# STACKED CORRELATION FILTERS FOR BIOMETRIC VERIFICATION

*Jonathon M. Smereka, Vishnu Naresh Boddeti, and B.V.K. Vijaya Kumar*

Carnegie Mellon University
5000 Forbes Ave, Pittsburgh, PA 15213

*Andres Rodriguez*

Air Force Research Laboratory
Dayton, OH 45433

## 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.

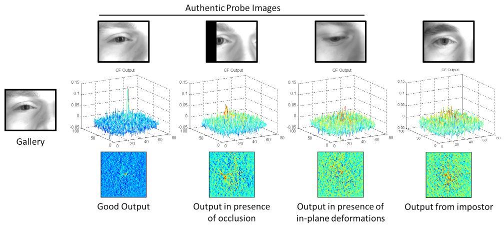


**Fig. 1**: Example correlation outputs when comparing a gallery template to several different probe images. A less discernible correlation output is obtained under difficult match scenarios while a good output (sharp peak at the location of best match) is obtained from the pair with few difficulties.

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

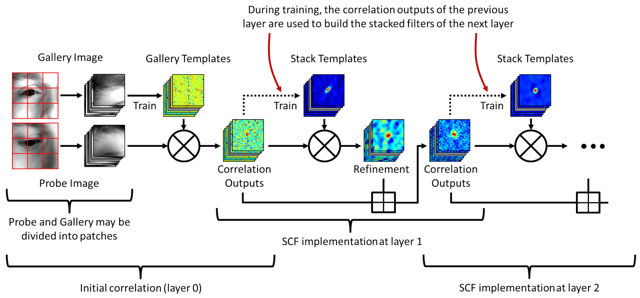

During training, the correlation outputs of the previous layer are used to build the stacked filters of the next layer
Gallery Image
Gallery Templates
Stack Templates
Stack Templates
Train
Train
Train
Probe Image
Correlation Outputs
Refinement
Correlation Outputs
Probe and Gallery may be divided into patches
SCF implementation at layer 1
Initial correlation (layer 0)
SCF implementation at layer 2

**Fig. 2**: Stacked Correlation Filter (SCF) Overview. Operating first on the outputs from an initial matching stage, additional sets of CFs are consecutively layered with each set designed to refine the previous layer's output to the desired ideal output.

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).

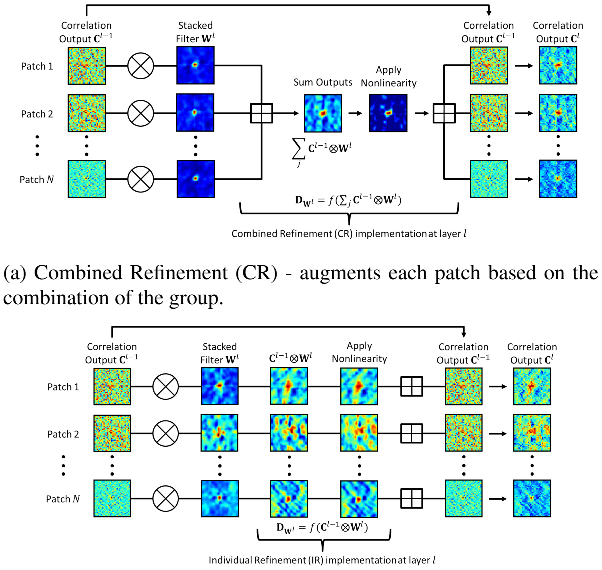
The use of sequential predictions (feeding the output of predictors from a previous stage to the next) has been revisited many times in the literature. In [16, 17] sequential prediction is applied to natural language processing tasks, while in [18] a face detection system was developed consisting of 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.


Correlation Output $\mathbf{C}^{l-1}$
Stacked Filter $\mathbf{W}^{l}$
Correlation Output $\mathbf{C}^{l-1}$
Correlation Output $\mathbf{C}^{l}$
Patch 1
Patch 2
Patch $N$
Sum Outputs
Apply Nonlinearity
$\sum_j \mathbf{C}^{l-1} \otimes \mathbf{W}^{l}$
$\mathbf{D}_{\mathbf{W}^l} = f(\sum_j \mathbf{C}^{l-1} \otimes \mathbf{W}^{l})$
Combined Refinement (CR) implementation at layer $l$

(a) Combined Refinement (CR) - augments each patch based on the combination of the group.

Correlation Output $\mathbf{C}^{l-1}$
Stacked Filter $\mathbf{W}^{l}$
$\mathbf{C}^{l-1} \otimes \mathbf{W}^{l}$
Apply Nonlinearity
Correlation Output $\mathbf{C}^{l-1}$
Correlation Output $\mathbf{C}^{l}$
Patch 1
Patch 2
Patch $N$
$\mathbf{D}_{\mathbf{W}^l} = f(\mathbf{C}^{l-1} \otimes \mathbf{W}^{l})$
Individual Refinement (IR) implementation at layer $l$

(b) Individual Refinement (IR) - augments each individual patch.

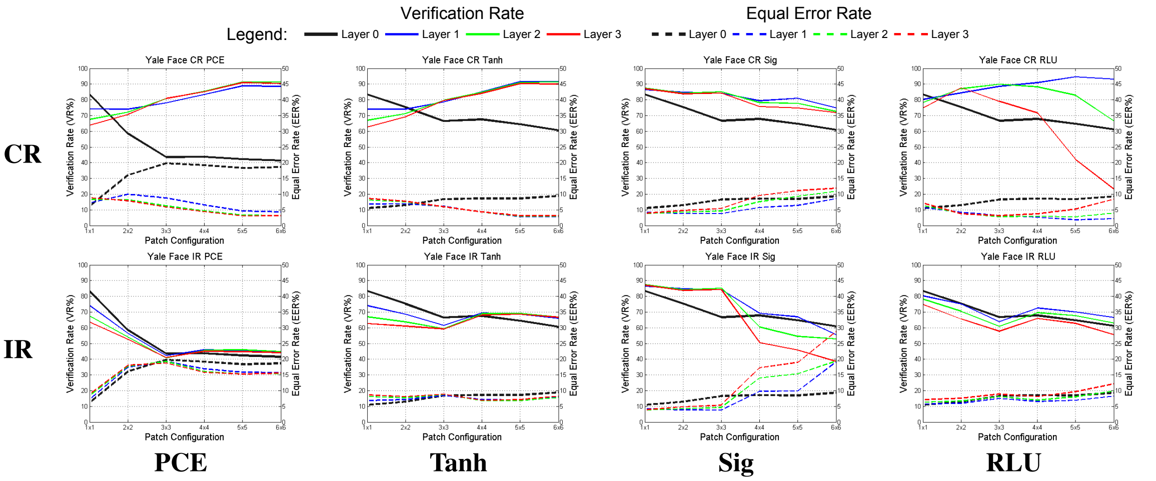
**Fig. 3**: Overview of each method for computing the refinement to be added back to the previous layer's outputs.

CFs are a class of classifiers which are generally designed for high localization performance and can even be built with a single training image. The correlation output from an authentic probe image and gallery template should exhibit a peak at the location of the best match and no such peak for an impostor. Each gallery template is designed to achieve such behavior on training data and this peak behavior is hypothesized to extend to testing data from the same user.

The main idea behind CF design is to control the shape of the correlation output between the training image and filter by minimizing the mean square error (MSE) between the actual and the desired correlation output for an authentic (or impostor) input. Thus, conceptually, CFs are regressors which map the image features to a specified output. Under very challenging conditions, a single CF may be insufficient to deal with the wide range of variations in the image features (e.g., see Fig. 1). The intuition behind SCFs is to use a layered approach to refine the correlation outputs from initial matching to provide better discrimination between authentic and impostor pairs. The 'stack' is built as a set of sequential CFs, the first layer is applied to the output after correlating the image features (referred to as 'layer 0'), and the subsequent layers are applied to the refined outputs of the previous layer, as in Fig. 2.

### 2.1. Correlation Output Refinement

We train the SCFs using only the correlation outputs and corresponding similarity/dis-similarity labels per match pair from the previous layer. In the case where the image is divided into multiple patches (e.g., Fig. 2), the SCF for the next layer is designed as a multi-channel CF with the correlation outputs of the patches ($N$ total) from the previous layer constituting the 'features' for each channel (see [24] for more details). Since correlation is linear, as in other multi-layer



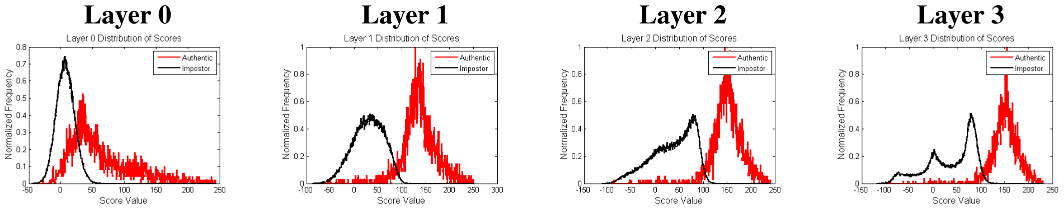
| | VR % | EER % |
|---|---|---|
| Layer 0 | 64.74 | 8.42 |
| Layer 1 | 94.62 | 1.84 |
| Layer 2 | 82.89 | 2.79 |
| Layer 3 | 42.06 | 5.21 |

Best CR Results: 5×5 RLU

| | VR % | EER % |
|---|---|---|
| Layer 0 | 66.71 | 8.33 |
| Layer 1 | 84.30 | 3.78 |
| Layer 2 | 85.18 | 4.62 |
| Layer 3 | 84.19 | 5.38 |

Best IR Results: 3×3 Sig

**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.



Initial refinement

Hard impostors emerging

Distribution separates

| | Fisher Ratio |
|---|---|
| Layer 0 | 1.16 |
| Layer 1 | 5.01 |
| Layer 2 | 4.31 |
| Layer 3 | 3.72 |

**Fig. 5**: Example of how ‘hard impostors’ can negatively effect 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.

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 correlated 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 ($f(\sum_{j=1}^{N}\mathbf{C}_j)$).
- **Individual Refinement (IR)** - A nonlinear function, $f$, is applied to each SCF output ($f(\mathbf{C}_j)$) 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. 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).[1] 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%.

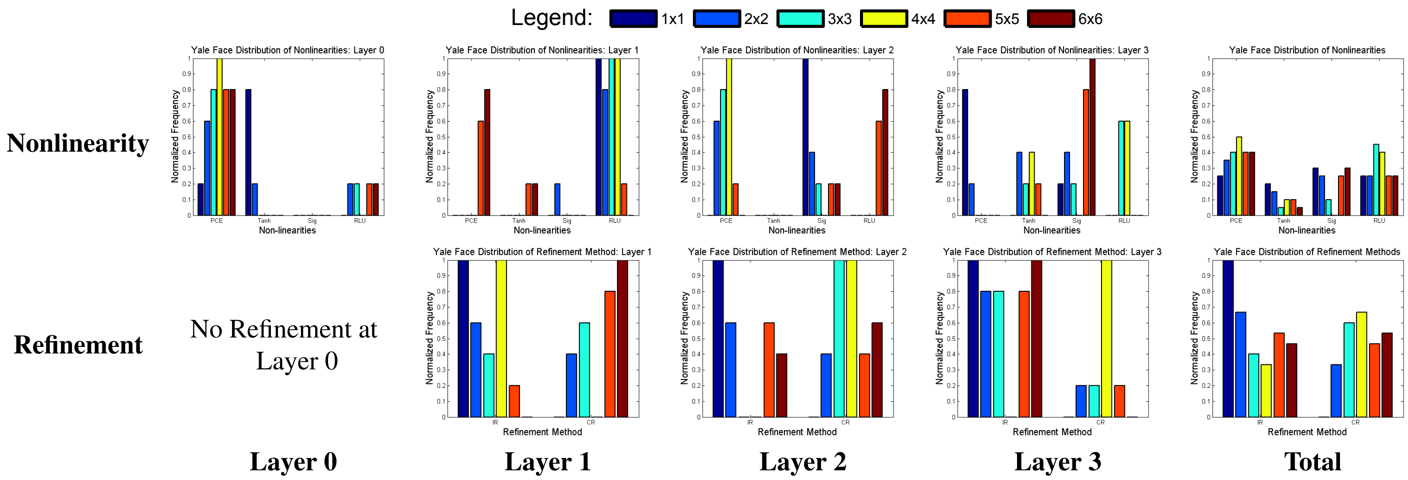


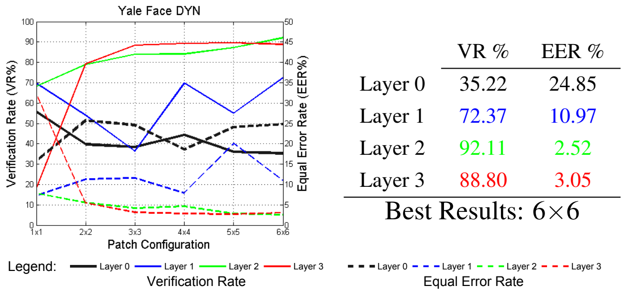
**Fig. 6**: Distribution across nonlinearities (top row) and refinement methods (bottom row) used at each layer when searching over each during cross-validation (referred to as Dynamic Refinement) for each patch configuration.

SCF layers over six patch configurations (non-overlapping rectangular patches, e.g., Fig. 2 displays a 3×3 configuration). Best results for CR (94.62% VR, 1.84% EER) are obtained using the first layer of a 5×5 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 3×3 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 training, and Fig. 8 displays the corresponding performance. Best results are obtained at the second layer of a 6×6 patch configuration (92.11% VR, 2.52% EER). While the best overall performance is achieved from CR, there is a significant improvement of the stability of the model (across all patches) when using DYN. Progress is no longer limited to the first or second layer and then quickly degrading, instead it becomes more gradual across several layers. Thus, rather than needing to empirically test each refinement method and nonlinearity, we can now largely ensure that improvement will occur as long the two images are divided into patches for matching.[2]



| | VR % | EER % |
|---|---|---|
| Layer 0 | 35.22 | 24.85 |
| Layer 1 | 72.37 | 10.97 |
| Layer 2 | 92.11 | 2.52 |
| Layer 3 | 88.80 | 3.05 |

Best Results: 6×6

**Fig. 8**: Performance of the Dynamic Refinement (DYN).

Finally, by examining the histograms in Fig. 6 we note that there is little correlation with regard to when one refinement or nonlinearity is better than another. Rather, what stands out is that the folds make similar choices, e.g., based on the experiments it is unlikely that the nonlinearity implemented at layer 2 of one fold will differ from that of another fold. However, we do notice that overall the Tanh and Sig operations are very rarely used after the initial CFs (layer 0).

## 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.

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

# References

[1] B. V. K. Vijaya Kumar, A. Mahalanobis, and R. Juday, *Correlation Pattern Recognition*, New York: Cambridge University Press, 2005.

[2] J.F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 37, no. 3, pp. 583–596, March 2015.

[3] M. D. Rodriguez, J. Ahmed, and M. Shah, "Action MACH a spatio-temporal maximum average correlation height filter for action recognition," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, June 2008, pp. 1–8.

[4] H. Kiani, T. Sim, and S. Lucey, "Multi-channel correlation filters for human action recognition," in *IEEE Int. Conf. on Image Processing*, Oct 2014, pp. 1485–1489.

[5] A. Rodriguez, V. N. Boddeti, B. V. K. Vijaya Kumar, and A. Mahalanobis, "Maximum margin correlation filter: A new approach for localization and classification," *IEEE Trans. on Image Processing*, vol. 22, no. 2, pp. 631–643, 2012.

[6] C. K. Ng, M. Savvides, and P. K. Khosla, "Real-time face verification system on a cell-phone using advanced correlation filters," in *IEEE Workshop on Automatic Identification Advanced Technologies*, Oct 2005, pp. 57–62.

[7] M. Zhang, Z. Sun, and T. Tan, "Perturbation-enhanced feature correlation filter for robust iris recognition," *Biometrics, IET*, vol. 1, no. 1, pp. 37–45, March 2012.

[8] J. M. Smereka, V. N. Boddeti, and B. V. K. Vijaya Kumar, "Probabilistic deformation models for challenging periocular image verification," *IEEE Trans. on Information Forensics and Security*, vol. 10, no. 9, pp. 1875–1890, Sept 2015.

[9] A. Meraoumia, S. Chitroub, and A. Bouridane, "Multimodal biometric person recognition system based on fingerprint & finger-knuckle-print using correlation filter classifier," in *IEEE Int. Conf. on Communications*, June 2012, pp. 820–824.

[10] P.H. Hennings-Yeomans, B. V. K. Vijaya Kumar, and M. Savvides, "Palmprint classification using multiple advanced correlation filters and palm-specific segmentation," *IEEE Trans. on Information Forensics and Security*, vol. 2, no. 3, pp. 613–622, Sept. 2007.

[11] George W. Quinn and Patrick J. Grother, "Performance of face recognition algorithms on compressed images," Tech. Rep. NISTIR 7830, National Institute of Standards and Technology (NIST), Dec. 2011.

[12] P. Jonathon Phillips, W. Todd Scruggs, Alice J. O'Toole, Patrick J. Flynn, Kevin W. Bowyer, Cathy L. Schott, and Matthew Sharpe, "FRVT 2006 and ICE 2006 large-scale results," Tech. Rep., National Institute of Standards and Technology (NIST), March 2007.

[13] P. J. Phillips, J. R. Beveridge, B. A. Draper, G. Givens, A. J. O'Toole, D. S. Bolme, J. Dunlop, Yui Man Lui, H. Sahibzada, and S. Weimer, "An introduction to the good, the bad, & the ugly face recognition challenge problem," in *IEEE Int. Conf. on Automatic Face Gesture Recognition and Workshops*, 2011, pp. 346–353.

[14] M. Alkanhal, B. V. K. Vijaya Kumar, and A. Mahalanobis, "Improving the false alarm capabilities of the maximum average correlation height correlation filter," *Optical Engineering*, vol. 39, pp. 1133–1141, 2000.

[15] Yi Li, Zhiyan Wang, and Haizan Zeng, "Correlation filter: an accurate approach to detect and locate low contrast character strings in complex table environment," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 26, no. 12, pp. 1639–1644, Dec 2004.

[16] William W. Cohen, "Stacked sequential learning," Tech. Rep., DTIC Document, 2005.

[17] H. Daumé Iii, J. Langford, and D. Marcu, "Search-based structured prediction," *Machine learning*, vol. 75, no. 3, pp. 297–325, 2009.

[18] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2001.

[19] D. Munoz, J. A. Bagnell, and M. Hebert, "Stacked hierarchical labeling," in *European Conf. on Computer Vision*, 2010, pp. 57–70.

[20] S. Ross, D. Munoz, Martial. Hebert, and J. A. Bagnell, "Learning message-passing inference machines for structured prediction," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2011, pp. 2737–2744.

[21] T. Murakami and K. Takahashi, "Accuracy improvement with high confidence in biometric identification using multihypothesis sequential probability ratio test," in *IEEE Int. Workshop on Information Forensics and Security*, 2009, pp. 67–70.

[22] V.P. Nallagatla and V. Chandran, "Sequential decision fusion for controlled detection errors," in *Conf. on Information Fusion*, 2010.

[23] K. Lee, J. Ho, and D. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 684–698, May 2005.

[24] V. N. Boddeti, T. Kanade, and B. V. K. Vijaya Kumar, "Correlation filters for object alignment," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, 2013.