IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 15 February 2023, accepted 5 March 2023, date of publication 8 March 2023, date of current version 23 March 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3253964

RESEARCH ARTICLE

CoRAE: Energy Compaction-Based Correlation Pattern Recognition Training Using AutoEncoder

M. DILSHAD SABIR[1], MUHAMMAD FASIH UDDIN BUTT[1], ALI HASSAN[2], SAAD REHMAN[3], MEHWISH MEHMOOD[1], AND ABDULAH JEZA ALJOHANI[4,5], (Senior Member, IEEE)

[1]Department of Electrical and Computer Engineering, COMSATS University Islamabad, Islamabad 45550, Pakistan
[2]Department of Computer and Software Engineering, College of Electrical and Mechanical Engineering (CEME), National University of Sciences and Technology, Islamabad 44000, Pakistan
[3]International Institute of Science, Arts and Technology (IISAT), Gujranwala 52250, Pakistan
[4]Department of Electrical and Computer Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia
[5]Center of Excellence in Intelligent Engineering Systems (CEIES), King Abdulaziz University, Jeddah 21589, Saudi Arabia

Corresponding author: M. Dilshad Sabir (dilshad.sabir@comsats.edu.pk)

ABSTRACT Automatic Target Recognition (ATR) using Correlation Pattern Recognition (CPR) in IoT-based applications encounters limitations like limited memory and inadequate computational resources. One reason is the required quantity of reference templates for each target/object to cover all features of a target/object. To mitigate the issue of reference templates per target/object without accuracy degradation, this paper proposes *energy compaction-based CPR autoencoder-training*. Additionally, a newly proposed performance metric known as *Peak Energy Gain (PEG)* estimates the quality of the correlation plane and the feature compression capability CPR methods. The proposed, composite filtering strategy, Eigen Maximum Average Correlation Height (EMACH), and Extended Eigen Maximum Average Correlation Height (E^2MACH) are vigorously validated using publicly available biometric and object image databases. By training a single reference template, the proposed training method achieves 97.97% mean accuracy with the second-best approach of E^2MACH that attains 53.04% mean accuracy on the Pose Estimation Database. For bio-metric fingerprint verification, the mean Equal Error Rate (EER) of the proposed approach and the composite strategy is 3% and 29.69%, respectively on the FVC2002DB1A database. Similarly, the mean EER of the proposed approach and the composite strategy is 10.55% and 26.32%, respectively on the FVC2006IA database. For FEI faces dataset, the proposed method achieves 1.41% mean EER, and the composite filtering approach achieves 21.43% mean EER. On the University of Tehran Iris database, the proposed autoencoder-based methodology obtains 19.07%, and 18.07% mean EER on the left and right side iris instances, respectively. The comparative results for each dataset demonstrate superiority of AE-based method over the state-of-the-art CPR methods.

INDEX TERMS Automatic target recognition, correlation pattern recognition, energy compaction, principle component analysis, autoencder, peak energy gain, bio-metric verification, object detection.

I. INTRODUCTION

Correlation Pattern Recognition (CPR) development came across many stages to address multiple challenges which impede its good recognition performance. Human visual perception is invariant to clutter, illumination, rotation, scale, occlusion, and noise, however, these are the real issues for artificial vision systems like, CPR. Multiple attempts have been made to resolve these challenges. The most common steps are pre-processing before training or testing CPR filters. Pre-processing steps may be log-polar transform, wavelet transform, or extraction of scale invariant features. Despite all these developments, the basic training framework of the advanced CPR filters e.g. Eigen Maximum Average

The associate editor coordinating the review of this manuscript and approving it for publication was Gianluigi Ciocca.

Correlation Height (EMACH) [1], Extended Eigen Maximum Average Correlation Heigh E^2MACH [2], and composite filtering strategy [3] approaches remain susceptible to a high number of training samples.

These advanced CPR techniques rely on the computation of covariance matrix training observations. Then the covariance matrix is decomposed using the linear transformation of Principle Component Analysis (PCA) or Singular Value Decomposition (SVD). The consequent matrices are the representation of the training observations in an alternative coordinate system. The resulting coordinates are uncorrelated with each other. However, the first coordinate has the largest possible variance and is most informative as compared to the others. Therefore, CPR techniques readily employ a scaled version of the first coordinate as a reference template or filter. It is also known as an eigenvector. The transformation has another property as it reduces multiple observations to a single reference template. The dimension reduction property could reconstruct the input observation using a single or only a few eigenvectors. Usually, a less reconstruction error implies a more accurate reconstruction of observations. Auto-encoders are neural networks with the same purpose of minimizing the reconstruction error. A single-hidden-layer autoencoder and activation function with minimum reconstruction error is related with PCA [4], [5], since weights of the autoencoder and the first few coordinates of PCA span the same PCA subspace. However, both are not identical. Autoencoder can handle greater sample complexity as compared to PCA. In this paper, we propose auto encoder-based training for CPR filters.

A. RELATED WORK

The matched filter (MF) [6], [7], [8] is a beginning point in a series of advanced correlation filters, which is ideal for recognizing the trained reference images in the presence of additive white noise. Its implementation using coherent optical computers has been described in the literature. Distortions in test images like rotations and scaling of the object, face, or fingerprint impede the performance of the MF. The lack of tolerance to distortions gave rise to the requirement of a separate MF for each appearance of an object, face, or fingerprint. Furthermore, several distinct filters are required for different poses, expressions, illuminations, occlusions, and various angles of a single subject. This unfeasible quantity of filters limits the practical deployment of MFs.

A new type of filter called synthetic discriminant function (SDF) [9] filter was introduced to address this problem. An earlier version of SDF was a weighted sum of MFs, which ensured the already specified output at the origin after correlation with training images. The value associated with trained images was ''one'' while for false targets, it was ''zero''. A peak identified the positive target detection at the origin of the output plane, if the object was centered in the test image. Despite, the attainment of a uniform peak response against training images, SDFs lacked tolerance against input noise that produced large side lobes in the output plane. To handle this challenge, a variant of SDF known as minimum variance synthetic discriminant function (MVSDF) [10], [11] has been proposed that emphasizes low frequencies. Another filter known as minimum average correlation energy (MACE) [12], [13] enhanced the peak sharpness in output images. However, it gave importance to high frequencies. An optimal trade-off [14] was possible between MVSDFs and MACE. SDFs considered hard assumptions, that were ''zero'' for false targets and ''one'' for trained images. On the contrary, an unconstrained filter [15] like Maximum Average Correlation Height Filter (MACH) can show better performance on non-trained images. MACH filter detects a false image that is relatively close to the mean of training samples as the target becasue of its dependency on the average of training samples. Eigen Maximum Average Correlation Height (EMACH) [1] and Extended Eigen Maximum Average Correlation Height (EEMACH) [2] filters resolved this problem by introducing a tunable parameter that lowers the contribution of the mean of training samples in the overall design of the filter. Distance classifier correlation filter (DCCF) [16] is another variation of the SDFs filter. This filter utilizes the distance between a prototype correlation output array and the resultant correlation array for the classification of the test image. This change of measure is more appealing to multi-class problems than using only a peak to detect the target.

Circular harmonic function (CHF) [17] filters utilized algebraic solution to attain distortion invariance like in-plane rotation. This support enables only one reference template to detect all in-plane rotation of the target. Furthermore, to extend the correlation filters to more complex and non-linear mapping of input, a statistical approach known as a polynomial correlation filter (PCF) [18] is employed. However, this approach involves computation of different powers of input values at the inference stage. Therefore, PCFs were computationally expensive and not feasible for low precision data types.

In addition to CHF, the log-polar transform [19] was proposed to accomplish distortion tolerance like in-plane and scale-invariance. By involving log-polar as a pre-processing step at the training phase of the MACH filter, in-plane rotation of object translated into a horizontal shift of peak while scaling of object resulted in a vertical shift of peak in output correlation plane. To achieve the out-of-plane invariance across the 360-degree range, pre-processed images were used to train the wavelet-modified MACH (WaveMACH) filter. However, its simulation results were limited to only three classes.

Kumar et al. [20] employed different advanced correlation approaches for biometrics verification. There were input images of the face, fingerprints, and iris with various illuminations, facial expressions, and additive white noise tested for these filters. The results were obtained for different distortion sets. Evaluations were considered for multiclass datasets.

Alam et al. [21] studied different types of filters that fall under the categories of matched filter-based correlation (MFC) and joint transform correlation (JTC). These filters have been using various measurement criterion for performance evaluation. However, the authors investigated a unified framework for eight performance parameters under the same constraints with the same SAR images. Several authors have claimed that the consequent analysis was supportive in deciding the relevance of these filters to the different practical fields. Alam and Bhuiyan [22] proposed a review of recent trends in CPR filters. The authors presented an investigation of these filters on Forward-Looking Infrared Imagery (FLIR) that experimented with the detection/tracking of single and multiple targets. Chiang et al. [23] investigated the statistical properties of SDF, MVSDF, and MACE filters for Synthetic Aperture Radar (SAR) images. Authors have proposed modified versions of these SDF-type filters. An analysis of the performance of these modified filters compared with standard filters in white and color noise is performed like degradation of performance due to misestimation of noise statistics and image normalization. Bhuiyan et al. [24] proposed a Power EMACH filter by introducing a power function on all training images before the training stage. The same was repeated with the test image at the inference stage. However, the outcomes demonstrated non-significant improvement in performance.

Akbar et al. [25] proposed the rotational invariant correlation filter for human motion detection. Akbar et al. [26] proposed an FPGA implementation of a correlation filter in LabVIEW that reduced the processing time with a loss in accuracy. Masood et al. [27], [28] suggested a Proximal Gradient (PG) filter and modified MACH for target tracking and detection. Tahsin et al. [29] claimed the fully invariant correlation filter against noise, clutter, occlusion, and in-plane rotation. The authors have pre-processed WMACE filter with the Difference of Gaussian (DoG) and logarithmic transform.

Rodriguez et al. [30] proposed the integration of Support Vector Machine (SVM) and CPR filter for enhacne localization of the target within the input scene. Fernandez and Vijaya Kumar [31] offered the partial-aliasing correlation filters to increase the performance of CPR filters.

Yang et. al. [32] proposed Neural Network Rotation Recognition Filter (NNRRF) based on the physical recognition process of a planar integrated 2F optical correlator. This filter design aims to overcome the limitation of existing filters in the case of a large rotation distortion. The results are compared with traditional optimal trade-off synthetic discriminant function (OTSDF) filters which demonstrate improvement in the recognition of optical correlators for rotating distortion targets. However, the suggested approach does not handle scale and color invariance issues. Xu et. al. [33] proposed a Neural Network based optical filter that emulates the working of a 4F optical correlator. The proposed approach achieves more than 300% average peak than OTSDF. The newly proposed technique enables high recognition ability in distorted environments. Xu et. al. [34], [35] proposed a simulation system established on the understanding of the lens-less integrated micro-optic correlator structure and 2F structure. OTSDF filter for different distortions is designed. Experimental results demonstrate that the lens-less coaxial integrated micro-optical correlator can accurately detect the target in presence of rotational and scaling distortion.

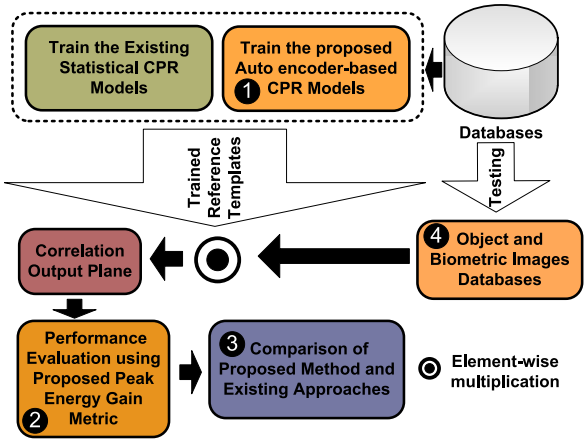


FIGURE 1. Optimization diagram of proposed Auto encoder-based *Correlation Pattern Recognition* training.

B. MOTIVATION AND RESEARCH CHALLENGES

1) HIGH NUMBER OF REFERENCE TRAINING TEMPLATES FOR A SINGLE TARGET

Typically, multiple trained templates are employed to detect a single target in CPR filters [1], [2], [3]. A reference template trains with a limited number of training samples. These samples merely cover a limited range of aspects/angles of the target. During the inference phase, many reference templates are required to incorporate all aspects of the target/object. Reducing the number of reference templates per target to minimum will greatly enhance the capacity of CPR training process.

2) COMPUTATIONAL COMPLEXITY OF INFERENCE PHASE

Multiple reference templates for a single target produces multiple responses. A high number of operations that involves cross-correlation between reference templates and the testing scene is required, i.e., the total number of operations is equal to the product of the number of operations per output for cross-correlation and the number of reference templates per target. Limiting the number of reference templates to one will reduce the total number of operations required for inference.

3) WORKLOAD REDUCTION FOR THE POST-PROCESSING OF INFERENCE

To localize the target in the output plane as a result of inference, post-processing of the output plane incurs extra

TABLE 1. Detailed description of variables used in the paper.

Variables	Metrics	Comments
E_x	Energy of sample x	E_x represents the energy of input before the Transfrom
E_y	Energy of sample y	E_y represents the energy of output after the Transfrom
σ_x^2	variance of sample x	Variance input of x vector before the Transform
σ_y^2	variance of sample y	Variance input of y vector after the Transform
A_k	Energy co-efficient for weight channel k	Ratio between variance of ratio of output variance and input variance of the Transform
D	Dataset matrix	Zero centered dataset containing samples in the column vectors
β	Beta	Parameter controls the contribution of mean within filter design
e	Required vector	Vector that minimizes the reconstruction error
Σ	Co-variance matrix	Covariance matrix of dataset D
N	Numebr of Samples	Total quantity of samples in dataset
C_1 & C_2	Constants	Constants value in equation
e_ρ^e	Weight vector	Weight vector for encoder of autoencoder
e_ρ^d	Weight vector	Weight vector for decoder of autoencoder
$\Delta\%$	Percentage difference	Percentage difference of reconstruction error between AE training method and other approaches
CP	Correlation plane	Matrix having correlation output plane
NCP	Normalized correlation plane	Matrix having normalized form of output correlation plane
M_{CP}	Mean of correlation plane	Average value of correlation output plane
ι	Peak intensity of the normalized correlation plane	Maximum value of in the normalized correlation plane
PEG	Peak energy gain	Performance parameter
σ_ϑ	Variance of the output ϑ	Variance of output plane ϑ
COPI	Correlation output peak intensity	Maximum value of the output correlation plane
Δacc	Accuracy difference	Difference of accuracy between PEG and COPI
C	Constant	Parameter controllers the energy gain bias
ACC_{PEG}	Accuracy due to PEG	Accuracy achieved after employment of PEG metric
ACC_{COPI}	Accuracy due to COPI	Accuracy achieved after employment of COPI metric
EER	Equal error rate	Rate where the false acceptance rate and false rejection rate are equal
FAR	False acceptance rate	Rate of invalid users matched to a valid user's biometric
FRR	False rejection rate	Rate of incorrectly rejected valid users

workload for the output of each reference template, e.g., normalizing the output plane before the target localization. The post-processing workload can be reduced to a minimum by limiting inference to a single reference template.

4) ANALYSIS OF PEAK CORRELATION ENERGY AND CLASSIFICATION ACCURACY OF EXISTING CPR ALGORITHMS

CPR approaches utilize training methods that have a limited number of training samples. Figure 2 analyses the mean classification accuracy *(Figure 2.(a))* and mean peak correlation energy *(Figure 2.(b))* metrices of these algorithms against ascending number of training samples per reference template. All the algorithms exhibit decaying average peak correlation energy and accuracy responses by increasing the number of samples for training. A reference template with four training samples has 100%, 85%, and 70% mean accuracy for EMACH [1], E^2MACH [2], and composite filtering strategy [3] filters, respectively. For the same sequence, the average PCE responses are 33, 22, and 10. On the contrary, by increasing the number of samples per reference template to 72, the mean accuracy response falls to almost 65%, 52%, and 50% for E^2MACH, composite filtering strategy, and EMACH filters, respectively. All the corresponding mean PCE responses drop to below 20. The analysis identifies the degradation in performance metrics for the high number of training samples. The proposed strategy demonstrates superior performance compared to existing approaches in handling the loss of performance. **This paper mainly focuses on enhancing the capacity of reference template.**

5) ASSOCIATED RESEARCH CHALLENGES

To achieve the above-specified goals, important scientific challenges which are associated with the existing CPR strategies need to be highlighted. The following challenges exist for the CPR algorithms:

1) *Few numbers of training instances per reference template:* All CPR statistical methods are limited to the small number of samples for training. These samples do not differ much from each other. Typically, each instance carries a slight variation from its neighbor samples while having a few-degree rotation (in-plane or out-of-plane). Increasing the number of instances per reference template with more variation of samples might be a challenge to solve.
2) *Loss of information in feature space of dataset:* Previous CPR approaches loose most of the feature information of training instances due to their dependency on the mean of samples or over-utilization of principle component analysis. Using the leading principle component in CPR leads to seizing the use of any feature information that is perpendicular to it.
3) *Hardware implementation issues and challenges:* Implementation of the state-of-art statistical methods of CPR faces shortage of hardware resources like limited memory availability, memory transmission rate [36], and capacity for parallel execution of operations. Reducing the number of reference templates eases the hardware implementation of these statistical methods of CPR.
4) *Exploration of possible CPR training strategies* to enhance the capacity of reference template for high

number of training samples without performance degradation.

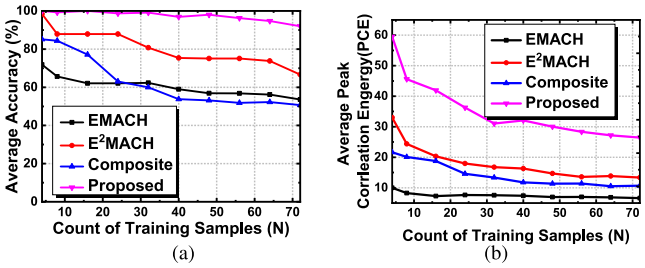


FIGURE 2. Classification accuracy and Peak Correlation Energy (PCE) comparison graphs of the existing state-of-the-art and proposed CPR training algorithms with increasing sample complexity (N).

Recent research trends in the CPR area emphasize increasing the accuracy and efficiency of these filters. Sabir et al. [37] proposed compression techniques of the power-of-two and dynamic fix point quantization schemes for spatial domain correlation filters. A modified method of training these filters has been introduced that results in low precision data type implementation of CPR. Additionally, a step enhanced the sparsity of these filters, which improved the target searching time. The proposed method handled the memory limitation and improved the efficiency of these filters. Gardezi et al. [38] claimed to achieve full invariance of spatial domain correlation filter using Affine Scale Shift Invariant Feature Transform (ASIFT). Awan et al. [39] separated the object from the background before correlation to reduce the side lobes in the outcomes. In order to achieve better performance, an auto-contour-based approach has been proposed.

C. NOVEL CONTRIBUTIONS

Subsequently, to handle the challenges mentioned earlier, following contributions have been made:

1) An autoencoder-based CPR training *(Step 1 in Figure 1)* has been proposed to enhance the recognition capability of training templates with a high sample complexity.
2) Peak Energy Gain (PEG) *(Step 2 in Figure 1)*, a performance metric based on energy compression of instances, is presented and compared with existing CPR approaches to demonstrate its superior performance.
3) Extensive performance investigation *(Step 3 in Figure 1)* of the proposed and existing CPR approaches using various object and biometric databases has been presented in this paper. The resultant analysis *(Step 4 in Figure 1)* demonstrates a high classification and identification performance advantage over the existing CPR techniques for both object and biometric applications.

This paper is organized into multiple sections. The complete methodology is comprehensively presented in Section II. Each step is explained in detail. Section III describes the mathematical framework of the energy compaction property of CPR. It establishes a metric to estimate the compression capability of new auto-encoder-based CPR training. Section IV provides the mathematical relationship between CPR objective function and reconstruction error. It also compares the reconstruction error of the proposed and existing approaches. The proposed method is validated by comprehensive experimentation of object and bio-metric databases in Section V.

II. METHODOLOGY

Despite the success of CPR in tackling the challenges of rotation, scale, occlusion, clutter, and lighting conditions, it requires multiple reference templates per object/target for training to cover each feature of the object/target. The computational complexity of inference increases due to the number and size of reference templates. The pre-processing for achieving invariance further adds to the computational complexity of the CPR process. Figure 3 exhibits step-to-step methodology of the proposed approach. The proposed method consists of two data flows. Training stream and inference stream. In training stream, instances with a black background and centered objects apply for training. These out-of-plane images cover every aspect object/target *(Step 1 Figure 3)* for a particular elevation angle. However, one training image is distinctive from another, i.e., each out-of-plane rotation of the same object/target comprises of a different instance. This variation increases as the image count increases in the training set. The consecutive out-of-plane training images incrementally appends the previous training images thus creating different training sets. These training sets contain instance's variations that support investigating the filter's response against a progressively enhanced sample variation. The training process picks one training set after another to train a reference template per set. Although, the CPR's training performs in either frequency or spatial domain. The proposed method opts for old fashion frequency-based training. The proposed approach reshapes Fast Fourier Transformed instances *(Step 2 Figure 3)* to bring convenience for employing the statistical method. These frequency samples *(Step 2.(a) in Figure 3)* are converted into vector columns before the mathematical transform. The resultant vectors for each instance concatenate to form a matrix. Subsequently, each frequency transformed instance *(Step 2.(b) Figure 3)* reshapes into a diagonal matrix. These matrices process mathematically to get an average matrix. The product of the inverse of this matrix and the transpose of the concatenated matrix *(Step 2.(a) Figure 3)* provides the resultant matrix that contains both the real and imaginary parts. However, to effectively train the auto-encoder, these parts separate into two matrices and concatenate to assemble a matrix for the next step. This matrix feeds into a single-layer auto-encoder *(Step 2 Figure 3)* to acquire Minimum Mean Square Error (MMSE) over a number of iterations. After training, the product of the weighted vector *(Step 2 Figure 3)* and the transpose of instances matrix reshapes to acquire a single reference

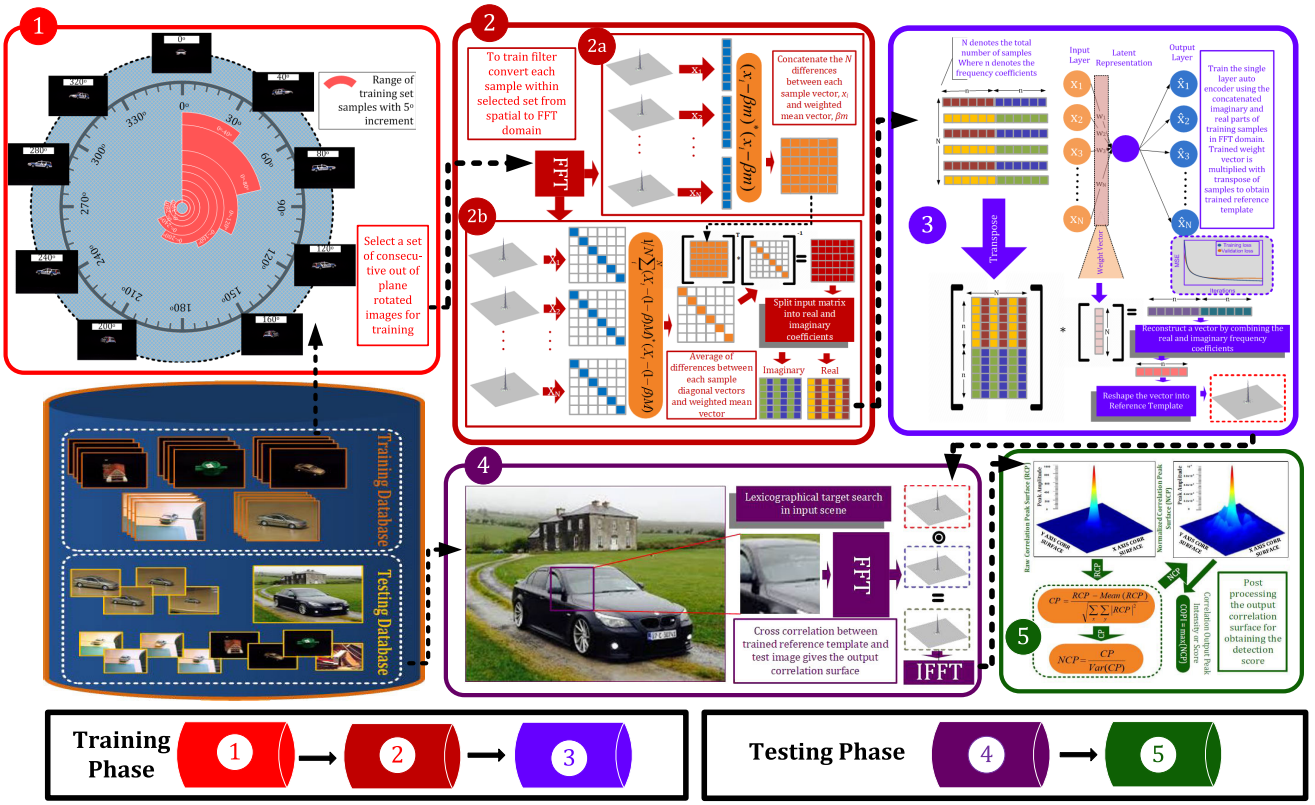


FIGURE 3. Complete block diagram of the step-by-step implementation of the proposed AE-based Correlation Pattern Recognition.

template. The training process (from step 1 to step 3) repeats itself for each set to produce a different reference template. In case of the inference stream, the approach searches the target over the input testing database. The target localization accomplishes matching the reference template over the input scene in a traditional windowing fashion *(Step 4 Figure 3)*. The method performs cross-correlation between the reference template *(Step 2 Figure 3)* and Fast Fourier Transform (FFT) of the selected window *(Step 4 Figure 3)* in the input scene. The Inverse Fast Fourier Transform (IFFT) of the correlation plane provides the correlation plane in the spatial domain. Post-processing *(Step 5 Figure 3)* of this output plane provides the detection score. Typically, this score is the peak output value of correlated surface known as correlation output peak intensity (COPI).

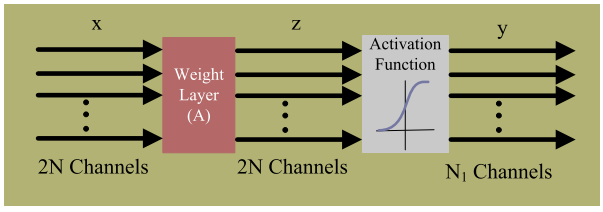


FIGURE 4. Architecture of a single neuron in auto-encoder. Note that in our case $N_1 = 1$.

III. ENERGY COMPACTION PROPERTY

This section provides a mathematical framework for the energy compaction property of the proposed CPR method based on an auto-encoder. A typical CPR training only squeezes a limited number of samples into a reference template. This limitation is due to eigen analysis, which reduces the dimensionality of data. Conventionally, the training process employs the PCA that acquires the eigenvector with the maximum eigenvalue. Eigen analysis merely learns a linear function that converts the input data into an alternative domain, with projection vectors specified by the data's variance. The projection vectors are known as eigenvectors. The vector with the maximum eigenvalue contains maximum variance. Restricting this vector as a reference template will considerably reduce the excessive dimensionality of data. However, the prime reason for considering this vector is to maximize the pre-defined objective function. To achieve good energy compaction, the transform should utilize a few coefficients to conserve the energy of the input signals. For instance, discrete cosine transform and discrete wavelet transform only exploit a few frequency coefficients to represent the number of images while causing others to have zero values. The higher number of image samples at the input, and more zero coefficients at the output, implies a better compaction property of transform. Similarly, the sound compression capability of the CPR technique is define by a high number of training instances per reference template for CPR training. We are interested in producing a reference template(s) that cover more samples in this paper.

Sub-band coding is the most widely used transform coding method that divides the input signal into constituent frequency bands. Usually, in a pre-processing step, this

transform decomposes the signal that reinforces the data compression capability. Any sub-band coding can be described mathematically using analysis and synthesis transform. Previous coding techniques [40] provide us with a mathematical framework that generalizes the bit allocation procedure of linear system energy. For any such linear sub-band coding system, the analysis transform is described in the mathematical form [41].

$$\sigma_z^2 = A\sigma_x^2 \tag{1}$$

where input vector that contains the corresponding frequency coefficient from each training sample and filter coefficients $h(n)$ decide the A_k during transmission and B_k is defined by filter coefficients $g(n)$ and quantized errors at the receiver. This paper is limited only to the transmission side A_k and single channel $k = 1$ due to the requirement of one training template for all training samples. However, unlike coding systems, auto-encoder training is a non-linear system. It is necessary to provide a solution to extend the same mathematical framework for auto-encoder. In contrast to coding, a non-linear activation function in an auto-encoder induces more complexity in the energy compaction property. For a single layer auto-encoder, an expression generalizes the energy representation for linear convolution operation as follows,

$$E_z = AE_x \tag{2}$$

where E_x and E_z denote energies of x and y due to linear convolution operation respectively. A is a vector with dimensions $N\times 1$ and N represents the number of training samples,

$$E_x = \mathbb{E}(x^2), \quad E_z = \mathbb{E}(z^2) \tag{3}$$

$$\mathbb{E}(z^2) = A\,\mathbb{E}(x^2) \tag{4}$$

Equations (1) and (4) become identical when average in equation (1) approaches to zero. The auto-encoder utilizes the Sigmoid activation function $f(z) = \frac{1}{1+e^{-z}}$, and the following equation is acquired:

$$\mathbb{E}(y^2) = \int_{-\infty}^{\infty} (\frac{1}{1+e^{-z}})^2 p(z)\,dz \tag{5}$$

We transform $f^2(z)$ using *Maclaurin series*,

$$f^2(z) = \frac{1}{4} + \frac{1}{4}z + \frac{1}{16}z^2 - \frac{1}{48}z^3 - \frac{1}{96}z^4 \ldots$$

By ignoring the higher order terms, we approximate $f^2(z)$,

$$f^2(z) \approx \frac{1}{4} + \frac{1}{4}z + \frac{1}{16}z^2$$

We can write equation (5) as,

$$\begin{aligned}\mathbb{E}(y^2) &= \int_{-\infty}^{\infty} \frac{1}{4}p(z)\,dz + \frac{1}{4}\int_{-\infty}^{\infty} zp(z)\,dz + \frac{1}{16}\int_{-\infty}^{\infty} z^2 p(z)\,dz \\ &= \mathbb{E}(\frac{1}{4}) + \frac{1}{4}\,\mathbb{E}(z) + \frac{1}{16}\,\mathbb{E}(z^2)\end{aligned}$$

$$\mathbb{E}(\frac{1}{4}) = \frac{1}{4},\ \mu_z \equiv \mathbb{E}(z) = 0,$$

We simplify the solution,

$$\mathbb{E}(y^2) = \frac{A}{16}\,\mathbb{E}(x^2) + \frac{1}{4} \tag{6}$$

Equation (6) can be generalized for M conventional layers,

$$\mathbb{E}(y^2) = \begin{cases} A\,\mathbb{E}(x^2), & \text{M=1} \\ \left(\frac{1}{16}\right)^{M-1}\left(\Pi_{m=1}^{M} A_m\right)\mathbb{E}(x^2) & \\ +\frac{1}{4}\sum_{l=2}^{M}\left(\left(\frac{1}{16}\right)^{M-l}\Pi_{m=1}^{M-l+1}A_m\right), & M > 1 \end{cases}$$

$\Pi_{m=1}^{M} A_m$ denotes the matrix multiplication that has output with dimensions $K \times 2N$ and input x has $1 \times 2N$ dimensions that is twice the number of corresponding frequency coefficients of training samples. In case of a single layered autoencoder, $K = 1$ as it has only one output channel. Further input is either the FFT of gray or three channel RGB image. The average energy of this input has the dimension equal to one. However, we have employed a single layered autoencoder that contains only one weight layer. Architecture of a single neuron is depicted in Figure 4. Therefore, A can be expressed as,

$$A = \frac{\mathbb{E}(y^2)}{\mathbb{E}(x^2)} \tag{7}$$

For large number of training frequency coefficients, average of training samples tends to be zero. As per definition of second moment $\mathbb{E}[(x-\mu)^2] = \sigma_x^2$, the equation (7) becomes,

$$A = \frac{\sigma_y^2}{\sigma_x^2} \tag{8}$$

where σ_x^2 denotes the variance of input that can be calculated using standard variance equation. Same is true of output variance. As per Parseval's theorem, energy is conserved in time and frequency domain, $\int_{-\infty}^{\infty}|E(f)|^2\,df = \int_{-\infty}^{\infty}|e(t)|^2\,dt$, thus equation (9) is valid for both domains. Figure 5 demonstrates the relationship between input and output variance.

IV. RECONSTRUCTION ERROR AND CORRELATION PATTERN RECOGNITION METHODS

Previous CPR techniques readily utilized PCA transform and mean vector approach to acquire a reference template for cross-correlation with inference images. Peak height in the correlation output plane is maximized by defining the criteria or objective function. However, the EMACH and EEMACH training processes create covariance metrics from defined criteria. In order to obtain a reference template, dimension reduction of these covariance matrices is a must. PCA provides the mechanism to convert these criteria matrices into eigenvalues λ and eigenvectors. However, among these eigenvectors, only the column vector corresponding to the maximum eigenvalue provides the reference template. This column vector maximizes the objective function as compared to other vectors having smaller eigenvalues. Additionally, this column vector also contains the highest variance of projection

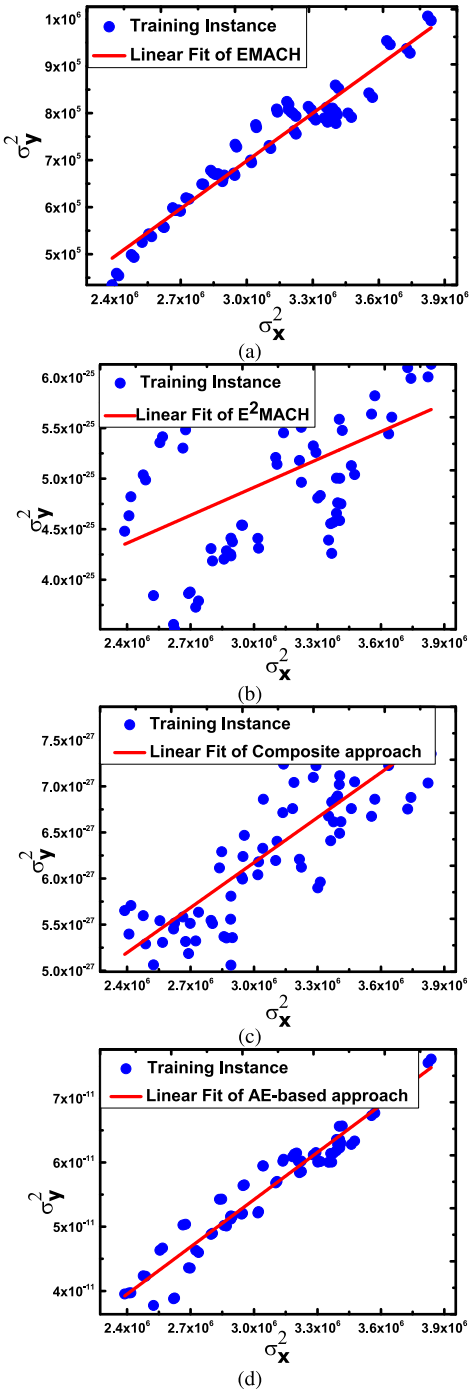


FIGURE 5. Transform analysis of (a) EMACH (b) E^2MACH (c) composite filtering strategy (d) AE-based approach.

along the data compared to the other vectors i.e., the vector contains the highest information about all training samples. However, this vector does not ensure the maximum discriminatory performance as it depends on the value of β.

Among many properties of PCA, one is dimension reduction, where for all possible linear transforms, it may perform the reconstruction of samples using a few eigenvectors with maximum eigenvalues. The approach can be extended to a single leading eigenvector with maximum eigenvalue instead of a few eigenvectors. Typically, a reference template of EMACH and E^2MACH utilizes the first principal component vector of PCA transform. However, this first principle component determines the vector in the feature domain along which projections have highest variance.

Let D be the matrix of centered data having N samples as rows and e is the unit vector that define an axis in feature space. Where covariance is defined as $\Sigma = \frac{1}{N-1}D^TD$. The main objective is to find e that results in maximum variance of projection De.

$$var(De) = \frac{1}{(N-1)}e^TD^TDe = e^T\Sigma e \tag{9}$$

The Euclidean or Frobenius norm F determines the reconstruction error between original points and the projected or estimated points.

$$\min_e \|D - ee^TD\|_F^2 \tag{10}$$

$$\|D - ee^TD\|^2 = tr((D - ee^TD)(D - ee^TD)^T) \tag{11}$$

Simplifying the equation(11),

$$\|D - ee^TD\|^2 = C_1 - C_2e^T\Sigma e \tag{12}$$

$$= C_1 - C_2var(De) \tag{13}$$

Because,

$$C_1 = tr(DD^T) + tr(Dee^Tee^TD)$$
$$C_2 = 2(N-1)$$

where C_1 and C_2 denote the constants. Only the first principle component vector provides the direction in feature space where the projections of this vector e deliver the highest variance. Therefore, only the first principle component in equation(12) ensures maximum variance and minimum reconstruction error.

EMACH filter provides a generalized mathematical framework for maximizing the objective function J_x^β in equation 16. Other methods like EEMACH and composite filtering strategy are the improved extensions of this framework.

$$C_x^\beta = \frac{1}{N}\sum_{i=1}^{N}(x_i - \beta m)(x_i - \beta m)^+ \tag{14}$$

$$S_x^\beta = \frac{1}{N}\sum_{i=1}^{N}[X_i - (1-\beta)M][X_i - (1-\beta)M]^+ \tag{15}$$

$$J_x^\beta = \frac{h^+C_x^\beta h}{h^+(1+S_x^\beta)h} \tag{16}$$

$$(1+S_x^\beta)^{-1}C_x^\beta h = \lambda h \tag{17}$$

where x_i and X_i denote the samples in spatial and frequency domains while m_i and M_i are sample averages in spatial and frequency domains. Small alphabets represent sample in vector form, and the capital alphabet shows the sample along a diagonal matrix. β controls the contribution of the

sample's mean in filter design. N denote the total number of samples.h represents the filter. Eigen value λ is the maximum value in eigenvalue matrix and its corresponding Eigen vector $(1+S_x^\beta)^{-1}C_x^\beta$ demonstrates the basic design of the EMACH filter.

AE-based CPR training provide the $(1+S_x^\beta)^{-1}C_x^\beta$ matrix to auto encoder instead of performing its Eigenvalue decomposition. In order to maximize the objective function J_x^β, the reconstruction error should be minimum. So, both C_x^β and S_x^β are modified in (18) and equation (19), respectively. Mathematical proof of these equations are given in appendix.

$$C_x^\beta = \frac{1}{N}\sum_{i=1}^{N}(x_i - \beta\hat{m} - \frac{N\sqrt{MSE - Var[\hat{x}_i]}}{(2-\beta)}) \times (x_i - \beta\hat{m} - \frac{N\sqrt{MSE - Var[\hat{x}_i]}}{(2-\beta)})^{+} \tag{18}$$

$$S_x^\beta = \frac{1}{N}\sum_{i=1}^{N}(X_i - (1-\beta)\hat{M} - \frac{N(1-\beta)\sqrt{MSE - Var[\hat{x}_i]}}{\beta(2-\beta)}) \times (X_i - (1-\beta)\hat{M} - \frac{N(1-\beta)\sqrt{MSE - Var[\hat{x}_i]}}{\beta(2-\beta)})^{+} \tag{19}$$

To investigate the effectiveness of AE-based CPR training, the Euclidean norm of the proposed technique should be less than previous approaches. Therefore, the following percentage difference is present for comparison between the proposed method and its counterparts.

$$\Delta\% = \frac{(\rho - \tau_i)}{\rho} \times 100 \tag{20}$$

where,

$$\rho = \|D - D_o\|^2 \tag{21}$$

$$D_o = \left|e_\rho^\varepsilon e_\rho^d\right| D \tag{22}$$

$$\tau_i = \|D - e_i e_i^T D\|^2 \tag{23}$$

$\Delta\%$ denotes the percentage difference of reconstruction error of the AE training method and the previous approaches. Equation (21) provides the Frobenius norm by utilizing the proposed method. e_ρ^ε and e_ρ^d denote the weights of encoder and decoder of AE in equation (22). Where equation (23) provides three options $i = EMACH, E^2MACH, compositefilteringstrategy$ to have a fair comparison of reconstruction error of previous techniques with the proposed approach ρ.

Figure 6 (a-c) is demonstrating less Euclidean norm value as compared to EMACH, E^2MACH, and the composite approach that identifies the distinction of the AE-based approach. For EMACH, Figure 6 (a) has the greatest difference between norm values that increases with a growth in the number of training samples. Conversely, in Figure 6 (b-c), this difference is decreased along with a rise in the count of training samples.

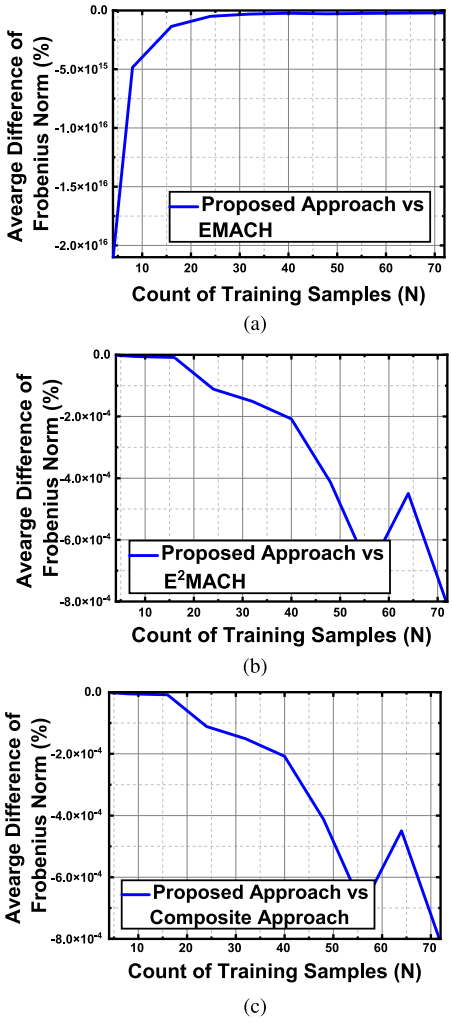


FIGURE 6. Frobenius norm comparison graphs of the proposed approach with the existing state-of-the-art CPR algorithms with growing sample complexity (a) EMACH (b) E^2MACH (c) composite filtering strategy. Each case presents a percentage difference below zero, which demonstrates that the proposed approach achieves less average Euclidean norm value during the reconstruction of input.

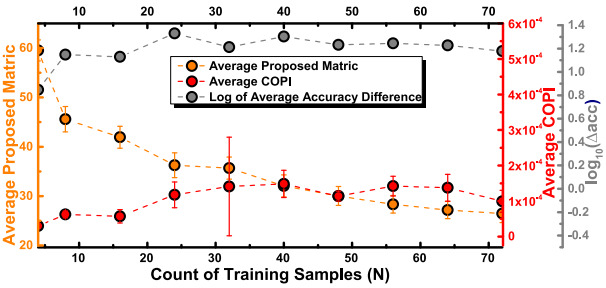


FIGURE 7. Comparison of performance metrics after proposed AE-based CPR training i.e. proposed average PEG and conventional metric, average Correlation Output Peak Intensity (COPI), and consequent accuracy difference with increasing sample complexity (N).

A. PEAK ENERGY GAIN (PEG)

Interpreting the output correlation plane for target detection requires defining some performance metric or score. This

performance measure describes the quality of the CPR algorithm, as it conveys the separation between the target from the background. Finding a metric that maximizes the difference between target and background responses is the key to good performance. Previous approaches [1], [2], [3] have widely employed the correlation output peak intensity (COPI) to detect and localize the target in the test scene. A threshold value is calculated by the weighted mean of COPI values of training samples. Any COPI response above the threshold is considered a target. Otherwise, it is background. Side lobes in the output correlation plane are always the issue as they could mislead the detection process. However, COPI does not consider the energy compaction property due to the filtering of the CPR algorithm. Here, we have introduced a new metric for target detection that assumes the energy compaction of the CPR filter. Figure 7 is the accuracy difference achieved after employing the newly proposed performance measure, peak energy gain (PEG). The performance is considerably enhanced over COPI due to the proposed metric.

$$\vartheta = \frac{CP - M_{CP}}{\sqrt{\sum_x \sum_y |CP|^2}} \tag{24}$$

$$NCP = \frac{\vartheta}{\sigma_\vartheta} \tag{25}$$

$$\iota = max(NCP) \tag{26}$$

$$PEG = -\frac{\iota}{\log_{10}(A) + C} \tag{27}$$

whereas CP denotes the correlation output plane and M_{CP} is the average of this plane. σ_ϑ is variance of the output ϑ. NCP is the normalized correlation plane. ι denotes the peak intensity of the normalized correlation plane. PEG denotes the peak energy gain. C illustrates the constant term it controllers the energy gain bias where $C \geq 5$ in our case. While equation (9) describes the A.

Figure 7 demonstrates the comparison of COPI metric with the proposed PEG measure. AE-based correlation filter has been trained for an increasing number of training samples, and their corresponding average values of COPI (red line) and PEG (saffron line) are compared along with their standard deviation. Similarly, Δacc (gray line) defines the difference in accuracy due to PEG and COPI.

$$\Delta acc = ACC_{PEG} - ACC_{COPI} \tag{28}$$

where ACC_{PEG} and ACC_{COPI} denote the accuracy because of PEG and COPI metrics. Note that the experiment setting for both readings is the same. The gray line in Figure 7 is above the value of one for almost all counts of training instances that demonstrate the better performance of the proposed metric.

V. EXPERIMENTAL RESULTS AND DISCUSSION

Performance evaluation of the target detection model is a key step in the model development process. It provides us with a good idea to establish the suitability of the detection model for a particular application. Previous approaches [1], [2], [3] are limited to assess the performance using simple correlation output peak intensity (COPI) score or through the graphs of ROC curves of a few classes. In biometric systems, equal error rate (EER) is vital along with accuracy, False Acceptance Rate (FAR), and False Rejection Rate (FRR). Where FAR is defined as the rate of identification of false or invalid users and FRR is defined as the system failure in identification of the valid user. However, for object detection, accuracy is the most vital metric for performance evaluation. Here, we have performed a more rigorous experimental evaluation of proposed and existing schemes of CPR by employing a large number of classes with a variety of sources e.g, object, fingerprint, face, and iris.

We have investigated the proposed and existing approaches using various publicly available databases. These datasets are included for biometric identification and object recognition. Evaluation of biometric identification involves fingerprint, face, and, iris classification while object recognition introduces the target recognition with a variety of object complexity, scalability and, poses. Table 2 demonstrates the details of these image databases. In order to grasp the maximum knowledge about the target, each class is trained on alternative samples from the dataset and tested on the rest of the samples from the class. The training set is 60% and the testing set is 40% of samples from each class.

Experimental databases contain multiple classes or individuals or categories, which alter the performance evaluation into a multi-classification problem. Traditionally, ROC analysis is sufficient to report CPR performance. However, ROC analysis is usually performed for binary classification problems. For biometric identification, performance evaluation based on EER and ROC analysis is widely employed for binary classification problems due to its analysis simplicity and ease of vision. However, the same is not true for the multi-classification problems where the utilization of a single threshold is not possible for each class. For n multi-class problem, employing a single threshold result in $n \times n$ confusion matrix, which contains n correct entries along diagonal and $(n^2 - n)$ possible errors excluding diagonal entries. In order to evaluate the multi-class problem, each class must hold its decision threshold, which increases the complexity of the decision thresholding process more than the binary classification problem. We have a multi-classification problem, where n classes are considered as independent classes. Each class is regarded as a binary classification problem, where the true instances are predicted positives and the rest are negative instances. Subsequently, n classes generate n different ROC graphs where each class corresponds to a ROC graph. Hence, we have n EER and accuracy values for n different thresholds.

Amsterdam Library of Object Images (ALOI) [42] is an object image database of 1000 classes. The details of datasets are shown in Table 2. The databases contain different datasets, where each dataset corresponds to 72 different viewpoints for each object, a wide variety of object images under illumination conditions, color illumination under temperature

TABLE 2. Details of databases used in experimental evaluation.

Source	Databases	Individuals/categories	Total Samples	Samples Per Individual	Sensor Type
Object	Amsterdam Library of Object Images (ALOI) -Object Viewpoint	1000	72000	72	Camera
Object	Amsterdam Library of Object Images (ALOI) -Illumination Direction	1000	24000	24	Camera
Object	Amsterdam Library of Object Images (ALOI) -Illumination Color	1000	12000	12	Camera
Object	Amsterdam Library of Object Images (ALOI) -Wide-baseline Stereo	750	2250	3	Camera
Object	Object Pose Estimation - Black Background	16	576	36	Camera
Object	Object Pose Estimation - Background	16	576	36	Camera
Finger Print	FVC 2002DB1A	100	800	8	Optical Sensor
Finger Print	FVC 2006IA	140	1680	12	Electric Field sensor
Finger Print	Level Three Synthetic Fingerprint -R1	148	1480	10	Synthetic
Finger Print	Level Three Synthetic Fingerprint -R2	148	1480	10	Synthetic
Finger Print	Level Three Synthetic Fingerprint -R3	148	1480	10	Synthetic
Finger Print	Level Three Synthetic Fingerprint -R4	148	1480	10	Synthetic
Finger Print	Level Three Synthetic Fingerprint -R5	148	1480	10	Synthetic
Faces	PICS-Pain expression subset	12	84	7	Camera
Faces	PICS-Stirling faces	35	312	9	Camera
Faces	FEI faces	200	2800	14	Camera
Iris	MMU-Iris-V1	46	460 (Left=230, Right=230)	10(Left=5, Right=5)	Camera
Iris	UTIRIS-V1	79	793 (Left=400, Right=393)	Left=4∼10, Right=4∼10	multi spectral

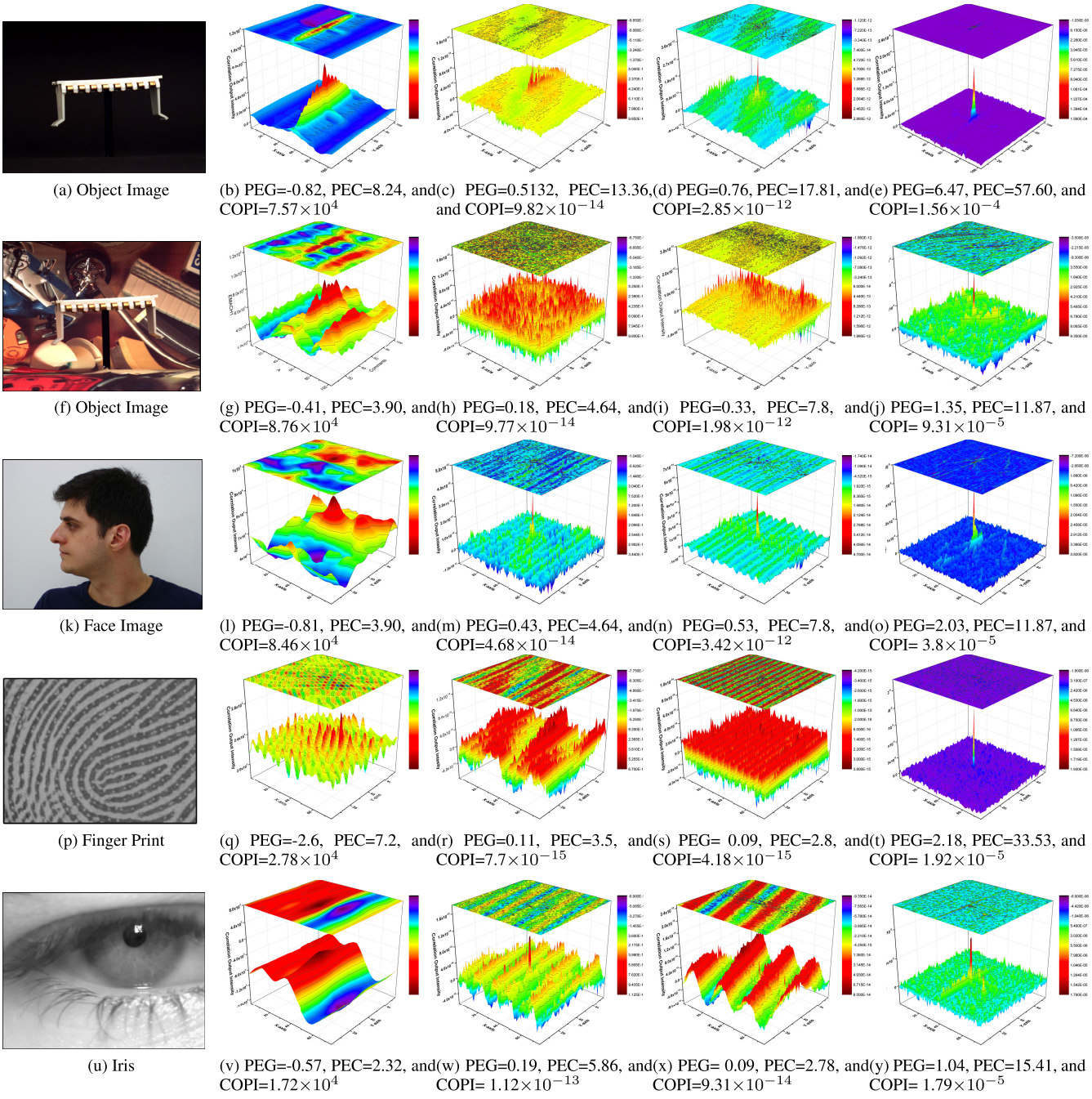
TABLE 3. Results of performance evaluation of autoencoder-based, EMACH, EEMACH, and composite strategy-based CPR methods for object recognition for one trained reference template per object.

Source	Databases	Techniques	Accuracy(%)	Precision	Recall	F1	Precision	Recall	F1
				Macro Average			Weighted Average		
Object	Object Pose Estimation - Black Background	AE-based	94.97	0.95	0.95	0.95	0.95	0.95	0.95
		Composite	27.60	0.39	0.28	0.28	0.39	0.28	0.28
		EEMACH	53.04	0.71	0.53	0.52	0.71	0.53	0.52
		EMACH	6.25	3.91×10^{-3}	0.06	0.01	3.91×10^{-3}	0.06	0.01
Object	Object Pose Estimation - Background	AE-based	41.49	4.96×10^{-1}	0.41	0.40	4.96×10^{-1}	0.41	0.40
		Composite	12.67	1.11×10^{-1}	0.13	0.11	1.11×10^{-1}	0.13	0.11
		EEMACH	9.29	2.51×10^{-1}	0.09	0.11	2.51×10^{-1}	0.09	0.11
		EMACH	6.25	3.91×10^{-3}	0.06	0.01	3.91×10^{-3}	0.06	0.01
Object	Amsterdam Library of Object Images (ALOI) -Object Viewpoint	AE-based	38.00	0.82	0.38	0.43	0.82	0.38	0.43
		Composite	8.49	0.45	0.08	0.12	0.45	0.08	0.12
		EEMACH	1.64×10^{-1}	3.66×10^{-6}	1.64×10^{-3}	7.31×10^{-6}	3.66×10^{-6}	1.64×10^{-3}	7.31×10^{-6}
		EMACH	1.63×10^{-1}	1.80×10^{-2}	1.63×10^{-3}	1.16×10^{-3}	1.80×10^{-2}	1.63×10^{-3}	1.16×10^{-3}
Object	Amsterdam Library of Object Images (ALOI) -Illumination Direction	AE-based	24.68	0.46	0.25	0.27	28.94	15.43	16.75
		Composite	3.05	0.09	0.03	0.04	5.59	1.90	2.27
		EEMACH	7.15	0.17	0.07	0.08	10.79	4.47	5.05
		EMACH	2.50×10^{-2}	8.31×10^{-7}	2.50×10^{-4}	1.66×10^{-6}	5.20×10^{-5}	1.56×10^{-2}	1.04×10^{-4}
Object	Amsterdam Library of Object Images (ALOI) -Illumination Color	AE-based	47.61	0.47	0.48	0.45	0.47	0.48	0.45
		Composite	10.38	0.09	0.10	0.09	0.09	0.10	0.09
		EEMACH	18.69	0.18	0.19	0.17	0.18	0.19	0.17
		EMACH	1.75×10^{-3}	7.63×10^{-9}	1.50×10^{-5}	2.55×10^{-8}	4.21×10^{-6}	1.13×10^{-4}	1.02×10^{-8}
Object	Amsterdam Library of Object Images (ALOI) -Wide-baseline Stereo	AE-based	36.62	0.36	0.27	0.29	0.48	0.37	0.38
		Composite	5.33	0.07	0.04	0.04	0.09	0.05	0.06
		EEMACH	11.02	0.14	0.08	0.09	0.18	0.11	0.12
		EMACH	3.40×10^{-3}	6.41×10^{-8}	1.25×10^{-5}	1.88×10^{-7}	5.20×10^{-8}	1.56×10^{-5}	1.04×10^{-8}

conditions, and wide-baseline stereo. The first dataset has 72 out-of-plan object images with 5° intervals covering each aspect of the object. The second dataset is generated using different camera positions and five lights by switching on

TABLE 4. Results of performance evaluation of autoencoder-based, EMACH, EEMACH, and composite strategy-based CPR methods for bio-metric identification one trained reference template per individual.

Source	Databases	Techniques	Avg. Accuracy(%)	Avg. EER(%)	max EER(%)	Avg. FAR(%)	Avg. FRR(%)
Finger Print	FVC 2002DB1A	AE-based	96.92	3.00	25.00	12.63	2.98
		Composite	70.37	29.69	53.41	31.75	29.61
		EEMACH	60.41	39.64	75.00	42.25	39.57
		EMACH	40.12	59.89	89.27	63.63	59.84
Finger Print	FVC 2006IA	AE-based	89.45	10.55	33.33	13.39	10.53
		Composite	73.70	26.32	50.00	27.86	26.29
		EEMACH	48.74	51.27	75.00	53.33	51.25
		EMACH	20.42	79.58	100.00	81.79	79.57
Finger Print	Level Three Synthetic Fingerprint -R1	AE-based	99.37	0.57	10.00	10.00	0.57
		Composite	71.62	28.42	60.00	30.00	28.37
		EEMACH	54.89	45.14	70.00	47.23	45.10
		EMACH	13.22	86.81	100.00	88.31	86.77
Finger Print	Level Three Synthetic Fingerprint -R2	AE-based	99.53	0.41	10.00	10.00	0.41
		Composite	73.61	26.42	60.00	28.24	26.38
		EEMACH	56.14	43.88	70.00	46.49	43.84
		EMACH	16.26	83.76	100.00	85.47	83.73
Finger Print	Level Three Synthetic Fingerprint -R3	AE-based	99.37	0.57	20.00	10.07	0.57
		Composite	74.33	25.70	50.00	27.36	25.66
		EEMACH	55.04	44.98	67.69	47.30	44.94
		EMACH	15.70	84.31	100.00	86.15	84.28
Finger Print	Level Three Synthetic Fingerprint -R4	AE-based	99.75	0.18	10.00	10.00	0.18
		Composite	73.93	26.10	52.79	28.24	26.06
		EEMACH	55.03	44.98	70.95	47.57	44.95
		EMACH	13.75	86.27	100.00	87.77	86.24
Finger Print	Level Three Synthetic Fingerprint -R5	AE-based	99.44	0.50	10.00	10.00	0.50
		Composite	74.36	25.66	51.70	28.11	25.62
		EEMACH	54.18	45.83	73.13	48.45	45.80
		EMACH	14.64	85.38	100.00	86.69	85.35
Face	PICS-Pain expression subset	AE-based	96.53	2.49	28.57	15.48	2.38
		Composite	62.80	37.45	70.13	40.48	36.90
		EEMACH	80.46	19.70	42.86	23.81	19.16
		EMACH	31.45	68.94	85.71	70.24	68.40
Face	PICS-Stirling faces	AE-based	96.78	2.98	22.22	12.65	2.95
		Composite	73.26	26.83	55.56	29.01	26.68
		EEMACH	91.57	8.32	33.33	14.81	8.25
		EMACH	42.96	57.01	74.60	60.80	56.93
Face	FEI faces	AE-based	98.57	1.41	7.72	7.18	1.40
		Composite	78.58	21.43	42.86	22.75	21.41
		EEMACH	87.58	12.43	37.90	14.14	12.42
		EMACH	30.04	69.96	97.49	71.75	69.95
Iris	U-left	AE-based	80.59	19.07	71.43	51.02	19.02
		Composite	51.56	48.35	93.17	63.38	48.26
		EEMACH	60.45	39.42	87.58	58.80	39.32
		EMACH	31.35	68.53	100.00	83.39	68.47
Iris	U-right	AE-based	81.55	18.07	81.29	56.48	18.01
		Composite	51.80	48.29	96.13	63.98	48.02
		EEMACH	64.27	35.67	91.56	59.16	35.47
		EMACH	30.88	69.03	100.00	84.45	68.93
Iris	MMU-left	AE-based	53.01	46.84	97.73	67.78	46.52
		Composite	46.99	53.41	97.73	62.22	52.80
		EEMACH	51.70	48.36	98.86	63.33	47.95
		EMACH	50.47	49.22	100.00	72.22	49.02
Iris	MMU-right	AE-based	53.43	46.57	86.36	64.44	46.16
		Composite	47.83	52.42	94.32	63.33	51.92
		EEMACH	49.23	50.61	95.45	71.11	50.30
		EMACH	48.10	52.02	100.00	65.56	51.59



(a) Object Image (b) PEG=-0.82, PEC=8.24, and COPI=7.57×10^{4} (c) PEG=0.5132, PEC=13.36, and COPI=9.82×10^{-14} (d) PEG=0.76, PEC=17.81, and COPI=2.85×10^{-12} (e) PEG=6.47, PEC=57.60, and COPI=1.56×10^{-4}

(f) Object Image (g) PEG=-0.41, PEC=3.90, and COPI=8.76×10^{4} (h) PEG=0.18, PEC=4.64, and COPI=9.77×10^{-14} (i) PEG=0.33, PEC=7.8, and COPI=1.98×10^{-12} (j) PEG=1.35, PEC=11.87, and COPI= 9.31×10^{-5}

(k) Face Image (l) PEG=-0.81, PEC=3.90, and COPI=8.46×10^{4} (m) PEG=0.43, PEC=4.64, and COPI=4.68×10^{-14} (n) PEG=0.53, PEC=7.8, and COPI=3.42×10^{-12} (o) PEG=2.03, PEC=11.87, and COPI= 3.8×10^{-5}

(p) Finger Print (q) PEG=-2.6, PEC=7.2, and COPI=2.78×10^{4} (r) PEG=0.11, PEC=3.5, and COPI=7.7×10^{-15} (s) PEG= 0.09, PEC=2.8, and COPI=4.18×10^{-15} (t) PEG=2.18, PEC=33.53, and COPI= 1.92×10^{-5}

(u) Iris (v) PEG=-0.57, PEC=2.32, and COPI=1.72×10^{4} (w) PEG=0.19, PEC=5.86, and COPI= 1.12×10^{-13} (x) PEG= 0.09, PEC=2.78, and COPI=9.31×10^{-14} (y) PEG=1.04, PEC=15.41, and COPI= 1.79×10^{-5}

FIGURE 8. *Visualization of output planes and their projections corresponding to the existing, and AE-based CPR training methods. 1^{st}, 2^{nd}, 3^{rd}, 4^{rth}, and 5^{th} columns represent source image, EMACH, E^2MACH, composite filtering strategy, and proposed approach respectively. Each graph holds its performance metrics of COPI, PCE, and PEG values.*

one at a time. Under three other conditions, eight lighting conditions, 24 instances are produced for each object. For the third dataset, the frontal view of each object under the 5k temperature interval from 2175K to 3075k is recorded. The left, right, and front angle orientations of each object are utilized to construct the fourth dataset. However, only 750 classes are included in this dataset. Table 3 compares the proposed approaches with the state-of-the-art CPR schemes for object-based recognition. In the case of viewpoint testing, the experimental investigation concludes the 38% accuracy with a 0.82 precision rate, 0.38 recall, and 0.43 F1 value for the proposed autoencoder-based approach. The composite filtering strategy approach achieves the nearest accuracy of 8.49%. For the illumination condition dataset, the AE-based method achieves 24.68% accuracy with the second-best EEMACH approach that attains 7.15% accuracy. Similarly, precision, recall, and F1 values are much better than EEMACH. Table 3 reports the 47.61% and 36.62% accuracy of the AE-based

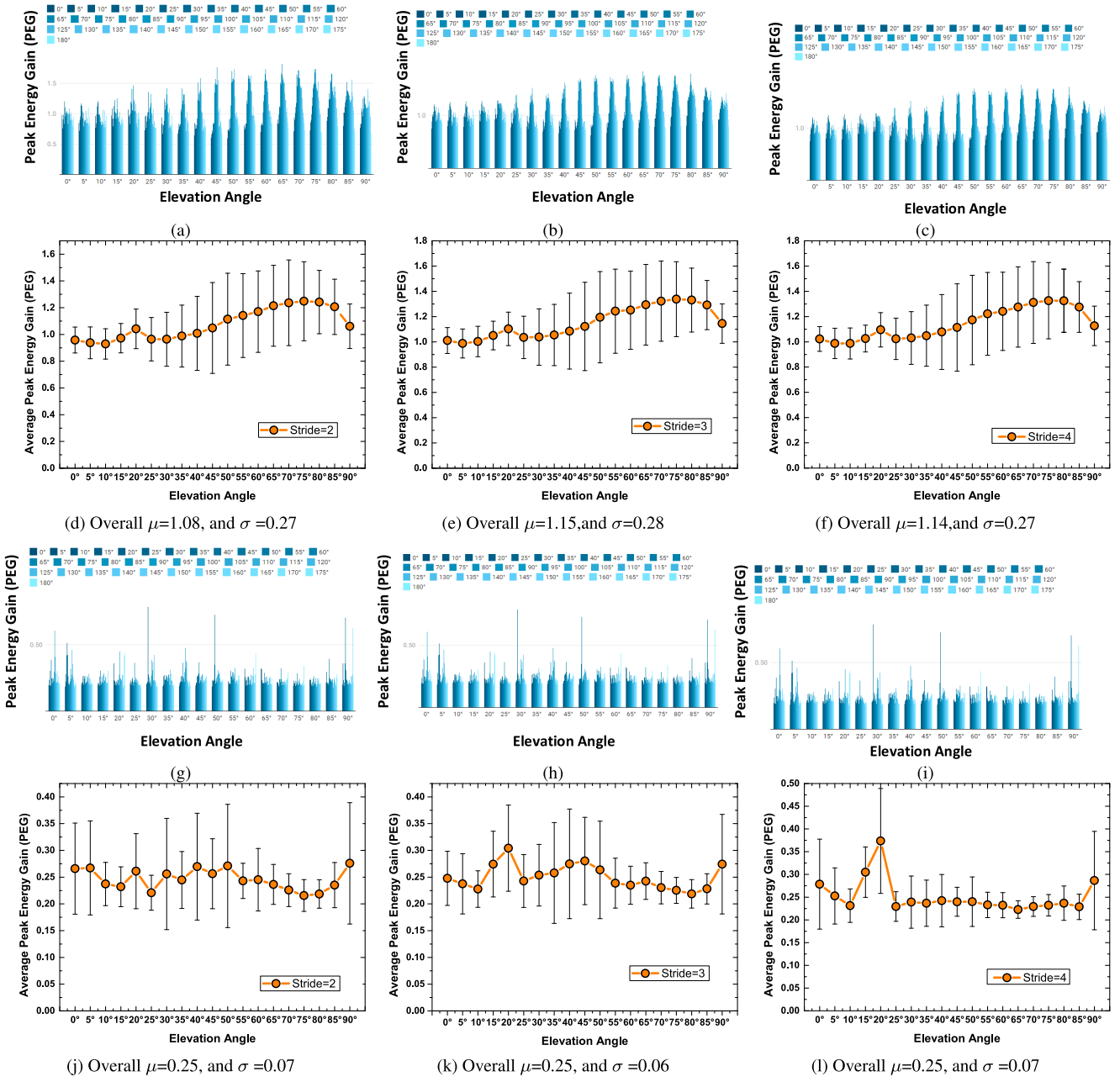


FIGURE 9. Peak energy gain responses of a single trained reference template testing using AE-based (1st row) and composite filtering strategy (3rd row) against elevation angles (0°-90°) with an interval of 5°. Each elevation angle has an output-of-plane rotation of (0°-180°) with an interval of 5. AE-based (2nd row) and composite filtering strategy (4^{rth} row) represent the mean and standard deviation of each elevation across (0°-90°) out-of-plane rotations. 1st, 2nd, and 3rd columns represent training strides of 2, 3, and 4 respectively.

approach for color illumination under different temperatures and wide-baseline stereo datasets. These accuracies, precision, recall rates, and F1 values are much better than the existing state-of-art CPR techniques. Note that the utilized object database contains multiple cases of the same object/class with different colors while CPR techniques inherently do not have color recognition ability. This may be the cause of low accuracy on the ALOI database. Object Pose Estimation [43] image database has 16 objects/classes with different elevation angles. Table 3 illustrates the performance measures of experimentation evaluations for the database. Each object/class has been recorded using 19 different elevation angles from 0° to 90° with 5° interval. 37 out-of-plane images have been recorded for each elevation angle with 5° interval. For experimentation evaluation, we have selected an elevation angle of 20° for training and testing the CPR methods for all objects. The database contains both black background and cluttered background image datasets. For black background, the

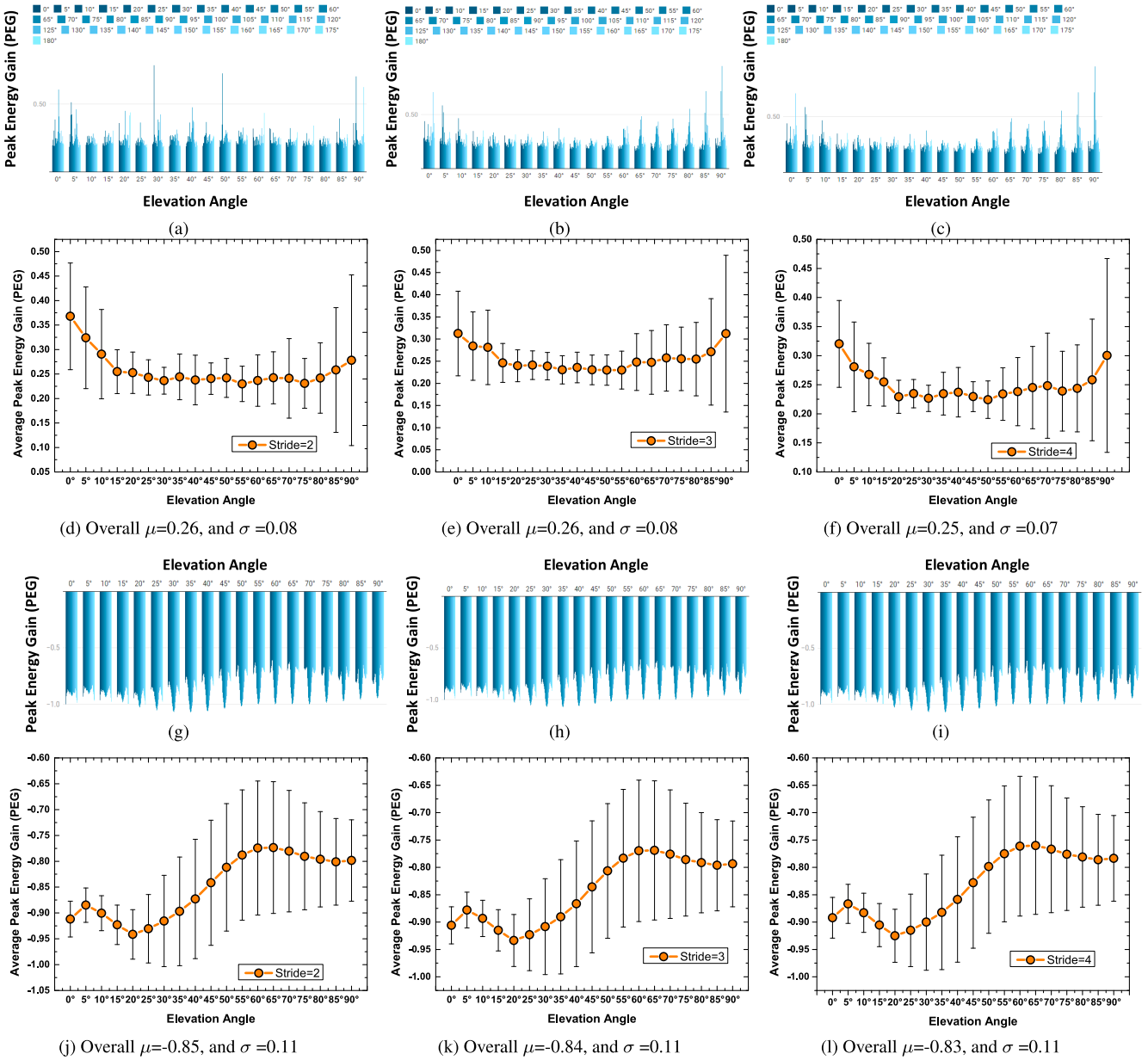


FIGURE 10. Peak energy gain responses of a single trained reference template testing using E^2MACH (1^{st} row) and EMACH (3^{rd} row) against elevation angles (0°-90°) with an interval of 5. Each elevation angle has an output-of-plane rotation of (0°-180°) with an interval of 5°. E^2MACH (2^{nd} row) and EMACH (4^{rth} row) represent the mean and standard deviation of each elevation across (0°-90°) out-of-plane rotations. 1^{st}, 2^{nd}, and 3^{rd} columns represent training strides of 2, 3, and 4 respectively.

AE-based approach attains 94.97% accuracy with 0.95 precision, recall rate, and F1 value. The most comparable approach of EEMACH achieves 53.04% accuracy with a 0.71 precision rate, 0.53 recall, and 0.52 F1 value. On the contrary, the cluttered background dataset achieves 41.49% accuracy with the nearest accuracy of 12.67% of the the composite strategy method. The reported accuracy degradation may be due to the object like clutter in the background. FVC 2002DB1A [44] and 20061A [45] are the databases for finger verification. Their details are mentioned in table 2. For the AE-based method, table 4 reports the averages of 96.92% accuracy, 3% EER, 12.63% FAR, and 2.98% FRR. The second best approach is composite strategy method which declares averages of 70.37% accuracy, 29.69% EER, 31.75% FAR, and 29.61% FRR. On the FVC 20061A database, the averages of 89.45% accuracy, 10.55% EER, 13.39% FAR, and 10.53% FRR are reported. The composite strategy method achieves the second highest averages of 73.70% accuracy, 26.32% EER, 27.86% FAR, and 26.29% FRR. Level Three Synthetic Fingerprint is the publicly available database of synthetic fingerprints. It contains five sets from R1 to R5. For each set, the average accuracies of the AE-based method stay

above 99% while EER and FRR remain equal or below 0.57%. The average FAR stays around 10%. The second best technique is the composite strategy method for all data sets. PICS-Pain expression subset [46] belongs to a database of faces of female and male individuals. Dataset maintains different facial expressions of each individual. CPR schemes are utilized to classify these individuals with a variety of expressions. The AE-based approach achieves averages of 96.53% accuracy, 2.49% EER, 15.86% FAR, and 2.38% FRR while EEMACH approach the obtains second-best averages of 80.46% accuracy, 19.70% EER, 23.81% FAR, and 19.16% FRR. On PICS-Stirling faces [46] database, the top two techniques of AE-based and EEMACH achieve the mean accuracies of 96.78% and 91.57%, respectively while the corresponding FAR are 12.65% and 14.81%, respectively. For AE-based and EEMACH methods, FRR are 2.95% and 8.25%, respectively. On FEI faces [47], the achieved accuracies of the top two approaches 98.57% and 87.58% mean accuracy, 1.41%, and 12.43% mean EER, 7.18%, and 14.14% mean FAR, and 1.40% and 12.41% mean FRR. The two best-performed methods are AE-based and EEMACH. University of Tehran Iris [48] database has instances covering both left and right-sided iris images with distribution is described in table 2. The corresponding average accuracies of left and right-sided iris images are 80.59% and 81.55% respectively, while the mean EER for the same datasets is 19.07% and 18.07% respectively as described in table 4. Likewise, the mean FAR for the left and right sides of the iris datasets are 51.08% and 56.48%, respectively, and for the same datasets, the average FRR is 19.02% and 18.01%, respectively. The second-best approach for both datasets is EEMACH. Multimedia University Iris [49] database contains the iris images of the left and right sides as described in Table 2. Left and right side samples are trained separately with a train reference template per individual. Table 4 shows the average accuracies for the left and right sides of the iris datasets are 53.01% and 53.43%, respectively, while the mean EER of the same iris datasets is 46.84% and 47.57%, respectively. The average FAR for the left and right sides of the iris datasets are 67.78% and 64.44%, respectively, and for the same datasets, FRR is 46.52% and 46.16%, respectively. For both datasets, the second-best results belong to the EEMACH approach. Figure 8 provides the visualization output plane with the COPI, PEG and, PEC of each CPR method.

A. RESILIENCE TESTS

Stringent environment conditions like noise and different illumination changes during the image acquiring process could severely degrade the target detection/prediction process. These conditions could modify the target perception and change the intensities of the target which may cause the predictor/detector to confuse the target with the background. In order to verify the invariance of the AE-based CPR training model against these harsh environmental conditions,

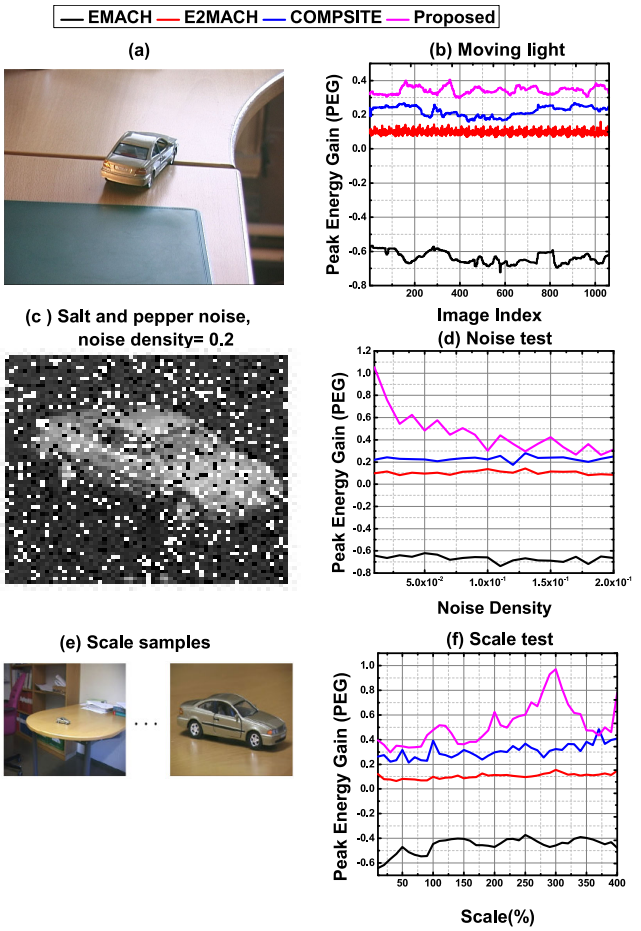


FIGURE 11. PEG comparison of existing and proposed CPR-based schemes for (a)&(b) moving lighting test, (c)&(d) noise test, and (e)&(f) scaled test.

we evaluated proposed and existing CPR statistical models on publicly available databases [43] against the artificial moving lighting conditions and scalability. Figure 11.a is illustrating the example from dataset 02 for illumination changes and Figure 11.e is demonstrating the object's different scalability with the background. In both graphs of Figure 11.b and Figure 11.f, the average PEG response of AE-based method is better than existing approaches. In addition to these variations, noise is another challenge for detection algorithms. Salt and pepper noise is induced with noise density from 0.01 to 0.2. Figure 11.c is illustrating a sample with noise density of 0.2. In Figure 11.d, despite the greater gradual fall in PEG value for the proposed approach with noise intensity, its value is still above the PEG values of existing approaches. The consequent value of PEG of these evaluations shows the resilience of the proposed approach against the variable environmental and scale instances.

B. ELEVATION ANGLES TEST

Experimental evaluation has been extended to acquire the PEG responses for different elevation angles. For each object, the database provides images with a 0°-90° degree range.

Further, for each elevation angle, 0°-180° degree out-of-plane images are recorded with 5° intervals. Therefore, 703 images are acquired for each category. We have contemplated three study cases of training in the car category. The first case considers a training set with a stride of 2, i-e, 10° difference between consecutive training samples. Similarly, the second and third cases are tested using strides 3 and 4. A single reference template is trained for each training method per case. In total, 12 reference templates are trained. Figure 9 demonstrates each study case for the AE-based proposed approach and composite filtering strategy. Figure 10 represents the same cases for EEMACH and EMACH filters. The proposed approach has the highest PEC response, while these scores remain invariant to each training stride. However, in the case of the composite filtering scheme and EEMACH, these responses have slightly variant PEG values for each training stride. While the EMACH filter has negative scores which are almost invariant across each stride. The experimental evaluation proves the resiliency of the AE-based approach as compared to the existing approaches.

VI. CONCLUSION

The designs for existing strategies for CPR statistical approaches are intrinsically constrained by a limited number of training features. Consequently, each target requires multiple reference templates for recognition, which increases the computational workload and memory prerequisites for inference. The existing state-of-the-art CPR training procedures rely heavily on PCA to mitigate the redundancy in feature space. However, PCA Transform only assesses the feature projection along the specific axis in the feature space. Thus, a new auto-encoder-based CPR training framework has been introduced in this paper with enhanced energy compression ability. The mathematical relationship has been established to depict the energy compression capability of the Transform. A new performance metric PEG, based on this relationship has been suggested, which summarizes the quality of the output plane using both COPI and energy compression capability. To perform a comparative analysis of the multi-classification ability of the existing and proposed CPR strategies, various publicly available databases of the object and biometric images are utilized. These databases maintain datasets which test the resilience against different viewpoints, scales, noises, and elevation angles. The experimental evaluation demonstrates the excellent performance of the proposed AE-based method over the existing CPR approaches. The proposed approach achieves multiclass accuracy of almost 94% for the Object Pose Estimation dataset with the black background. Similarly, attained accuracy with background drops to nearly 41%. However, in the case of the Amsterdam Library of Object Images (ALOI) which contains 1000 object categories, 38% accuracy is observed. Increasing the database size and complexity results in a significant drop in accuracy. The AE-based approach achieves 80% accuracy for most of the fingerprint, face, and Iris multi-class datasets. However, for the MMU dataset, the reported accuracy is nearly 53%. However, these datasets have less number of categories as compared to ALOI.

In order to handle more complex datasets, attention-based compression [50] or variational autoencoder may be considered for future work. Similarly, using the convolutional auto-encoders for CPR training instead of single-layer auto-encoders could be explored to enhance the energy compression capability of CPR methods. Scale-invariant Feature Transform (SIFT) can also be considered for pre-processing before CPR training for future work. Subsequently, auto-encoder-based CPR training can be used for enhancing accuracy.

APPENDIX
RELATIONSHIP BETWEEN MEAN SQUARE ERROR AND OBJECTIVE FUNCTION

$$C_x^{\beta} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \beta m)(x_i - \beta m)^{+} \tag{29}$$

$$\begin{aligned}
\hat{C}_x^{\beta} &= \frac{1}{N}\sum_{i=1}^{N}(\hat{x}_i - \beta \hat{m})(\hat{x}_i - \beta \hat{m})^{+} \\
&= \frac{1}{N}\sum_{i=1}^{N}(\hat{x}_i - \beta \hat{m})^2 \\
&= \frac{1}{N}\sum_{i=1}^{N}[\hat{x}_i^2 + \beta^2 \hat{m}^2 - 2\beta \hat{x}_i \hat{m}] \\
&= \frac{1}{N}[\sum_{i=1}^{N}\hat{x}_i^2 + \sum_{i=1}^{N}\beta^2 \hat{m}^2 - 2\beta\sum_{i=1}^{N}\hat{x}_i \hat{m}] \\
&= \frac{1}{N}[\sum_{i=1}^{N}\hat{x}_i^2 + N\beta^2(\frac{1}{N}\sum_{i=1}^{N}\hat{x}_i)^2 - \frac{2\beta}{N}(\sum_{i=1}^{N}\hat{x}_i)^2] \\
&= \frac{1}{N}[\sum_{i=1}^{N}\hat{x}_i^2 + \frac{\beta^2}{N}(\sum_{i=1}^{N}\hat{x}_i)^2 - \frac{2\beta}{N}(\sum_{i=1}^{N}\hat{x}_i)^2] \\
&= \frac{1}{N}[\sum_{i=1}^{N}\hat{x}_i^2 + \frac{\beta}{N}(\beta - 2)(\sum_{i=1}^{N}\hat{x}_i)^2] \\
&= \frac{1}{N}\sum_{i=1}^{N}\hat{x}_i^2 + \frac{\beta}{N^2}(\beta - 2)(\sum_{i=1}^{N}\hat{x}_i)^2
\end{aligned}$$

$$\frac{\partial \hat{C}_x^{\beta}}{\partial \hat{x}_i} = 0$$

$$\frac{2}{N}\sum_{i=1}^{N}\hat{x}_i + \frac{2\beta}{N^2}(\beta - 2)\sum_{i=1}^{N}\hat{x}_i = 0$$

$$\sum_{i=1}^{N}\hat{x}_i + \frac{\beta}{N}(\beta - 2)\sum_{i=1}^{N}\hat{x}_i = 0$$

$$\sum_{i=1}^{N}\hat{x}_i + \beta(\beta - 2)\hat{m} = 0$$

$$\hat{m} = \frac{\sum_{i=1}^{N}\hat{x}_i}{\beta(2 - \beta)} \tag{30}$$

Similarly,

$$m = \frac{\sum_{i=1}^{N} x_i}{\beta(2-\beta)} \tag{31}$$

Difference between eq. (30) and (31),

$$m - \hat{m} = \frac{1}{\beta(2-\beta)} \sum_{i=1}^{N} (x_i - \hat{x}_i)$$

$$m = \hat{m} + \frac{1}{\beta(2-\beta)} \sum_{i=1}^{N} (x_i - \hat{x}_i) \tag{32}$$

We know that Mean Square Error is,

$$\begin{aligned} MSE &= Var[\hat{x}_i] + Bias^2[\hat{x}_i] \\ \because Bias[\hat{x}_i] &= E(x_i - \hat{x}_i) \\ E(x_i - \hat{x}_i) &= \sqrt{MSE - Var[\hat{x}_i]} \end{aligned} \tag{33}$$

Substituting eq. (33) into eq. (32),

$$m = \hat{m} + \frac{N\sqrt{MSE - Var[\hat{x}_i]}}{\beta(2-\beta)} \tag{34}$$

Substituting eq. (34) into eq. (29),

$$\begin{aligned} C_x^{\beta} = \frac{1}{N} \sum_{i=1}^{N} &(x_i - \beta\hat{m} - \frac{N\sqrt{MSE - Var[\hat{x}_i]}}{(2-\beta)}) \\ &\times (x_i - \beta\hat{m} - \frac{N\sqrt{MSE - Var[\hat{x}_i]}}{(2-\beta)})^{+} \end{aligned} \tag{35}$$

$$\begin{aligned} S_x^{\beta} = \frac{1}{N} \sum_{i=1}^{N} &(X_i - (1-\beta)\hat{M} - \frac{N(1-\beta)\sqrt{MSE - Var[\hat{x}_i]}}{\beta(2-\beta)}) \\ &\times (X_i - (1-\beta)\hat{M} - \frac{N(1-\beta)\sqrt{MSE - Var[\hat{x}_i]}}{\beta(2-\beta)})^{+} \end{aligned} \tag{36}$$

REFERENCES

[1] M. Alkanhal, B. V. Kumar, and A. Mahalanobis, "Improving the false alarm capabilities of the maximum average correlation height correlation filter," *Proc. SPIE*, vol. 39, no. 5, pp. 1133–1141, 2000, doi: 10.1117/1.602492.

[2] B. V. Kumar and M. Alkanhal, "Eigen-extended maximum average correlation height (EEMACH) filters for automatic target recognition," *Proc. SPIE*, vol. 4379, pp. 424–431, Oct. 2001.

[3] A. B. Awan, S. Rehman, and A. D. Bakhshi, "Composite filtering strategy for improving distortion invariance in object recognition," *IET Image Process.*, vol. 12, no. 8, pp. 1499–1509, Aug. 2018. [Online]. Available: https://digital-library.theiet.org/content/journals/10.1049/iet-ipr.2017.1147

[4] E. Plaut, "From principal subspaces to principal components with linear autoencoders," 2018, *arXiv:1804.10253*.

[5] S. Ladjal, A. Newson, and C.-H. Pham, "A PCA-like autoencoder," 2019, *arXiv:1904.01277*.

[6] S. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory* (Fundamentals of Statistical Signal Processing). Upper Saddle River, NJ, USA: Prentice-Hall, 1993. [Online]. Available: https://books.google.com.pk/books?id=aFwESQAACAAJ

[7] B. V. K. V. Kumar and E. Pochapsky, "Signal-to-noise ratio considerations in modified matched spatial filters," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 3, no. 6, pp. 777–786, Jun. 1986. [Online]. Available: http://josaa.osa.org/abstract.cfm?URI=josaa-3-6-777

[8] A. V. Lugt, "Signal detection by complex spatial filtering," *IEEE Trans. Inf. Theory*, vol. IT-10, no. 2, pp. 139–145, Apr. 1964.

[9] D. Casasent, "Unified synthetic discriminant function computational formulation," *Appl. Opt.*, vol. 23, no. 10, pp. 1620–1627, May 1984. [Online]. Available: http://ao.osa.org/abstract.cfm?URI=ao-23-10-1620

[10] B. V. K. V. Kumar, "Tutorial survey of composite filter designs for optical correlators," *Appl. Opt.*, vol. 31, no. 23, pp. 4773–4801, Aug. 1992. [Online]. Available: http://ao.osa.org/abstract.cfm?URI=ao-31-23-4773

[11] S. I. Sudharsanan, A. Mahalanobis, and M. K. Sundareshan, "Unified framework for the synthesis of synthetic discriminant functions with reduced noise variance and sharp correlation structure," *Proc. SPIE*, vol. 29, no. 9, pp. 1021–1028, 1990, doi: 10.1117/12.55698.

[12] A. Mahalanobis and D. P. Casasent, "Performance evaluation of minimum average correlation energy filters," *Appl. Opt.*, vol. 30, no. 5, pp. 561–572, Feb. 1991. [Online]. Available: http://ao.osa.org/abstract.cfm?URI=ao-30-5-561

[13] A. Mahalanobis, B. V. K. V. Kumar, and D. Casasent, "Minimum average correlation energy filters," *Appl. Opt.*, vol. 26, no. 17, pp. 3633–3640, Sep. 1987. [Online]. Available: http://ao.osa.org/abstract.cfm?URI=ao-26-17-3633

[14] P. Refregier, "Optimal trade-off filters for noise robustness, sharpness of the correlation peak, and horner efficiency," *Opt. Lett.*, vol. 16, no. 11, pp. 829–831, Jun. 1991. [Online]. Available: http://opg.optica.org/ol/abstract.cfm?URI=ol-16-11-829

[15] A. Mahalanobis, B. V. K. V. Kumar, S. Song, S. R. F. Sims, and J. F. Epperson, "Unconstrained correlation filters," *Appl. Opt.*, vol. 33, no. 17, pp. 3751–3759, Jun. 1994. [Online]. Available: http://ao.osa.org/abstract.cfm?URI=ao-33-17-3751

[16] A. Mahalanobis, B. V. Kumar, and S. R. F. Sims, "Distance classifier correlation filters for distortion tolerance, discrimination, and clutter rejection," *Proc. SPIE*, vol. 2026, pp. 325–335, Nov. 1993, doi: 10.1117/12.163625.

[17] B. V. K. V. Kumar, A. Mahalanobis, and A. Takessian, "Optimal tradeoff circular harmonic function correlation filter methods providing controlled in-plane rotation response," *IEEE Trans. Image Process.*, vol. 9, no. 6, pp. 1025–1034, Jun. 2000.

[18] M. Alkanhal and B. V. K. V. Kumar, "Polynomial distance classifier correlation filter for pattern recognition," *Appl. Opt.*, vol. 42, no. 23, pp. 4688–4708, Aug. 2003. [Online]. Available: https://opg.optica.org/ao/abstract.cfm?URI=ao-42-23-4688

[19] A. Aran, N. K. Nishchal, V. K. Beri, and A. K. Gupta, "Log-polar transform-based wavelet-modified maximum average correlation height filter for distortion invariance in a hybrid digital-optical correlator," *Appl. Opt.*, vol. 46, no. 33, pp. 7970–7977, Nov. 2007. [Online]. Available: http://opg.optica.org/ao/abstract.cfm?URI=ao-46-33-7970

[20] B. V. K. V. Kumar, M. Savvides, C. Xie, K. Venkataramani, J. Thornton, and A. Mahalanobis, "Biometric verification with correlation filters," *Appl. Opt.*, vol. 43, no. 2, pp. 391–402, Jan. 2004. [Online]. Available: http://opg.optica.org/ao/abstract.cfm?URI=ao-43-2-391

[21] M. S. Alam, J. Khoury, P. P. Banerjee, W. M. Durant, D. M. Martin, and G. T. Nehmetallah, "Performance evaluation of optimal filters for target detection using SAR imagery," *Proc. SPIE*, vol. 9094, Apr. 2014, Art. no. 90940C.

[22] M. Alam and S. Bhuiyan, "Trends in correlation-based pattern recognition and tracking in forward-looking infrared imagery," *Sensors*, vol. 14, no. 8, pp. 13437–13475, Jul. 2014. [Online]. Available: https://www.mdpi.com/1424-8220/14/8/13437

[23] H.-C. Chiang, R. L. Moses, and S. C. Ahalt, "Statistical properties of linear correlators for image pattern classification with application to synthetic aperture radar (SAR) imagery," *Proc. SPIE*, vol. 2490, pp. 266–277, Mar. 1995, doi: 10.1117/12.205784.

[24] S. M. A. Bhuiyan, J. F. Khan, and M. S. Alam, "Power enhanced extended maximum average correlation height filter for target detection," in *Proc. IEEE SOUTHEASTCON*, Mar. 2014, pp. 1–4.

[25] N. Akbar, S. Tehsin, A. Bilal, S. Rubab, S. Rehman, and R. Young, "Detection of moving human using optimized correlation filters in homogeneous environments," *Proc. SPIE*, vol. 11400, pp. 73–79, May 2020, doi: 10.1117/12.2559578.

[26] N. Akbar, S. Tehsin, H. U. Rehman, S. Rehman, and R. Young, "Hardware design of correlation filters for target detection," *Proc. SPIE*, vol. 10995, pp. 71–79, May 2019, doi: 10.1117/12.2519497.

[27] H. Masood, S. Rehman, A. Khan, F. Riaz, A. Hassan, and M. Abbas, "Approximate proximal gradient-based correlation filter for target tracking in videos: A unified approach," *Arabian J. Sci. Eng.*, vol. 44, no. 11, pp. 9363–9380, Nov. 2019.

[28] H. Masood, S. Rehman, M. Khan, Q. Javed, M. Abbas, M. Alam, and R. Young, "A novel technique for recognition and tracking of moving objects based on E-MACH and proximate gradient (PG) filters," in *Proc. 20th Int. Conf. Comput. Inf. Technol. (ICCIT)*, Dec. 2017, pp. 1–6.
[29] S. Tehsin, S. Rehman, F. Riaz, O. Saeed, A. Hassan, M. Khan, and M. S. Alam, "Fully invariant wavelet enhanced minimum average correlation energy filter for object recognition in cluttered and occluded environments," *Proc. SPIE*, vol. 10203, May 2017, Art. no. 1020307, doi: 10.1117/12.2262434.
[30] A. Rodriguez, V. N. Boddeti, B. V. K. V. Kumar, and A. Mahalanobis, "Maximum margin correlation filter: A new approach for localization and classification," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 631–643, Feb. 2013.
[31] J. A. Fernandez and B. V. K. V. Kumar, "Partial-aliasing correlation filters," *IEEE Trans. Signal Process.*, vol. 63, no. 4, pp. 921–934, Feb. 2015.
[32] T. Yang, M. Chen, Y. Xiao, H. Xu, and P. Xu, "Research on 2F optical correlator based on neural network filter for recognizing large-angle rotation distortion target," *IEEE Photon. J.*, vol. 12, no. 2, pp. 1–10, Apr. 2020.
[33] P. Xu, T. Yang, K. Liang, X. Li, S. Long, and H. Huang, "A novel method to realize optical correlation recognition based on neural network," *IEEE Photon. J.*, vol. 10, no. 4, pp. 1–10, Aug. 2018.
[34] P. Xu, X. Li, T. Yang, Z. Sun, H. Huang, J. Yu, and G. Cheng, "Research on distortion invariant recognition based on lensless coaxial integrated micro-optic correlator," *J. Opt.*, vol. 20, no. 8, Aug. 2018, Art. no. 085801, doi: 10.1088/2040-8986/aacab1.
[35] P. Xu, T. Yang, H. Huang, X. Li, Y. Xiao, Y. Pan, and W. Yang, "Research on distortion invariant recognition of a planar integrated optical correlator," *IEEE Photon. J.*, vol. 10, no. 2, pp. 1–10, Apr. 2018.
[36] P. D. S. Manoj, J. Lin, S. Zhu, Y. Yin, X. Liu, X. Huang, C. Song, W. Zhang, M. Yan, Z. Yu, and H. Yu, "A scalable network-on-chip microprocessor with 2.5D integrated memory and accelerator," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 64, no. 6, pp. 1432–1443, Jun. 2017.
[37] D. Sabir, M. A. Hanif, A. Hassan, S. Rehman, and M. Shafique, "Weight quantization retraining for sparse and compressed spatial domain correlation filters," *Electronics*, vol. 10, no. 3, p. 351, Feb. 2021. [Online]. Available: https://www.mdpi.com/2079-9292/10/3/351
[38] A. Gardezi, U. Malik, S. Rehman, R. C. D. Young, P. M. Birch, and C. R. Chatwin, "Enhanced target recognition employing spatial correlation filters and affine scale invariant feature transform," *Proc. SPIE*, vol. 10995, pp. 145–160, May 2019, doi: 10.1117/12.2520555.
[39] A. B. Awan, A. D. Bakhshi, M. Abbas, and S. Rehman, "Active contour-based clutter defiance scheme for correlation filters," *Electron. Lett.*, vol. 55, no. 9, pp. 525–527, May 2019.
[40] P. N. Nuggehally and S. Jayant, *Digital Coding of Waveforms*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1984.
[41] J. Katto and Y. Yasuda, "Performance evaluation of subband coding and optimization of its filter coefficients," *J. Vis. Commun. Image Represent.*, vol. 2, no. 4, pp. 303–313, Dec. 1991. [Online]. Available: https://www.sciencedirect.com/science/article/pii/1047320391900114
[42] J.-M. Geusebroek, G. J. Burghouts, and A. W. M. Smeulders, "The Amsterdam library of object images," *Int. J. Comput. Vis.*, vol. 61, no. 1, pp. 103–112, Jan. 2005.
[43] F. Viksten, P.-E. Forssén, B. Johansson, and A. Moe, "Comparison of local image descriptors for full 6 degree-of-freedom pose estimation," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2009, pp. 2779–2786.
[44] D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, and A. K. Jain, "FVC2002: Second fingerprint verification competition," in *Proc. Int. Conf. Pattern Recognit.*, 2002, pp. 811–814.
[45] R. Cappelli, M. Ferrara, A. Franco, and D. Maltoni, "Fingerprint verification competition 2006," *Biometric Technol. Today*, vol. 15, no. 7, pp. 7–9, Jul./Aug. 2007. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0969476507701406
[46] *2D Face Sets*. Accessed: Jan. 25, 2023. [Online]. Available: http://pics.stir.ac.uk/2D_face_sets.htm
[47] *Centro Universitario da Fei, Fei Face Database*. Accessed: May 13, 2019. [Online]. Available: http://www.fei.edu.br/~cet/facedatabase.html
[48] *University of Tehran IRIS (UTIRIS) Database*. Accessed: Dec. 13, 2018. [Online]. Available: https://utiris.wordpress.com/
[49] *Multimedia University IRIS Database (MMU)—V1 and V2*. Accessed: Oct. 1, 2020. [Online]. Available: https://mmuexpert.mmu.edu.my/ccteo
[50] S. Meier, A. Erkan, N. Thielen, S. Klarmann, and J. Franke, "Attention-based image compression in sensor assembly," in *Proc. IEEE 28th Int. Symp. Design Technol. Electron. Packag. (SIITME)*, Oct. 2022, pp. 136–141.

M. DILSHAD SABIR received the degree in computer engineering from COMSATS University Islamabad (CUI), Islamabad, and the M.S. and Ph.D. degrees in computer engineering from the National University of Sciences and Technology, Islamabad, in 2022. He is currently a Faculty Member with CUI. His research interests include the Internet of Things (IoT), machine learning, deep learning, correlation pattern recognition, convolution neural networks, and network security.

MUHAMMAD FASIH UDDIN BUTT received the B.E. degree from the National University of Science and Technology (NUST), Pakistan, in 1999, the M.E. degree from the Center for Advanced Studies in Engineering, University of Engineering and Technology (UET), Taxila, Pakistan, with specialization in digital communication/computer networks, in 2003, and the Ph.D. degree in electrical and electronics engineering, with specialization in wireless communication systems from the Communications Research Group, School of Electronics and Computer Science, University of Southampton, U.K., in June 2010. He was a Postdoctoral Researcher in the area of machine learning for the massive Internet of Things with the University of Southampton, focusing on smart cities applications. He joined COMSATS University Islamabad (CUI), Pakistan, in 2002, where he is currently a tenured Associate Professor with the Department of Electrical and Computer Engineering. He is also an innovative and experienced researcher with a proven track record of prestigious peer-reviewed journals and conferences. He has published over 50 research papers in various reputed journals and conference proceedings. His research interests include radio over fiber technologies, physical-layer security, channel coding, cooperative cognitive radio networks, and efficient hardware implementation of high-throughput decoder.

ALI HASSAN received the B.E. and M.S. degrees in computer engineering from the College of Electrical and Mechanical Engineering (CEME), National University of Sciences and Technology (NUST), Islamabad, Pakistan, in 2004 and 2007, respectively, and the Ph.D. degree in electrical engineering from the University of Southampton, U.K., in 2012. He has been the Head of the Department of Computer and Software Engineering, College of Electrical and Mechanical Engineering, NUST, where he is currently a Professor. He has been actively involved in establishing linkage with international universities. He has done several research projects with local and international collaborators ranging from developing EEG-based rehabilitation system for stroke patients, gastroenterology screening-based computer-aided decision systems, and industry 4.0 process virtualization systems. He has published more than 70 research papers in international journals and conferences. His research interests include machine learning, image and signal processing in the domain of biomedical signal processing, and the IoT.

SAAD REHMAN received the Ph.D. degree from the University of Sussex, U.K., in 2009. He is currently the Vice Chancellor of the International Institute of Science, Arts and Technology, Gujranwala, Pakistan. Since then, he has been an active researcher and has published more than 100 papers in journals and conferences of international repute.

MEHWISH MEHMOOD received the B.S. degree in electrical (computer) engineering and the M.S. degree in electrical engineering from COMSATS University Islamabad (CUI), Islamabad, Pakistan, in 2015 and 2018, respectively. She joined the Department of Electrical and Computer Engineering, CUI, as a Research Associate, in 2016, where she is currently a Laboratory Engineer. Her research interests include medical image processing, deep learning, and computer vision.

ABDULAH JEZA ALJOHANI (Senior Member, IEEE) received the B.Sc. (Eng.) degree in electronics and communication engineering from King Abdulaziz University, in 2006, and the M.Sc. and Ph.D. degrees in wireless communication from the University of Southampton, Southampton, U.K., in 2010 and 2016, respectively. He is currently a Research and Innovation Consultant and an Associate Professor with the Department of Electrical and Computer Engineering, King Abdulaziz University. His research interests include channel coding, cooperative communications, free-space optical communication, and MIMO systems.