ABSTRACT
Correlation filters (CFs) are a well-known pattern classification approach used in biometrics. A CF is a spatial-frequency array that is specifically synthesized from a set of training patterns to produce a sharp correlation output peak at the location of the best match for an authentic image comparison and no such peak for an impostor image comparison. The underlying premise when using CFs is that this correlation output peak behavior on training data ideally extends to testing data. Yet in 1:1 verification scenarios, where there is limited training data available to represent pattern distortions, the correlation output from an authentic comparison can be difficult to discern from the correlation output from an impostor. In this paper we introduce Stacked Correlation Filters (SCFs), a simple and powerful approach to address this problem by training an additional set of classifiers which learn to differentiate correlation outputs from authentic and impostor match pairs. This is done by training a series of stacked modular CFs with each layer refining the output of the previous layer. Our basic premise is that since correlation outputs have an expected shape, an additional CF can be trained to recognize such shape and refine the final output. As previous works with CFs have only focused on individual filter design or application, which assumes the CF to provide a sharp peak, this is a new CF paradigm that can benefit many existing CF designs and applications.

1. INTRODUCTION
A correlation filter (CF) is a spatial-frequency array (equivalently, a template in the image domain) designed from a set of training patterns to discriminate between similar (authentic) and non-similar (impostor) match pairs. The CF design goal is to produce a correlation output displaying a sharp peak at the location of the best match from an authentic comparison and no such peak for an impostor comparison. As traditional design and usage focuses on the correlation outputs where (after the peak height and/or location are extracted) the remainder of the correlation shape is discarded. In this paper, we demonstrate a novel technique for improving the effectiveness of CFs by using the insight that the expected shape of a correlation output can be recognized. Moreover, the process of identifying an authentic correlation shape can be used to refine the correlation outputs after the initial matching for improved discrimination.

There are a large number of CFs that have been developed [1, 2, 3, 4, 5] for image matching problems and have been previously shown to perform well in biometric recognition scenarios like face [6], iris [7], periocular [8], fingerprint [9], and palm print [10]. However, the matching challenge is noticeably more difficult when only a single image is available for the gallery template, e.g., as in real-world applications (such as when matching crime-scene face images to face images in surveillance videos) and in several NIST biometric competitions [11, 12, 13] designed to mimic such real-world scenarios. CFs can implicitly and efficiently leverage shifted versions of an image as negative training samples. Therefore CFs are better suited for the 1:1 matching problem in comparison to other classifiers like Support Vector Machines (SVMs) and Random Forests which are designed to discriminate between two or more classes. However, in challenging matching scenarios (e.g., due to the presence of in-plane deformations, occlusions, etc.) an authentic correlation output may be difficult to discern from an impostor correlation output as shown in Fig. 1. The failure occurs due to a lack of training data and/or discriminative content between the probe and gallery. This problem is not new or unique to biometrics, and usual efforts to address it include varying features, changing the method of recognizing a peak (e.g., peak-to-sidelobe ratio, peak-to-correlation-energy, etc.), and filter design, e.g., the Extended Maximum Average Correlation Height (EMACH) [14, 15].

We address this problem by proposing a new architecture for 1:1 image matching referred to as Stacked Correlation Filters (SCFs). This architecture consists of a series of sequential classifiers which are trained to discriminate between
authentic and impostor correlation outputs for improved class separation. Operating first on the outputs from an initial CF (referred to as ‘layer 0’), additional sets of CFs are consecutively layered with each set designed to refine the output from the previous layer to the desired correlation output. What we present is a simple and powerful technique that can be applied iteratively to continuously improve results by simplifying the matching process to a series of sequential predictions (see Fig. 2). As previous works with CFs have only focused on individual filter design or application, this is a new paradigm in CF research (SCFs can be applied to all types of CF designs).

The use of sequential predictions (feeding the output of predictors from a previous stage to the next) has been revisited many times in the literature. In [16, 17] sequential prediction is applied to natural language processing tasks, while in [18] a face detection system was developed consisting of a cascaded series of classifiers. More recently the inference machines architecture [19, 20] was proposed that reduces structured prediction tasks, traditionally solved using probabilistic graphical models, to a sequence of simple machine learning sub-problems. Within biometrics, sequential predictions have been applied to perform score fusion [21, 22]. SCFs operate on a similar intuition (iteratively applying weak classifiers to improve the final output), however offer a novel approach in both biometric recognition as well as in CF application. To the best of our knowledge, no one has:

1. Studied the application of an additional CF (or set) to refine the initial correlation outputs.
2. Built an approach for shaping the correlation output by use of an additional classifier.
3. Used sequential predictors on an individual match score for biometric recognition. The SCF concept is not fusing the outputs from several different classifiers/features.

We demonstrate the effectiveness of SCFs through extensive experimentation on the Extended Yale B [23] facial database, achieving substantial performance gains over a single CF.

2. STACKED CORRELATION FILTERS

Correlation filters (CFs) are well explained in previous publications [1, 5, 8] and hence we provide only a brief summary.
classifiers, a nonlinear operation is implemented to separate the layers (otherwise the ‘stack’ is equivalent to learning a single filter and there would be no advantage to learning or using a stack of CFs). In our design, we considered four nonlinear operations: peak correlation energy (PCE), hyperbolic tangent (Tanh), sigmoid function (Sig), and the rectified linear unit (RLU), where the nonlinear operation is applied to the output(s) of the SCFs when combined with the previous layer’s outputs. Recall that the purpose of the SCFs at each layer is to refine the previous layer’s correlation output, to this end; we developed two refinement methods (see Fig. 3):

- **Combined Refinement (CR)** - A nonlinear function, \( f \), is applied to summed SCF outputs \( f(\sum_{j=1}^{N} C_j) \).

- **Individual Refinement (IR)** - A nonlinear function, \( f \), is applied to each SCF output \( f(C_j) \) individually.

The refinements are added to the previous layer’s outputs.

### 2.2. Implementation

As we will show, achieving optimal performance by manually encoding a single layer or set of layers to a specific refinement method and nonlinearity is a non-trivial task. Thus, during training we determine the best selection with cross-validation, a procedure we designate as **Dynamic Refinement (DYN)**. The result allows the architecture to actively adjust to the quality of the outputs of the previous layer.

---

**Fig. 4**: Performance of the Combined (CR) and Individual (IR) refinement methods with each nonlinearity. For each patch configuration (x-axis), each plot displays VR on the left y-axis (solid lines) and EER on the right y-axis (dashed lines). The tables display the best results and illustrate numerically how aggressive refinement can quickly cause performance to diverge.

**Fig. 5**: Example of how ‘hard impostors’ can negatively affect the SCF output. The shown impostor score distribution splits into multiple modes, separating impostors by quality and causing the higher layers to perform poorly. The neighboring table contains the fisher ratio (measuring the separation between the authentic and impostor distributions) at each layer.

**Fig. 7**: Sample images from the Yale face database.

We examine the various approaches to applying the SCFs using the Extended Yale B face dataset [23]. Composed of 2414 frontal-face images from 38 subjects, the images capture 9 lighting directions and 64 illumination conditions for each user (see Fig. 7). Traditionally the dataset is divided into 5 subsets, however for the presented experiments all of the images were treated equally to eliminate any bias.

In the presented tests each refinement method and nonlinearity is evaluated in a 1 : 1 image-to-image matching scenario using 5-fold cross validation. As a measure of overall system performance we report equal error rates (EERs) and verification rates (VRs) at 0.001 False Acceptance Rate (FAR) from the scores obtained by concatenating the associated folds (excluding self-comparisons). Finally, we preprocess the images by a simple histogram normalization and resize each to 128×128 pixels for computational efficiency.

Fig. 4 displays the resulting EERs (right y-axis, dashed lines) and VRs (left y-axis, solid lines) from running three

---

1 Rank-1 identification rate is not included as it is always > 99%.
SCF layers over six patch configurations (non-overlapping rectangular patches, e.g., Fig. 2 displays a $3 \times 3$ configuration). Best results for CR (94.62% VR, 1.84% EER) are obtained using the first layer of a $5 \times 5$ patch configuration and RLU nonlinearity. The best results for IR are obtained using the first (3.78% EER) and second layers (85.18% VR) of a $3 \times 3$ patch configuration and Sig nonlinearity.

From the plots in Fig. 4 we notice that there isn’t a single patch configuration or nonlinearity that consistently outperforms the others. Nonetheless, some relationships do emerge when focusing on each method individually. For CR, employing more patches generally produces better performance. This is because, by taking the sum of the set, patches with poor performance can be strengthened by those with better performance. Thus, adding patches will produce a larger response. While IR requires fewer patches for better performance due to relying on each patch to perform similarly (i.e., no specific mechanism is in place for adjusting poor performing patches).

The experiments also revealed what we refer to as the ‘hard impostor’ phenomenon. Fig. 5 displays an example in which the impostor score distribution separates into multiple modes. This occurs when a set of false peaks from impostor match pairs are refined/sharpened similar to authentic comparisons. Continuing to iterate with each layer only further perpetuates the problem and pushes more impostor scores closer to the authentic distribution (i.e., causing more false positives and thereby decreasing the VR, but not necessarily affecting the rank-1 identification rate since a large number of authentic scores are well above the EER and VR score thresholds). This is mitigated by cross-validating over refinement and nonlinearity for each layer.

Fig. 6 shows the distribution of nonlinearities and refinements from searching over each during cross-validation (referred to as Dynamic Refinement) for each patch configuration.

![Fig. 6: Distribution across nonlinearities (top row) and refinement methods (bottom row) used at each layer when searching over each during cross-validation (referred to as Dynamic Refinement) for each patch configuration.](image)

![Fig. 8: Performance of the Dynamic Refinement (DYN).](image)

Correlation filters (CFs) are designed to specify a desired output for authentic and impostor matches and are widely used in many biometric applications. In this paper we presented Stacked Correlation Filters (SCFs), a fundamentally new CF paradigm where instead of a single CF, we use a cascaded stack of filters to achieve the desired CF outputs. Extensive experimentation demonstrates the effectiveness of SCFs, achieving substantial performance gains over a single CF under 1 : 1 image matching scenarios.

3. CONCLUSIONS

Correlation filters (CFs) are designed to specify a desired output for authentic and impostor matches and are widely used in many biometric applications. In this paper we presented Stacked Correlation Filters (SCFs), a fundamentally new CF paradigm where instead of a single CF, we use a cascaded stack of filters to achieve the desired CF outputs. Extensive experimentation demonstrates the effectiveness of SCFs, achieving substantial performance gains over a single CF under 1 : 1 image matching scenarios.

---

2Similar to CR, the DYN method works best with more patches.
References


