
Fig. 8 Proposed system architectures. **a** Overview of recognition approach. **b** Architecture for 2-D HWT IP core with transpose-based computation. **c** Input data for images with $x, y \in [0, 1, \ldots, 7]$. **d** Transpose matrix after transpose with $x, y \in [0, 1, \ldots, 7]$

HWT. After transposition of the resultant matrix, another 1-D HWT is performed on the coefficients which are stored in memory to yield the 2-D HWT coefficients. Algorithm 1 gives the description of the 2-D HWT process.

Algorithm 1 The 2-D HWT pseudo-code

```
1: for row = 1 to norows do
2:    Apply a 1-D HWT column-wise
3: end for
4: for col = 1 to nocols do
5:    Apply a 1-D HWT row-wise
6: end for
```

4.4 2-D HWT with DPR

In this study, the ISE Design Suite 9.2PR and PlanAhead 10.1 [25] are used. With module-based DPR [16], this method has the limitation that all design files and reconfigurable modules must be available to the build environment to build partial modules.

Reconfigurable architectures using DPR technique comprises of several reconfigurable processing modules (RPM), a reconfigurable interface, an off-chip memory and micro blaze (μblaze). The system is connected to the host personal computer (PC) via peripheral component interconnect (PCI) express [16]. μblaze is a soft processor core designed for Xilinx FPGAs [25].

The reconfigurable processing modules allow hardware acceleration and can be reconfigured based on the system demand, whilst the communication interface is used to build the interconnection between RPM and the other components.

4.5 FPGA implementation and result analysis

FPGA implementation results for both architectures, analysis and an overview of the advantages offered with DPR technique are presented in the following sections. In this study, Xilinx early access partial reconfiguration (EAPR) design flow [26] is used as a design flow reference, and these two architectures are implemented on the Xilinx Virtex-5 (XC5VLX110T-3FF1136).

In the face recognition system, the inputs are various size of images, hence different transform sizes ($N = 8, 16, 32, 64$ and 128) have been used to examine the relationship of the transform sizes on the area (slices), power consumption (mW) and maximum speed (MHz).

In Table 5, results for both architectures are listed. As an example, for $N = 128$, the implementation with DPR technique yields a significant achievement with better resources used for area as well as better power consumption by 46.67 and 15.96%, respectively. On top of that, DPR technique also gives 4.59% better maximum frequency than without DPR.

To underline the influence of different transform size on area, power consumption and maximum frequency, Figs. 10, 11 and 12 illustrate the relationship for each performance indicator. Results obtained are clearly shown that the proposed 2-D HWT IP core without DPR consumes more area and power. Using DPR technique, better area and power saving can be achieved between 36.68–46.67 and 6.78–15.96%, respectively. Additionally, to visualise the impact of non-partial and partial reconfiguration chip layouts for $N = 16$ and 64 are given in Fig. 13.

DPR is a promising technique for reducing the hardware required as well as improving the performance of the

Fig. 9 Proposed top architecture of 2-D HWT IP core. **a** Without DPR. **b** With DPR
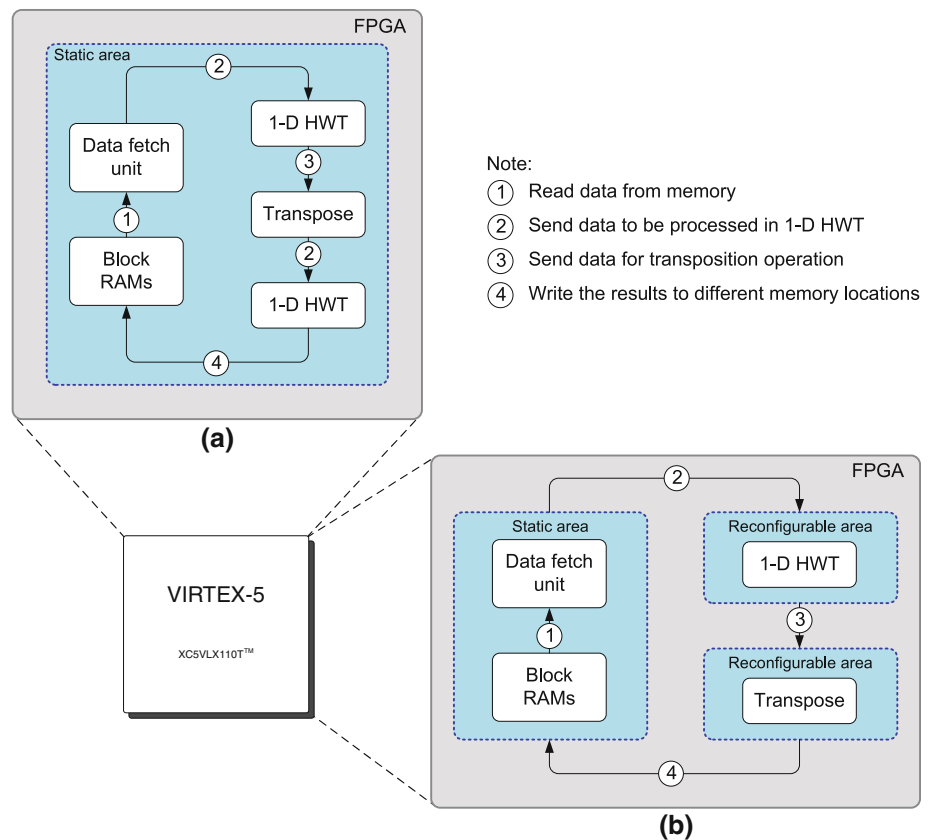
Table 5 Resources utilisation and overall proposed architectures performance on XC5VLX110T-3FF113

Parameters	Proposed 2-D HWT			
	Without DPR		With DPR	
	$N = 8$	$N = 128$	$N = 8$	$N = 128$
Area (slices)	2,180 (3.15%)	38,261 (55.35%)	1,376 (2.00%)	20,403 (29.51%)
Power consumption (mW)	902.64	1772.83	762.55	1489.81
Maximum frequency (MHz)	213.82	164.59	271.15	172.51

system. With this technique, the design can be divided into sub-designs that fit into the available hardware resources and can be uploaded into the reconfigurable hardware when needed [16].

In SRAM-based FPGAs, full-device reconfiguration is required upon power-up [27]. The process of initialisation involves the FPGAs to be programmed with a configuration bitstream file. Partial reconfiguration concept appears after initialisation and works to modify a fraction of the resources by programming the FPGA with a partial bitstream file. Obviously, a full bitstream size is very massive whereas a partial bitstream may represents only 2% of the full bitstream [16, 27–29]. With smaller bitstreams, several advantages can be achieved: reduced reconfiguration time, reduced storage requirements, and dynamic allocation of functionality.

An implementation of the PCA-DWT based face recognition system has been also carried out on the RC10 FPGA prototyping board equipped with the low-power Spartan 3l1500 Xilinx FPGA device and an integrated complementary metal oxide semiconductor (CMOS) camera, which can be deployed for faces acquisition to have a complete SoC solution.

Tests have been performed on the AT&T database with 40 subjects and have revealed a 90% recognition rate with an acceleration of four times compared to the software recognition running on a Intel Dual-Core with 2.42 GHz. The computation time for face recognition on the RC10 board using the AT&T database is 6.45 ms. It is worth mentioning that a further acceleration can be achieved by partitioning the execution of the two building blocks of the proposed system using a software-hardware co-design approach with an efficient host-FPGA communication.
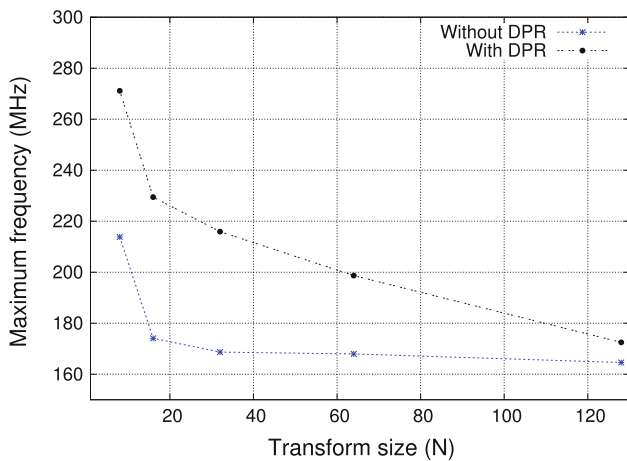
The application of face recognition requires several building blocks for its computationally intensive processes
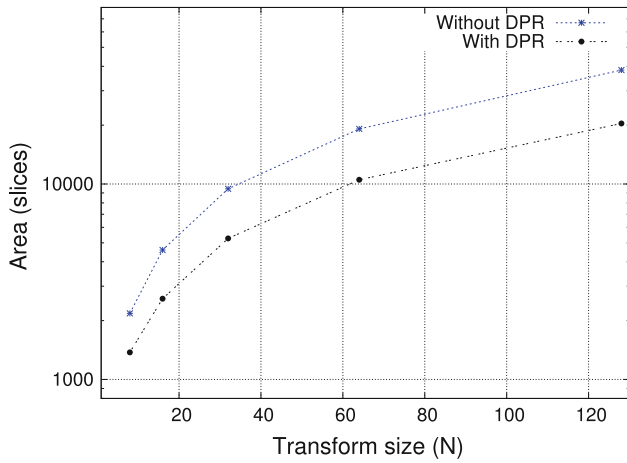
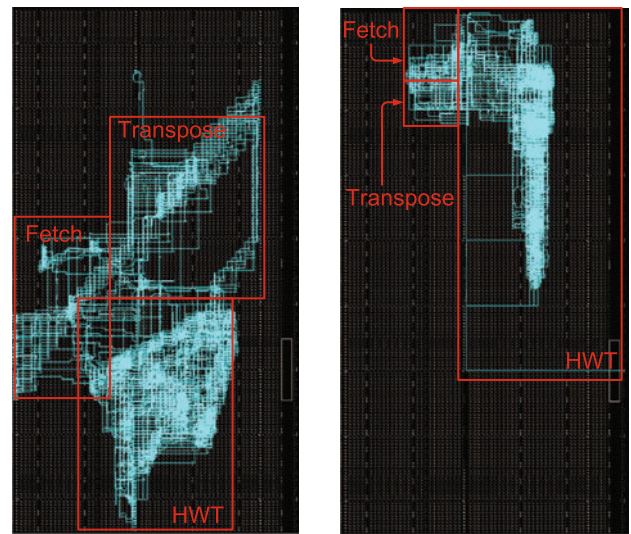**Fig. 10** Influence of transform size on area (slices)
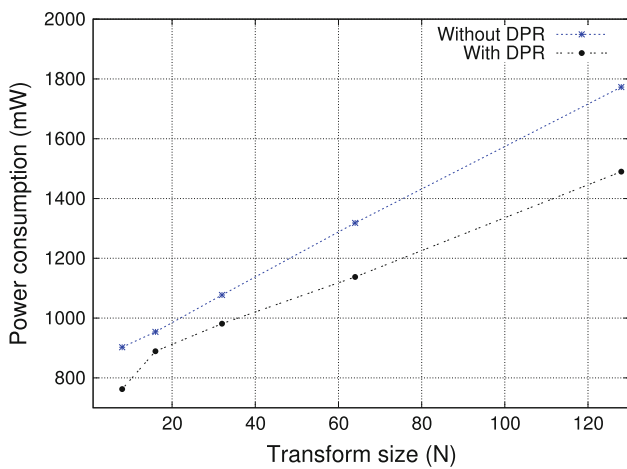
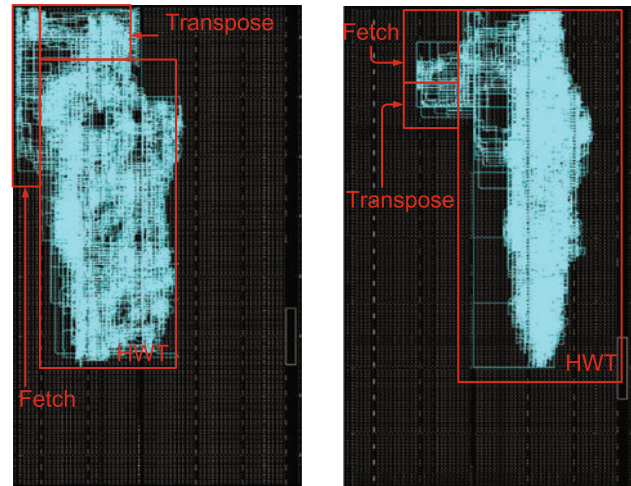**Fig. 11** Influence of transform size on power consumption (mW)

Fig. 12 Influence of transform size on maximum frequency (MHz) for 1-D HWT modules

Fig. 13 Comparison of chip layout for different transform sizes on XC5VLX110T-3FF113

to perform matrix transformation operations. Moreover, complexity in addressing and accessing large databases have resulted in vast challenges from a hardware implementation point of view. To cope with these issues, an FPGA-based architecture with efficient reconfigurability techniques is a promising solution to meet the demands of these applications in terms of speed, size (area), power consumption and throughput.

5 Conclusions

In this research study, two main issues have been addressed: the software simulation of a novel feature vectors construction approach for face recognition using DWT and the IP core implementation of transform block in the face recognition systems.

The first set of experiments performed focused on the choice of DWT features. It reveals that, where direct

coefficient values were used for recognition, the LL quadrant provided the best results. For the wavelet filters tested, the highest recognition rate achieved for this quadrant was 95%. The highest accuracies for the HL, LH and HH quadrants were 78, 74 and 66%, respectively. However, these tests did not provide enough information to indicate whether particular scales perform consistently better than others.

The second set of tests has been designed to examine which wavelet filters were the most effective at extracting features for face recognition with the specified database. The maximum recognition rates were compared for five wavelet filters each from the Daubechies, symlet, Coiflet and biorthogonal wavelet families. LL coefficients were used as features, with the first five scales investigated. The results indicated that there was no strong link between choice of wavelet family and recognition rate, although Coiflet wavelets produced the most consistent performance, across various filters and scales. When the results from all wavelet families and filters were examined together, there was no obvious correlation between the support size of the scaling filter and the maximum recognition rates.

The choice of scale did appear to have some effect, with the second, third and fourth scales outperforming the first scale by a small margin and the fifth scale by a significant margin. In case of feature optimisation by coefficient selections, the results show that DWT coefficient selection has increased maximum recognition rate in 16 out of the 20 cases tested. For instance, recognition accuracy increased from 94 to 97% for the Coiflet 3 wavelet, first scale.

For the feature threshold, two approaches have been investigated, which are PMA and ORA. Results obtained shown that the PMA is an ineffective approach, with recognition accuracy decreasing by an average of 0.025% from the results obtained without DWT coefficient selection. Unlikely, results for ORA approaches indicate better recognition accuracy by an average of 0.6%.

On the contrary, two architectures for 2-D HWT IP cores have been proposed for the transform in the proposed face recognition system based on transpose computation and partial reconfiguration. To sum up, comparative study for both non-partial and partial reconfiguration processes has shown that DPR offers many advantages and lead to a promising solution for implementing computationally intensive applications such as face recognition systems. Using DPR, several large systems are mapped to small hardware resources and the area, power and maximum frequency are optimised and improved.

References

1. Amira, A., Farrell, P.: An automatic face recognition system based on wavelet transforms. In: IEEE International Symposium on Circuits and Systems, 2005. ISCAS 2005, vol. **6**, 6252–6255 (2005)
2. Nicholl, P., Ahmad, A., Amira, A.: A novel feature vectors construction approach for face recognition. In: Gavrilova, M., Tan, C., Moreno, E. (eds.) Transactions on Computational Science XI. Lecture Notes in Computer Science, vol. **6480**, pp. 223–248. Springer, Berlin
3. Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face recognition: a literature survey. ACM Comput. Surv. **35**(4), 399–458 (2003)
4. http://www.cognitec-systems.de. Accessed 10 Jan 2009 (online)
5. http://www.l1id.com. Accessed 10 Jan 2009 (online)
6. http://www.geometrix.com. Accessed 10 Jan 2009 (online)
7. http://www.genextech.com. Accessed 10 Jan 2009 (online)
8. http://www.animetrics.com. Accessed 10 Jan 2009 (online)
9. Dang, P.: VLSI architecture for real-time image and video processing systems. J Real Time Image Process. **1**, 57–62 (2006)
10. Stokes, M.L.: A brief look at FPGAs, GPUs and cell processors. J. Int. Test Eval. Assoc. (ITEA), 9–11 (2007)
11. Todman, T.J., Constantinides, G.A., Wilton, S.J.E., Mencer, O., Luk, W., Cheung, P.Y.K.: Reconfigurable computing: architectures and design methods. IEE Proc. Comp. Digital Tech. **152**(2), 193–207 (2005)
12. Rousseau, B., Manet, P., Galerin, D., Merkenbreack, D., Legat, J.-D., Dedeken, F., Gabriel, Y.: Enabling certification for dynamic partial reconfiguration using a minimal flow. In: Design, Automation Test in Europe Conference Exhibition, 2007. DATE '07, pp. 1–6 (2007)
13. Stollnitz, E.J., DeRose, T.D., Salesin, D.H.: Wavelets for computer graphics: a primer, part 1. IEEE Comput. Graph. Appl. **15**(3), 76–84 (1995)
14. Samaria, F., Harter, A.: Parameterization of a stochastic model for human face identification. In: IEEE Workshop on Applications of Computer Vision, Sarasota, FL (1994)
15. Bajaj, C., Ihm, I., Park, S.: 3D RGB image compression for interactive applications. ACM Trans. Graph. **20**(1), 10–38 (2001)
16. Ahmad, A., Krill, B., Amira, A., Rabah, H.: Efficient architectures for 3D HWT using dynamic partial reconfiguration. J. Syst. Archit. **56**(8), 305–316 (2010)
17. Nefian, A., Hayes M.: Hidden Markov models for face recognition. In: ICASSP98, pp. 2721–2724 (1998)
18. Kim, J., Choi, J., Yi, J., Turk, M.: Effective representation using ICA for face recognition robust to local distortion and partial occlusion. IEEE Trans. Pattern Anal. Mach. Intell. **27**(12), 1977–1981 (2005)
19. Wang, H.Y., Wu, X.J.: Weighted PCA space and its application in face recognition. In: Proceedings of 2005 international conference on machine learning and cybernetics, 2005, Washington, DC, USA, 2005, pp. 4522–4527. IEEE Computer Society, USA
20. Ayinde, O., Yang, Y.H.: Face recognition approach based on rank correlation of Gabor-filtered images. Pattern Recognit. **35**(6), 1275–1289 (2002)
21. Samaria, F.: Face Recognition using Hidden Markov Models. PhD thesis, Cambridge University Engineering Department (1994)
22. Xue, Y., Tong, C.S., Chen, W.S., Zhang, W.: A modified nonnegative matrix factorization algorithm for face recognition. In: ICPR '06: Proceedings of the 18th International Conference on Pattern Recognition, Washington, DC, USA, 2006, pp. 495–498. IEEE Computer Society, USA
23. Lu, J., Tan, Y.P.: Enhanced face recognition using tensor neighborhood preserving discriminant projections. In: 15th IEEE International Conference on Image Processing, 2008. ICIP 2008, pp. 1916–1919 (2008)
24. Ersi, E.F., Zelek, J.S.: Local feature matching for face recognition. In: CRV '06: Proceedings of the 3rd Canadian Conference

on Computer and Robot Vision (CRV '06), Washington, DC, USA, 2006, p. 4. IEEE Computer Society, USA

25. http://www.xilinx.com. Accessed 10 Jan 2008 (online)

26. Lysaght, P., Blodget, B., Mason, J., Young, J., Bridgford, B.: Invited paper: enhanced architectures, design methodologies and CAD tools for dynamic reconfiguration of Xilinx FPGAs. In: International Conference on Field Programmable Logic and Applications, 2006. FPL '06, pp. 1–6. August 2006

27. Parris, M.G.: Optimizing dynamic location realizations of partial reconfiguration of FGPAs. Master thesis, School of Electrical Engineering and Computer Science, University of Central Florida Orlando, 2009

28. Huang, J., Parris, M., Lee, J., Demara, R.F.: Scalable FPGA-based architecture for DCT computation using dynamic partial reconfiguration. ACM Trans. Embed. Comput. Syst. **9**(1), 9–18 (2009)

29. Krill, B., Ahmad, A., Amira, A., Rabah, H.: An efficient FPGA-based dynamic partial reconfiguration design flow and environment for image and signal processing IP cores. Signal Process. Image Commun. **25**(5), 377–387 (2010)

Author Biographies

Afandi Ahmad is a lecturer at the Universiti Tun Hussein Onn Malaysia (UTHM) within the Department of Computer Engineering in the Faculty of Electrical and Engineering. He joined UTHM as a tutor in May 2002 and was appointed as a lecturer in December 2003. He obtained his B.Sc. in electrical engineering (computer technology) from the Universiti Teknologi Malaysia (UTM-ITTHO) and M.Sc. in microelectronics from the Universiti Kebangsaan Malaysia (UKM) in 2002 and 2003, respectively. He received his Ph.D. in electronic and computer engineering from the Brunel University, London in 2010. He has been awarded a number of prizes and travel grants and has published in impact factor's journals as well as five star international conferences. He also actively contributes as a reviewer in IEEE Transaction on Circuit and Systems of Video Technology, IEEE Transaction of Very Large Scale Integration (VLSI), Transactions on Computational Science Journal Springer-Verlag, Springer Journal of Signal Processing Systems and Hindawi Journal of Artificial Intelligence. He is a graduate member of the IEM, professional member of ACM, and member of the IEEE, IET, ASEE and IAENG. His research interests include: embedded computing systems, VLSI DSP, 3-D medical image analysis and diagnosis, partial and dynamic reconfigurable architectures, 3-D compression, forensic computing and also engineering education.

Dr. Abbes Amira has been recently appointed as Associate Professor in the College of Engineering at Qatar University, Qatar. He also holds a readership in embedded systems in the Nanotechnology and Integrated BioEngineering Centre (NIBEC) at the University of Ulster, United Kingdom. From May 2006 to March 2010, he was a senior lecturer at the Brunel University, West London, within the division of Electronic and Computer Engineering. Before he joined Brunel University, he has held a lectureship in computer science at Queen's University, Belfast (QUB) since November 2001. He received his Ph.D. in computer engineering from Queen's University Belfast in 2001. He has been awarded a number of grants from government and industry and has published over 160 publications during his career to date. He has been invited to give talks and tutorials at universities and international conferences and to be chair, program committee for a number of conferences. He was the Conference Chair of ECV2011, one of the tutorial presenters at ICIP 2009, Program Chair of ECVW2010, and Program Co-Chair of DELTA 2008 and IMVIP 2005. He is also one of the 2008 VARIAN prize recipients. Dr. Amira currently holds a visiting associate professor position at the University of Tun Hussein Onn in Malaysia, and has held a visiting professor position at University of Nancy-France in April 2011. He is a senior member of the IEEE, senior member of ACM, Fellow of IET and Fellow of the Higher Education Academy. His research interests include: embedded systems, connected health, high performance reconfigurable computing, image processing, and medical and security applications.