

# A Still-to-Video Face Verification System Using Advanced Correlation Filters

Chunyan Xie<sup>1</sup>, B. V. K. Vijaya Kumar<sup>1</sup>, S. Palanivel<sup>2</sup>, and B. Yegnanarayana<sup>2</sup>

<sup>1</sup>Dept. of ECE., Carnegie Mellon University, Pittsburgh, PA, 15213, USA  
{kumar, chunyanx}@ece.cmu.edu

<sup>2</sup>Dept. of CSE, Indian Institute of Technology, Madras, Chennai-600 036, India.  
{yegna, spal}@cs.iitm.ernet.in

**Abstract.** In this paper we propose to use the technology of multiple correlation filters to address the issues involved in still-reference-to-video-sequence face verification. In particular, we assume that the gallery consists of one or a few digital images of a person's face as reference and the probe set consists of video sequences as test data. The test data consists of a sequence of face images of a naturally moving person, captured from one or more cameras. An automatic face recognition system involves face/head tracking and cropping, correlation filter based face matching, and combining the evidence from multiple frames of the face images.

## 1 Introduction

Biometric verification is being investigated for many access control and e-commerce applications as biometric signatures offer the advantage that they cannot be stolen or forgotten like passwords. Examples of biometrics include face images, fingerprints, voice patterns, iris patterns, etc. Biometric verification (or authentication) refers to the task of determining whether a stored biometric matches a live biometric or not and thus determining whether the claimant of an identity is an authentic or an impostor. A closely related problem is biometric identification, in which a live biometric may be matched against a database of biometrics to determine the identity of the subject. Sometimes the term recognition is used to refer both of them. In this paper, we focus on the problem of subject verification based on face images.

Face verification is attractive because face images can be captured in a non-intrusive way, but the recognition performance drops for many face recognition methods when the test image does not enjoy the same geometric and photometric conditions as the training images, e.g., frontal view vs. side view. To address the problem of pose difference, video-to-video face recognition approaches have been proposed [1,2], where both gallery and probe sets consist of video sequences containing the face in some frames. Such video-based face recognition systems exploit much more information than still image based face recognition systems. However, most current video-based face recognition methods have been tested only for the cases where the training video sequence and the test video sequence are captured in the same conditions. Also, the video-to-video approaches require video

sequences for training, which may not be available in some applications. These are some of the factors that limit the application of video-to-video face recognition.

We propose in this paper a still-to-video face verification system based on correlation filters. The gallery consists of a few high-quality digital images of a person's face as reference and the probe set consists of video sequences of a person moving normally in an area. The test subject may not be cooperative, in the sense of providing defined views to the camera. In this scenario, our hypothesis is that those frames that have a view approximately corresponding to the views in the training images are able to provide partial evidence of the identity of the test subject. Combining those partial evidences from multiple frames may lead to a better verification decision.

For this method to succeed, the matching algorithm for a single frame should have a very low false acceptance rate, otherwise it is very easy to accumulate a lot of false evidence. Correlation filters such as the minimum average correlation energy (MACE) filters [3] usually emphasize the high frequency components of the training image, so there is less chance for a MACE filter to produce high confidence match with an imposter image. So the MACE filter is suitable in our application.

The rest of this paper is organized as follows. In Sec. 2, we introduce the concept of face recognition using correlation filters and introduce the three correlation filters used in this work. In Sec. 3, a still-to-video face verification system is described and each component is introduced in detail. Our numerical experiments results are presented in Sec. 4 and final discussion is in Sec. 5.

## 2 Advanced Correlation Filters

The basic concept of still-to-still face verification [3] using a correlation filter is illustrated in Fig. 1. There are two stages, namely the enrollment stage and the verification stage. During the enrollment stage, one or multiple images of an individual's face are acquired. These multiple training images should reflect the expected variability (perhaps due to rotation, scale changes, illumination changes, etc.) in the face image. The 2-D Fourier transforms (FT) of these training face images are used by a correlation filter design algorithm to determine a single frequency-domain array (called a correlation filter) that is stored. In the verification stage, the user presents a live face image and its 2-D FT is multiplied by the stored frequency domain array and the inverse 2-D FT of this product results in the correlation output. If the filter is well designed, we should observe a strong peak in the correlation output when the input image is from an authentic and no such discernible peak if the input is from an imposter. The location of the peak indicates the position of the input image and thus provides automatic shift-invariance. One measure of the peak sharpness is the peak-to-sidelobe-ratio (PSR), [3] defined as follows

$$PSR = \frac{Peak - mean(sidelobe)}{std(sidelobe)}, \quad (1)$$

where the sidelobe refers to a small region centered at, but excluding the peak.

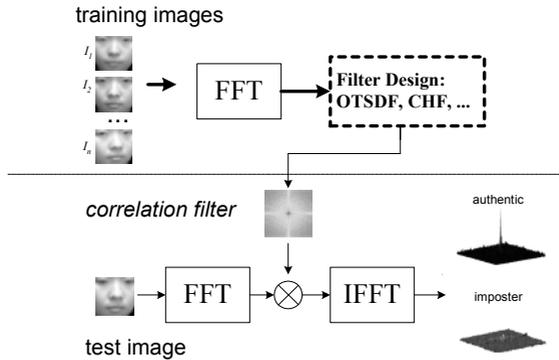


Fig 1. The concept of still-to-still face verification using correlation filter

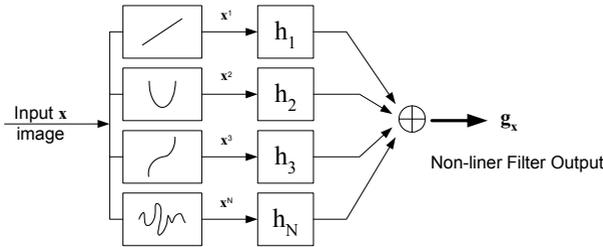


Fig 2. N-th Order polynomial correlation filter architecture [4]

Many advanced correlation filter have been developed with different criteria such as maximal noise tolerance, and maximal peak sharpness [3-6].

**Optimal Tradeoff Filter (OTF).** The minimum variance synthetic discriminant function (MVSDF) filter [3] was developed to maximize tolerance to input noise and the minimum average correlation energy (MACE) filter [3] was developed to suppress sidelobes in correlation output and thus produce a sharp peak. However, the MVSDF filter emphasizes low spatial frequencies whereas the MACE filter emphasizes high spatial frequencies and thus they are conflicting. Using the multi-criteria optimization approach, an optimal tradeoff filter (OTF) that optimally trades off distortion tolerance against discrimination was introduced in [5].

**Optimal Tradeoff Circular Harmonic Function Filter (OTCHF)** [6]. A common distortion in acquiring face images is in-plane rotation of the face image. The coordinate transformations such as polar mappings can be combined with Fourier transforms to design circular harmonic function (CHF) [6] filters that can handle in-plane rotation. Since the face images are mapped into polar coordinates, in-plane rotations appear as shifts in the polar domain and shifts can be handled by correlation filters. The CHF filter design optimally trades off among various correlation filter performance criteria while achieving specified in-plane rotation response of the correlation peak. Although polar mappings are used during filter design, filter use

during the verification stage is just like in all other correlation filter cases and no coordinate transformation is needed during the verification stage.

**Polynomial Correlation Filter (PCF).** The main idea underlying PCFs [4] is illustrated in Fig. 2. Although any point nonlinearity can be used for PCFs, we consider powers (e.g.  $x^2$ ,  $x^3$ , etc.) for reasons of analytical simplicity. The analysis and the form of the solution remain the same irrespective of the non-linearity used, with the only restriction that the nonlinearity should be point nonlinearity. The objective is to find the filters  $h_i(m,n)$  such that the structure shown in Fig. 2 optimizes a performance criterion of interest. More details of PCF filter design can be found elsewhere [4].

### 3 Still-to-Video Face Verification System

The proposed still-to-video face verification system has two stages: the training process based on one or more still images and the automatic face verification based on video sequences. A face template is cropped from one training image of a subject, and multiple correlation filters are then built from that template. The test system consists of three components: a face detecting and tracking block to crop the face image from each frame, a face verification block that uses multiple correlation filters to derive partial evidences from each frame, and an evidence accumulation block to summarize evidence and yield a score for final face verification decision.

Cropping face images from each frame is implemented in two steps. In step one, a region of interest (ROI) of human head is detected from each frame by using the motion information derived from difference image of two consecutive frames [7]. In step two, we use a template matching method to accurately locate the face region in the ROI. This exhaustive search is not very efficient, but still computationally affordable since the searching range has been narrowed by the ROI.

The verification decision is based on a video sequence, but at frame level we use correlation filters to derive partial evidences. OTSDF filter has been shown good verification performance for still-to-still face verification across illumination and expression variation, so it is used as a baseline filter in this paper. For each frame, three correlation filters are used to generate three PSRs. The PSR values are used as partial evidences. We use a naïve combination method that produces a weighted sum of matching scores, for which the weights are chosen by trial and error. Since different correlation filter methods focus on different performance criteria, the combination of multiple correlation filters can lead to a better decision.

$$s_i = \begin{cases} 0 & PSR_i < T_1 \\ PSR_i - T_1 & T_1 \leq PSR_i < T_2 \\ \exp(PSR_i - T_2) + (T_2 - 1 - T_1) & \text{otherwise} \end{cases}, \quad Score = \sum_{i=1}^N s_i \quad (2)$$

One property of MACE-type correlation filter is that it usually emphasizes high frequency components, and therefore there is little chance for a correlation filter to generate high PSR values with impostors, which implies that a higher PSR value indicates matching with high confidence. Based on this property, we propose an evidence accumulation method that maps each PSR value to a score by giving more

weight to the higher PSR and then sums up those scores to generate a final score for each test sequence. The mapping function is a two-threshold monotonically increasing function shown in Eq. (2).

## 4 Experiments

In order to test the proposed system, we collected a video sequence database with 10 subjects. During the data collection, the subject is asked to pretend to be working in front of a computer with a webcam on top of the monitor. The subject may change his/her position, expression and orientation of the head. For each subject, 300 frames are collected as a test video sequence. Whenever possible, we asked the subject to submit several high quality still images with a frontal face. Otherwise, we take their pictures using a digital camera. The cropped gallery images and some examples from the probe video sequence of these 10 subjects are shown in Fig. 3.

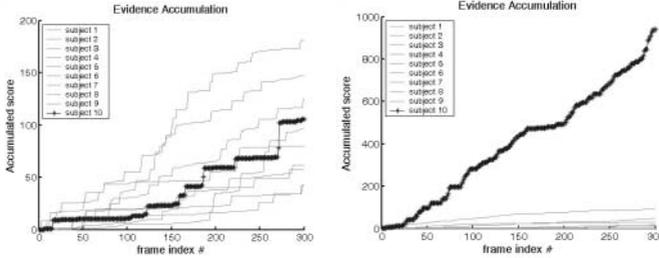
We tested the OTSDF filter, the OTCHF filter, the PCF and the combination of the OTCHF filter and the PCF for each sequence. The OTSDF filter is used as a baseline method for comparison. For evidence accumulation, we tested a baseline method of direct summation of all PSR values for each sequence to generate a final score. We also tested the proposed nonlinear PSR accumulation method, for which we need to choose two thresholds for the two-level-threshold mapping function. These thresholds can be chosen based on imposters' face video sequences. In our experiment, we additionally collected other 4 people's face video sequences as imposter sequences. We tested these imposter sequences for each person in the 10-subject database and get the PSR values for imposters in each case. From the mean and variance of these PSR values we compute the proper thresholds for each subject in the database.



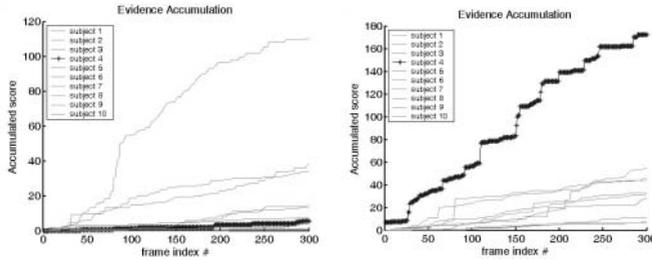
**Fig. 3.** The gallery set: including 10 templates manually cropped for 10 subjects; the probe set 100 sample images from the probe set: randomly chosen 10 for each subject

**Table 1.** Number of verification errors

	OTSDF	PCF	OTCHF	PCF+OTCHF
Direct summation	9	7	6	6
Weighted summation	6	1	6	0



**Fig. 4.** The accumulated scores over frame index number of subject 10 for the OTSDF filter (left) and the PCF (right)



**Fig. 5.** The accumulated score over frame index number of subject 4, PCF filter (left) and OTCHF filter (right)

In our experiments if the sequence with the highest score has the same class label as the template, it represents a correct verification; otherwise, we record an error. The face verification performances are shown in Table 1. In each cell we show the number of errors.

We can see from the Table 1 that the weighted summation outperforms the direct summation, which means the proposed evidence combination method is able to effectively exploit video sequence information to get better verification performance. We also see that the advanced correlation filters, e.g., PCF filter and OTCHF filter, perform better than the OTSDF filter. By combining the results of different correlation filters, we get even better verification rate.

In Figs. 4 and 5, we illustrate the performance of the PCF filter and OTCHF filter by means of two examples. In each plot, the thicker curve is for authentic and the lighter curves are for imposters. Fig. 4 shows evidence accumulation curves of an OTSDF filter and a fifth order PCF filter for subject-10. We can see that by using OTSDF filter there are three imposters whose accumulated verification scores are larger than the authentic, but by using PCF we essentially apply a nonlinear classifier for each frame and get the better verification performance. In our experiments, the

PCF generally outperforms its linear counterpart. In Fig. 5, we see the evidence accumulation curves for subject-4 by using the PCF and the OTCHF filter. In subject 4's testing sequence, a continuous tilt of the head is present, resulting in the in-plane rotation variation of his cropped face images. OTCHF filter is particularly designed to have in-plan rotation tolerance so it shows better performance than the PCF.

## 5 Conclusions

This paper introduced a still-reference-to-video-sequence face verification system. This system is useful in applications requiring the recognition of a subject in a video sequence when only one or a few still training images are available. The automatic face verification system consists of face detection and cropping, face matching and evidence accumulation procedures. Numerical experiments show that by using multiple correlation filters and nonlinear evidence accumulation method, the proposed system is able to effectively verify a video test sequence based on a still reference image. In the follow-on work, we will further investigate the multiple correlation filter combination technology and exploit temporal information in evidence accumulation methods.

## Acknowledgement

This research in the Improved Face Identification from Video (IFIVE) project at CMU is supported in part by the Technology Support Working Group (TSWG). The authors also acknowledge each individual appearing in our face database.

## References

- [1] S. Zhou, V. Krueger and R. Chellappa, "Probabilistic Recognition of human faces from video", Computer Vision - ECCV 2002, 7th European Conference on Computer Vision, Copenhagen, Denmark, Part III, pp. 681-697, 2002.
- [2] X. Liu and T. Chen, "Video-based face recognition using adaptive hidden Markov models." IEEE International Conference on Computer Vision and Pattern Recognition 2003, Madison, Wisconsin, June 2003.
- [3] B. V. K. Vijaya Kumar, M. Savvides, C. Xie, K. Venkataramani, J. Thornton and A. Mahalanobis: "Biometric verification with correlation filters". Applied Optics, Vol. 43, No.2, Jan. 2004.
- [4] B. V. K. Vijaya Kumar and A. Mahalanobis, "Recent advances in composite correlation filter designs," *Asian Journal of Physics*, Vol. 8, No. 4, pp. 407-420, 1999
- [5] Ph. Refregier, "Optimal trade-off filters for noise robustness, sharpness of the correlation peak, and Horner efficiency," *Opt. Lett.* 16, pp. 829-831, 1991.
- [6] B. V. K. Vijaya Kumar, A. Mahalanobis, and A. Takessian, "Optimal tradeoff circular harmonic Function correlation filter methods providing controlled in-plane rotation response". IEEE Trans. Image Processing, Vol. 9, No. 6, pp.1025-1034, 2000.
- [7] S. Palanivel, B. S. Venkatesh and B. Yegnanarayana, "Real time face authentication system using autoassociative neural network models". IEEE International Conference on multimedia and Expo., pp. 257-260, Baltimore, July 2003.