Neuromorphic Hardware Accelerated Adaptive Authentication System

Manan Suri, Vivek Parmar, Akshay Singla, Rishabh Malviya, Surag Nair
Department of Electrical Engineering
Indian Institute of Technology, Delhi
Email: manansuri@ee.iitd.ac.in

***Abstract*—In this paper we present a multimodal authentication (person identification) system based on simultaneous recognition of face and speech data using a novel bio-inspired architecture powered by the CM1K chip. The CM1K chip has a constant recognition time irrespective of the size of the knowledge base, which gives massive time gains in learning and recognition over software implementations of similar methods. We demonstrate a system utilizing the CM1K chip as a neural network accelerator along with data pre-processing done by a desktop PC. The system realized consumes energy of the order: 668 μJ for learning and 487 μJ for recognition, while operating at 25 MHz. The classification test accuracy of the system is approximately 91%.**

I. Introduction

Hardware implementation of neural networks and bio-inspired evolvable systems has undergone rapid development over the last few years. Unlike conventional Von-Neumann architecture that is sequential in nature, artificial neural networks (ANNs) profit from massively parallel processing. Despite tremendous growth in the digital computing power of general-purpose processors, dedicated neural network hardware stands promising for certain specialized real-time, data-intensive, asynchronous applications, such as image processing, speech synthesis and analysis, pattern recognition, high energy physics and so on [1]. Recent implementations range from purely digitized neural networks, analog circuit based solvers to emerging hybrid CMOS/non-CMOS designs involving non-volatile memory (NVM) technologies such as MRAM, OXRAM, etc.[2][3][4][5].

In this paper, we propose a neuromorphic hardware based approach for person identification (authentication), and compare its performance with standard software based implementation. We implemented the authentication system using an architecture built on the CM1K neuromorphic chip. Due to its inherent parallelism, the CM1K chip performs recognition in constant time irrespective of the size of the knowledge base. The CM1K has been used in numerous applications such as pattern recognition, target tracking and industrial automation.

We choose the case of person identification via simultaneous face and speech recognition, as it finds a variety of applications in surveillance, authentication and security systems. Compared to other biometric identification techniques such as fingerprint analysis or iris detection which require active cooperation of participants, face and speech recognition are easier as they often do not require the participants to cooperate [6]. Moreover exponentially increasing database sizes call for faster and more power efficient implementations of face and speech recognition algorithms. Software implementations are impaired by the current paradigm of Von Neumann computing resulting in slower training and recognition times[7].

For a comprehensive analysis, we used existing face and speech recognition algorithms utilizing techniques such as dimensionality reduction and feature extraction to train a fast and efficient classifier on the CM1K chip. For face recognition, several pre-processing techniques like Wavelet and Gabor Transforms were compared. Popular dimensionality reduction techniques like PCA and LDA were also implemented.

Section II explains basic features and working of the neuromorphic CM1K chip. Section III describes implementation of the Speech Recognition technique we used and results obtained. Section IV describes the implementation of our Face Recognition technique, and the results. Section V discusses the combined simultaneous Face and Speech Recognition application and complete authentication application results. Section VI presents the key conclusions.

II. CM1K Chip

A. Introduction

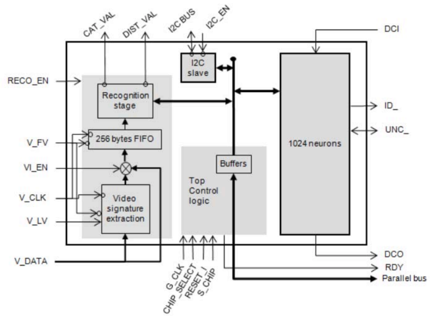
The CM1K chip developed by General Vision is based on the ZISC (Zero Instruction Set Computing) architecture[8]. Each chip consists of 1024 identical "neurons" which are capable of storing and recognizing vectors of length up to 256 bytes. All neurons operate in parallel and collaborate with each other through a bi-directional neuron bus. Each neuron incorporates information from all the other neurons into its own learning logic and into its response logic as shown in Fig. 2.

If several neurons recognize a pattern (or fire), their responses can be retrieved automatically in increasing order of distance from the broadcast vector. This retrieval is independent of the number of training points in the knowledge base. Multiple CM1K chips can be daisy-chained to increase the number of available neurons, without affecting

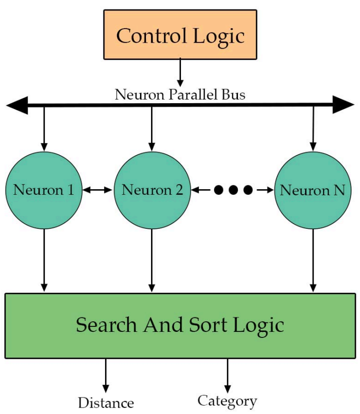
978-1-4799-7560-0/15/$31 ©2015 IEEE

978-1-4799-7560-0/15 $31.00 © 2015 IEEE
DOI 10.1109/SSCI.2015.173

the recognition or learn time. The information which can be read from a firing neuron includes its distance, category and neuron identifier as shown in Fig.1b. If the response of several or all firing neurons is polled, this data can be consolidated to make a more sophisticated decision weighing the cost of the uncertainty.



(a) Structure of CM1K chip Neural Network [9].



(b) Block Diagram of CM1K chip recognition [7].

Fig. 1: CM1K chip structural and functional block diagram.

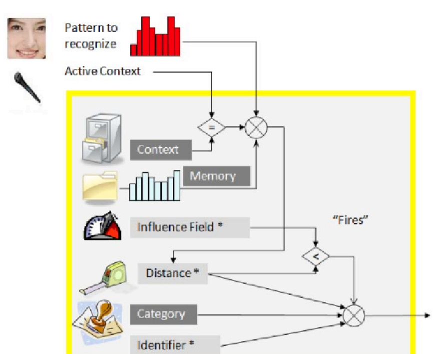


Fig. 2: CM1K neuron response [10].

B. Learning Algorithm and Methodology

The CM1K chip architecture allows the neurons to maintain multiple separate unrelated knowledge bases simultaneously by associating a context with each neuron. When a vector is broadcast, only those neurons react which have matching contexts with that of the vector. This allows recognition across multiple unrelated knowledge bases, e.g. speech and face image data of people to happen simultaneously. The calculation of distance between the stored (S) and broadcast vector (V) can be calculated using two norms: L1 (Manhattan distance) and Lsup, where

$$D_{\mathrm{L1}} = \Sigma|V_i - S_i| \quad (1)$$

$$D_{\mathrm{Lsup}} = \max|V_i - S_i| \quad (2)$$

The CM1K chip offers two different classifiers for neuron learning and recognition: K-Nearest Neighbours algorithm (k-NN) and Radial Basis Function (RBF). The decision space mapping resulting from the two different classifiers is shown in Fig.3.

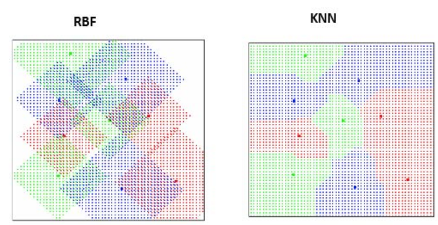


Fig. 3: CM1K Neuron Decision Space Mapping Based on learning algorithm. [11]

1) K-Nearest Neighbours algorithm: This method classifies objects based on the closest matches in the knowledge base. Since the decision space mapping doesn't get continuously modified during learning, the knowledge base is simply loaded to the neurons. The parallel architecture of CM1K allows it to retrieve the top match in constant time irrespective of the number of stored examples in the knowledge base. As the components of the input vector are broadcast one by one, the neurons update their distance values simultaneously. The K nearest neighbours can then be read successively from the Distance Register, and each read takes a fixed time [10].

2) Radial Basis Function: This classifier allows the formation of a complex non-linear decision space mapping which uses radial basis functions as activation functions. This requires a model generator internal to the neurons which is used when the knowledge base is learned vector by vector. Each neuron has its own influence field. When a vector is broadcast to be recognised, only those neurons fire for which the distance of the vector being broadcast from their stored vector is less than their influence field. If no neuron fires, the recognition status is 'Unknown'. However, when multiple neurons fire, the recognition status is

'Identified' if the firing neurons have the same category value, and 'Uncertain' if all the firing neurons do not have the same category value.[10].

III. Speech Recognition

For speech recognition, we used a subset of the data from the CSTR VCTK (Center for Speech Technology Research Voice Cloning Toolkit) Speech Corpus which includes speech data uttered by 109 native speakers of English with various accents [12].

As our speech recognition system was intended to be text-dependent, ten instances of the word 'rainbow' (uttered abundantly in the Rainbow Passage) were spliced out from the original recording of each speaker to make a custom dataset for 15 speakers. These signals were then pre-processed into vectors of 256 elements before training/testing on the CM1K chip.

A. *Pre-processing*

The features extracted from each signal were the Mel-Frequency Cepstral Coefficients (MFCCs) [13] [14] [15]. The speech signals for this experiment (utterances of single words) were small enough to be broken into sixteen 20-40ms pieces, which contributed 16 MFCCs each. This gave a 256-element vector for each data point. The steps involved in this particular implementation were:

1. Fast Fourier Transform (FFT) with a Hamming windowing function converted the waveform in n-space to a distribution of the power amplitudes in k-space (across different frequencies)
2. Converting the power frequency spectrum to a log scaled (Mel-Frequency scale) spectrum. This mimics the behavior of the human ear. By using an implementation which employed the Mel triangular filter-banks, this step was able to extract discrete 20-element spectra from each of the originally continuous power spectra:

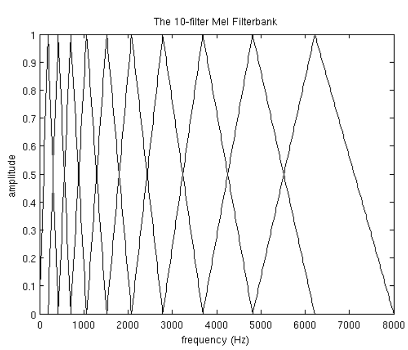


Fig. 4: A typical $8000Hz$ Mel-frequency filterbank. The equation characterizing the conversion of frequencies to the Mel-frequency scale is: $M(f) = 1125\ln(1 + \frac{f}{700})$[16].

3. Power magnitudes (on the y-axis of the power spectrum) were also log-scaled. Again, this was done to mimic the response of the human ear.
4. Finally, a Discrete Cosine Transform of the scaled spectra yielded the Mel-Frequency Cepstral Coefficients for each speech signal.

B. *Word Recognition (RBF Neural Network)*

In addition to identifying the speaker, we modified the system so that it could also verify that the word being spoken was indeed 'rainbow'. This exploited the RBF neural network capability of the hardware. For testing the word recognition of the RBF neural network, three utterances of the words 'blue cheese', 'train station', and 'raindrop' were appended to every speaker's testing dataset. These words were chosen to sound increasingly similar to 'rainbow'.

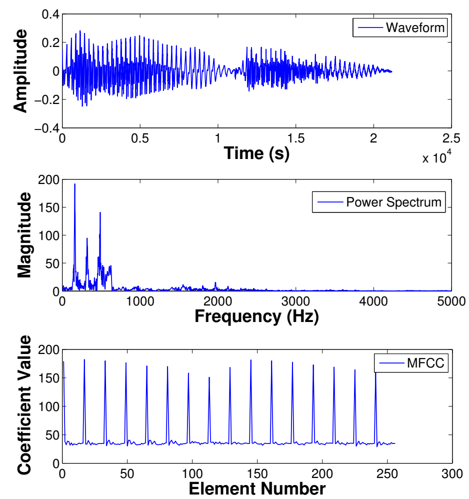


Fig. 5: The pre-processing transformations of the first utterance of 'rainbow' by the first speaker. The above behavior is typical of all of the speech signals.

The final work-flow for the combined system (word and speaker recognition) is outlined below:

1. A neural network was created for speaker categorization. This was done in SR (Save and Restore) mode, wherein the training vectors' information is simply written to the neurons without explicitly optimizing the decision-space mapping.
2. The training set is used to train another binary-output RBF neural network that can tell whether the word spoken is 'rainbow' or something completely different. Learning is done in iterative RBF mode.
3. Speaker recognition is done using the kNN classifier. The classifier employed classified the testing vectors by computing the mode of the nearest k neighbors. The system was found to work best for $k = 1$, and moderately well for $k = 5$.
4. For word recognition, the second network determines whether or not the distance of the input vector from

the network's neurons is below a neuron-specific threshold (determined during learning of these neurons). If so, these neurons fire and the RBF classifier returns the closest k firing neurons. If no neurons fire, the feature vector being tested is classified as 'not-rainbow'.

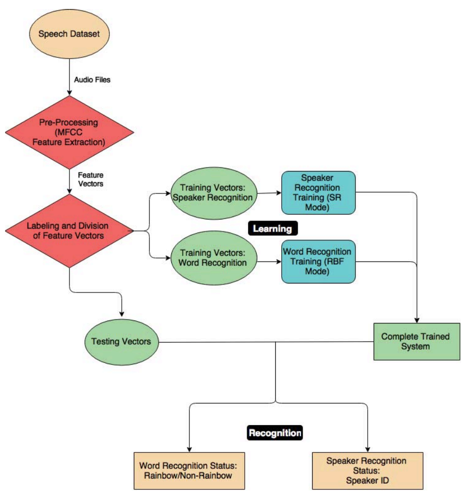


Fig. 6: A schematic of the working of the speaker + word recognition system.

C. Results

Given the small number of vectors in the training/testing datasets, there is a possibility of overfitting. To test this, 50 different random divisions of the procured data into training and testing subsets was made and run through the system. For every division, each speaker's 10 utterances of 'rainbow' were divided such that the first 7 vectors went into the training set, while the remaining 3 were combined with the non-rainbow utterances to create a test set. The performance statistics for the 50 runs were determined for 15 speakers (see Tab.I). Mean accuracy was around 88%.

TABLE I: Accuracy statistics for random runs with 15 speakers.

Mean	87.98
Standard Deviation	2.23
Minimum	83.33
Maximum	92.22

Accuracy decreased with increasing number of speakers (Fig. 7). With increasing number of speakers the decision space becomes more complex, thereby increasing the amount of confusion. The timing curves (Fig. 8) highlight the advantage of using CM1K. The hardware implementation of the speaker recognition system (kNN recognition mode) was compared against the ANN C++ library and an implementation of kNN classifier algorithm

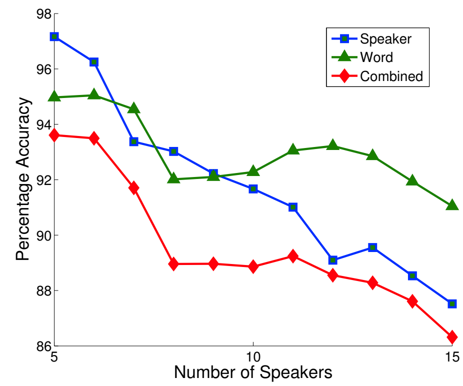


Fig. 7: Test accuracy vs number of speakers.

in MATLAB. Both software implementations were tested on an Intel i7 4th Generation PC with 16 GB RAM running Windows 7. The results clearly show that we achieve a speedup of about 5 times in recognition even while including the transport delay introduced by the use of a non-optimal USB.

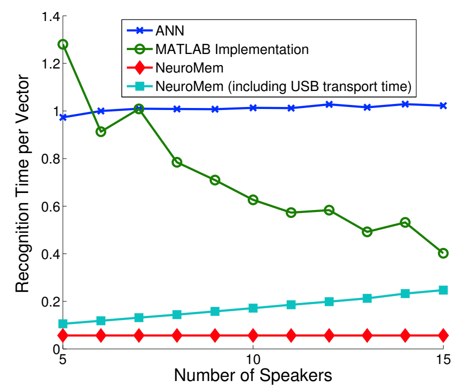


Fig. 8: Timing performance for speaker recognition.

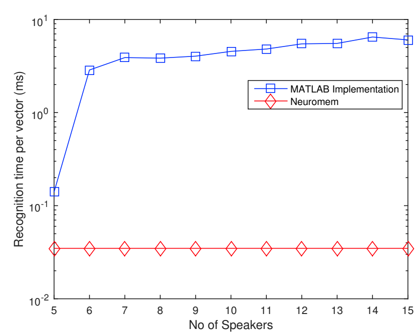


Fig. 9: Word recognition timing performance.

Similar comparison of the word recognition time (Fig. 9) was done for the hardware and a native MATLAB implementation. The ANN library was not included in this comparison, as the hardware implemented word recognition through RBF neural networks, while the ANN library only implemented an approximate nearest neighbor algorithm. Fig. 9, (recognition in RBF mode) clearly shows the superiority of the CM1K in mapping and using *complex non-linear decision spaces* efficiently.

IV. Face Recognition

We used the Yale Face Database A [17], [18] and ORL (AT&T) Database [19], [20] for all simulations. The images were divided into training and test sets randomly. Different pre-processing techniques like Discrete Wavelet Transform (DWT), Discrete Cosine Transform (DCT) and feature extraction through Gabor filters were applied along with dimensionality reduction techniques like Linear Discriminant Analysis (LDA) and Principal Component Analysis (PCA). This allowed data compression which reduced the number of neurons required for learning. Data pre-processing is done on PC and then fed to CM1K for training/classification. [21], [22], [23], [24], [25].

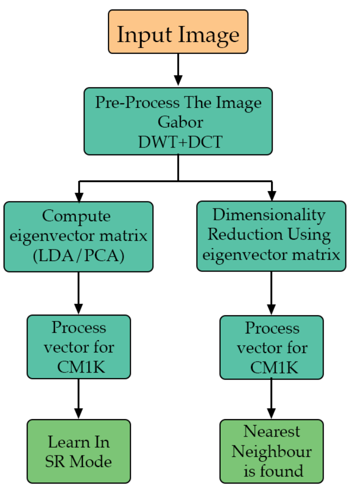


Fig. 10: Steps for Face Recognition.

A. Feature Extraction

We pre-processed the images using the following techniques :

- Gabor Filters : A set of Gabor filters is used with 5 spatial frequencies and 8 distinct orientations, this makes 40 different Gabor filters[26].
- Discrete Wavelet Transform & Discrete Cosine Transform : We have used the wavelet from Daubechies family, Db4 wavelet and after applying DWT upto 3 levels, we have applied DCT for further feature extraction[7],[27].
- Dimensionality Reduction
  - Linear Discriminant Analysis
  - Eigenfaces

B. Experiment

The entire experiment flow in shown in Fig. 10. The Yale database contains 165 grey scale images of 15 individuals, each individual has 11 images. The images demonstrate variations in lighting conditions and expressions, as shown in Fig. 11(b). The images are manually cropped to 32 x 32 pixels (closed crop), with 256 grey levels per pixel. For the simulations, we divided the dataset such that each individual has 5 training images and 6 test images, chosen randomly from the 11 images. 50 such random splits are made.

(a) ORL Database.

(b) Yale A Database.

Fig. 11: Sample individual photos.

TABLE II: Accuracy and Timing of Yale A Dataset.

Training : Test Per Individual	5 : 6
Accuracy(%)	85.63
Standard Deviation(%)	2.13
Components	14
Pre-Process Time	0.00061 ms
Recognition Time with USB Transport	7.12 ms
Recognition Time On Hardware	0.003375 ms

The ORL face database contains 400 grey scale images of 40 individuals, each individual has 10 images. The images demonstrate variations in orientation of face, as shown in Fig. 11(a). The size of each image is 112x92 which is scaled to 128x128 with 256 grey levels per pixel. We performed the simulations by separating the data into training sets and test sets through random selection. Training set sizes of 3,4,5 and 6 per individual (where rest of the images form the test set) were used. 10 such random splits are made for each case.
Results for both datasets using different techniques are shown in Tab II-IV and Fig.12-14.

An SVM classifier trained in libSVM gave comparable classification accuracy (98.7% for 6 training images using

TABLE III: Accuracy for ORL with varying training ratio and techniques.

Training	None	DWT+DCT	GABOR
3	88.7±1.9	93.3±1.2	91.7±1.7
4	92.0±1	95.2±1.1	95.3±0.9
5	94.9±1.8	97.7±0.7	97.3±0.8
6	96.2±1.6	98.1±0.5	98.1±1

DWT). The recognition times taken by the CM1K hardware is compared to ANN C++ library, knnsearch and fitcknn (both MATLAB functions) , over varying component size (Fig. 14) of each image and varying training size (Fig. 13). The software implementation were done on an Intel i7 4th Generation PC running Windows 7, 16GB RAM. Fig. 13-14 clearly indicate that the CM1K shows superior recognition timings w.r.t both training size and dimensionality. The increase in speed is approximately 5 times.

TABLE IV: Timing analysis for ORL with varying techniques.

	None	DWT + DCT	Gabor
Components	45	100	65
Pre-Process Time(ms)	0.0007	3.6025	71.1554
Recognition Time with USB Transport (ms)	7.1424	7.1145	7.1167
Recognition Time On Hardware (ms)	0.0054	0.0088	0.0066

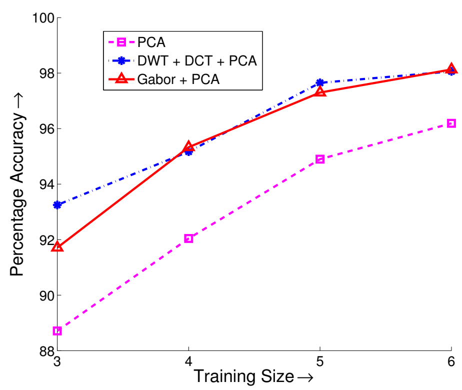


Fig. 12: Percentage accuracy.

V. Data Fusion : Face and Speech

A. Algorithm

The approach we used for data fusion is derived from Bayesian inference concepts and the ideas behind complementary filters often used in basic robotic sensory data fusion [28]. For each test case, every system determines the distances and categories of the k-nearest neighbors to construct a "confidence vector" $\bar{v}$ for every test case, which

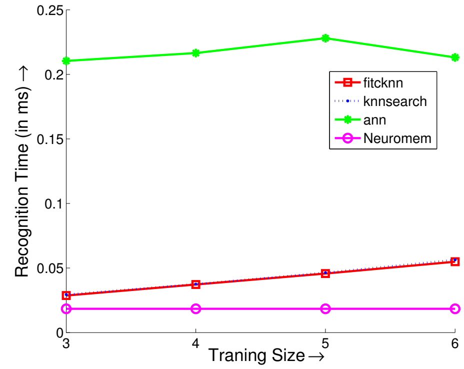


Fig. 13: Time variation with training size.

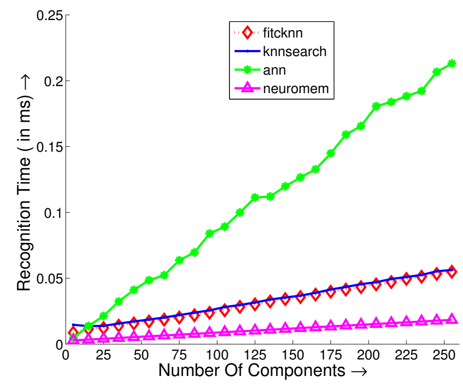


Fig. 14: Time variation with number of components.

denotes the confidence of the system in its categorisation of the test case as belonging to a certain category.

To compute confidence vector, $\bar{v}$, the validation set is used to create a confusion matrix C, which is used to calculate the conditional confidence.

$$\begin{aligned} v_{i|j} &= \text{Prob}\,(\text{person} = i \mid \text{prediction} = j) \\ &= \frac{C_{i,j}}{\sum_{i'} C_{i',j}} \end{aligned} \tag{3}$$

The confidence value for the i^{th} category ,v_i, is the product of inverse distance weight, $\frac{1}{d_k}$, of the K nearest neighbours and the conditional confidence summed over all categories.

$$v_i = \sum_{j \in \text{cat}} \alpha_j * v_{i|j} \tag{4}$$

where

$$\alpha_j = \frac{\sum_{k=1}^{K} \frac{1}{d_k} \delta_{kj}}{\sum_{k=1}^{K} \frac{1}{d_k}} \tag{5}$$

Where δ_{kj} is the Kroenecker Delta function, equal to 1 only when $k = j$, and 0 otherwise.

We then take the average of the confidence vector outputs from each system and give a final confidence vector. The category with the highest value in this vector is designated as the output of the system. In this way, the final confidence vector always draws upon the more confident system. The utility of this is that either system can change the number of nearest neighbours it pulls to optimise its own performance and still be able to output a universally compatible confidence vector.

B. Experiment

Yale A Faces dataset and VCTK audio dataset (as described in Section IV and III) are used to form a hybrid model of person identification using face and speech data. Both datasets have 15 individuals, and each individual in one dataset is assigned to an individual in the other dataset.

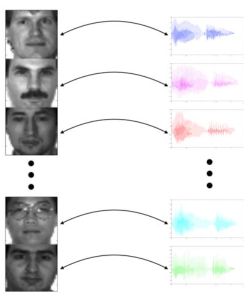


Fig. 15: Mapping of each individual from Yale A database to an individual of the VCTK database for Face-Speech Recognition.

As the Yale dataset has 11 images of each person while the VCTK has 10 sample of audio, 10 images are chosen randomly for each person for every simulation. Then, the 10 data vectors of each individual is split into 5 training, 2 validation and 3 test cases randomly for both face and speech. Each input data consists of one face image and one speech vector assigned to the same category.

The two models are fused using the algorithm described above, and 50 simulations were run for different random splits of the training, validation and test cases. For both Face and Speech, the simulations are done for the 5 nearest neighbors.
To run the simulation on the CM1K hardware, the two datasets were loaded onto the hardware under different contexts. Thus, when recognizing under the face context, only the neurons associated with it fire. Similarly we can expand the number of features to make the system more robust as shown in Fig.2.

As can be seen from the Table.V and Fig. 16, the combined accuracy shows a gain of about 6%, w.r.t individual Face and Speech accuracies.

C. Energy Considerations

The chip operates at 25 MHz with a power dissipation of 275 mW in the active state [9]. The learning time and recognition time were found to be approximately 2.43 ms and 1.77 ms respectively. Thus energy dissipated for learning and classification is 668 μJ and 487 μJ respectively.

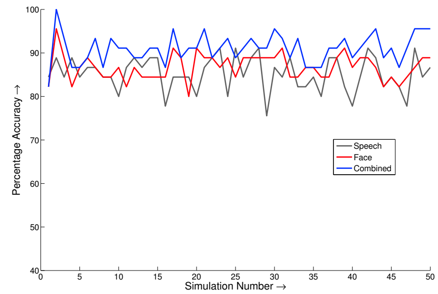


Fig. 16: Accuracy results over 50 simulations.

TABLE V: Results of 50 random splits of Face-Speech Recognition.

	Face	Speech	Combined
Accuracy	86.75	85.24	91.02
Standard Dev.	3.01	3.93	3.33
Maximum	95.55	91.11	100.00
Minimum	75.56	80.00	82.22

D. Limitations of current system and future directions

The current hardware suffers a considerable loss in actual speed of operation due to the slow USB transfer rates, which can be increased considerably or even removed if all the computation is done on and via hardware itself. Also the CM1K chip is currently fabricated at the 130 nm node with MRAM used as the non-volatile storage. Neuron density, power-efficiency and individual neuron storage capacity can be easily improved by advancing the CM1K design to more recent CMOS nodes (ex 45/28/14 nm). System capability and efficiency can be further enhanced by integrating emerging resistive memory (RRAM) technologies for the on-chip non-volatile functionality. Integration of RRAM would also open the possibility of directly expanding the learning kernels as RRAM has been widely shown to mimic synaptic emulation and a variety of learning rules in advanced mixed-signal neuromorphic hardware[2][3].

VI. Conclusion

In this paper, we demonstrated the implementation and methodology of a fast, efficient/accurate and low power authentication system using the CM1K chip. The size and scope of current application is of proof-of-concept nature, and can be improved further. Parallelism of the CM1K hardware chip allows us to have significant increase in the recognition time compared to standard software based solutions. We have demonstrated Speech Recognition, various alterations of Face Recognition and finally combined the two sets of data to realize a multimodal classifier on the CM1K chip to obtain a more robust and accurate authentication system. We compared the recognition times of the CM1K chip with that of various software libraries. The results clearly show that the recognition times are much faster even for small datasets (5 times) while still maintaining accuracy comparable to the software solution.

The scalability of the CM1K chip and the constant recognition times as opposed to the linearly increasing recognition times of computers (w.r.t increasing size of dataset) provides a huge advantage for real time computations, making them virtually independent of the dataset size. Although the number of neurons certainly need to be increased. The energy consumption values of the chip for learning and classification were 667 μJ and 448 μJ respectively. The CM1K chip, while promising, has ample scope of further performance/functionality improvement through advancing the CMOS design node and integration of new emerging NVM technologies.

VII. Acknowledgement

This research project is partially supported and sponsored by Department of Science & Technology (DST), Government of India.

References

[1] Y. Liao, "Neural networks in hardware: a survey."

[2] M. Suri and V. Parmar, "Exploiting intrinsic variability of filamentary resistive memory for extreme learning machine architectures," *Nanotechnology, IEEE Transactions on*, 2015.

[3] M. Suri, "Technologies émergentes de mémoire résistive pour les systèmes et application neuromorphique," Ph.D. dissertation, Université de Grenoble, 2013.

[4] R. Genov and G. Cauwenberghs, "Kerneltron: support vector" machine" in silicon," *Neural Networks, IEEE Transactions on*, vol. 14, no. 5, pp. 1426–1434, 2003.

[5] S. K. Kim, L. C. McAfee, P. L. McMahon, and K. Olukotun, "A highly scalable restricted boltzmann machine fpga implementation," in *Field Programmable Logic and Applications, 2009.FPL 2009.International Conference on.* IEEE, 2009, pp. 367–372.

[6] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: a literature survey," *ACM computing surveys (CSUR)*, vol. 35, no. 4, pp. 399–458, 2003.

[7] S. Sardar, G. Tewari, and K. Babu, "A hardware/software co-design model for face recognition using cognimem neural network chip," in *Image Information Processing (ICIIP), 2011 International Conference on.* IEEE, 2011, pp. 1–6.

[8] J. Y. Boulet, D. Louis, C. Godefroy, A. Steimle, P. Tannhof, and G. Paillet, "Innovative neuron circuit architectures," Jan. 29 1996, cA Patent App.CA 2,149,478. [Online]. Available: https: //www.google.com/patents/CA2149478A1?cl=en

[9] CM1K Datasheet. [Online]. Available: http: //general-vision.com/datasheet/DS_CM1K.pdf

[10] NeuroMem Ref Manual. [Online]. Available: http: //general-vision.com/documentation /TM_NeuroMem_Technology_Reference_Guide.pdf

[11] CM1K Decision Space mapping. [Online]. Available: http: //general-vision.com/documentation /TM_NeuroMem_Decision_Space_Mapping.pdf

[12] J. Yamagishi. (2010) English multi-speaker corpus for cstr voice cloning toolkit. [Online]. Available: http: //homepages.inf.ed.ac.uk/jyamagis/page3/ page58/page58.html

[13] S. Gupta, J. Jaafar, W. F. wan Ahmad, and A. Bansal, "Feature extraction using mfcc," *Signal & Image Processing*, vol. 4, no. 4, p. 101, 2013.

[14] K. Prahallad. Speech technology: a practical introduction topic: spectrogram, cepstrum and mel-frequency analysis. [Online]. Available: http: //www.speech.cs.cmu.edu/15-492/slides/03_mfcc.pdf

[15] M. Kleinsteuber. Feature extraction: mel-frequency cepstral coefficients. [Online]. Available: http: //recognize-speech.com/feature-extraction/mfcc

[16] J. Lyons. Mel frequency cepstral coefficient (mfcc) tutorial. [Online]. Available: http: //practicalcryptography.com/miscellaneous/machine-learning/ guide-mel-frequency-cepstral-coefficients-mfccs/

[17] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs.fisherfaces: recognition using class specific linear projection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, no. 7, pp. 711–720, 1997.

[18] Yale Face Database A. [Online]. Available: http: //vision.ucsd.edu/datasets/yale_face_dataset_original /yalefaces.zip

[19] F. S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," in *Applications of Computer Vision, 1994., Proceedings of the Second IEEE Workshop on.* IEEE, 1994, pp. 138–142.

[20] ORL(AT&T) Database. [Online]. Available: http: //www.cl.cam.ac.uk/Research/DTG/attarchive/pub /data/att_faces.tar.Z

[21] D. Cai, X. He, Y. Hu, J. Han, and T. Huang, "Learning a spatially smooth subspace for face recognition," in *Computer Vision and Pattern Recognition, 2007.CVPR'07.IEEE Conference on.* IEEE, 2007, pp. 1–7.

[22] D. Cai, X. He, J. Han, and H.-J. Zhang, "Orthogonal laplacianfaces for face recognition," *IEEE Transactions on Image Processing*, vol. 15, no. 11, pp. 3608–3614, 2006.

[23] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face recognition using laplacianfaces," *IEEE Trans.Pattern Anal.Mach.Intelligence*, vol. 27, no. 3, pp. 328–340, 2005.

[24] V. Štruc and N. Pavešić, "Gabor-based kernel partial-least-squares discrimination features for face recognition," *Informatica (Vilnius)*, vol. 20, no. 1, pp. 115–138, 2009.

[25] N. Pavešić and V. Štruc, "The complete gabor-fisher classifier for robust face recognition," *EURASIP Advances in Signal Processing*, vol. 2010, p. 26, 2010.

[26] Y. B. Jemaa and S. Khanfir, "Automatic local gabor features extraction for face recognition," *arXiv preprint arXiv: 0907.4984*, 2009.

[27] M. Wang, H. Jiang, and Y. Li, "Face recognition based on dwt/dct and svm," in *Computer Application and System Modeling (ICCASM), 2010 International Conference on*, vol. 3. IEEE, 2010, pp. V3–507.

[28] D. L. Hall and S. A. McMullen, *Mathematical techniques in multisensor data fusion.* Artech House, 2004.