Face Recognition from Video by Matching Image Sets

Tat-Jun Chin* James U Konrad Schindler David Suter
Institute of Vision Systems Engineering,
Monash University, Victoria, Australia.
{tat.chin | james.u | konrad.schindler | d.suter}@eng.monash.edu.au

Abstract

As opposed to still-image based paradigms, video-based face recognition involves identifying a person from a video input. In video-based approaches, face detection and tracking are performed together with recognition, as usually the background is included in the video and the person could be moving or being captured unknowingly. By detecting and raster-scanning a face sub-image to be a vector, we can concatenate all extracted vectors to form an image set, thus allowing the application of face recognition algorithms based on matching image sets. It has been reported that linear subspace-based methods for face recognition using image sets achieve good recognition results. The challenge that remains is to update the linear subspace representation and perform recognition on-the-fly so that the recognition-from-video objective is not defeated. Here, we demonstrate how this can be achieved by using a well-studied incremental SVD updating procedure. We then present our online face recognition-from-video framework and the recognition results obtained.

1. Introduction

It is conjectured that many still image-based face recognition algorithms achieve uninspiring results in practice because they rely heavily on the assumption of a fixed pose and expression by the subject as well as conducive background and illumination conditions. Given that many existing camera surveillance systems require to operate unobtrusively and without subject cooperation, a track-and-recognize framework would be more suitable for the face recognition task. With possible variations in subject appearance and imaging conditions encoded in the video sequence, it stands to reason that video-based face recognition algorithms should be more superior to their still image counterparts. Furthermore, temporal continuity in a video sequence can be exploited to aid in the recognition decision.

Contemporary face recognition from video algorithms are dominated by probabilistic approaches such as [18, 29, 28, 15, 16]. Here, we adopt a non-probabilistic direction towards achieving face recognition from video. More specifically, we observe that a video sequence can be treated as a set of images, and hence we apply face recognition algorithms based on using linear subspaces to represent and distinguish image sets [13, 9, 26]. We argue for the case of using linear subspace approaches due to their simplicity and comparable accuracy, and commercial face recognition products such as the Toshiba FacePass™ are testaments to their effectiveness.

We employ currently available face detection algorithms [19, 20, 24] which are able to achieve high accuracy and detection rates to detect faces in our video input, thus unburdening the task of tracking from our face recognition module. In order to update our face representation and perform recognition online, we make use of well-known SVD or PCA incremental updating procedures [5, 6, 17]. Unlike previous approaches, for the purpose of matching or distinguishing image sets using linear subspaces, we utilize a subspace distance metric that does not involve explicit computation of subspace *principal angles*. In this paper we present our framework for online computing, updating and distinguishing linear subspaces in order to achieve robust recognition-from-video. The results obtained demonstrates the effectiveness of our method.

2 Previous work

Traditionally, the video-based face recognition paradigm was approached from the still image perspective i.e. find an instance of a face in the video stream that is suitable under certain quality measures (e.g. pose, illumination, size etc.) and apply conventional still image-based recognition algorithms. Methods such as these are not considered true

*Tat-Jun Chin is a recipient of the Australia-Asia Awards 2004 conferred by the Department of Education, Science and Training (DEST) of the Government of Australia.

video-based face recognition in the sense that they do not coherently exploit both spatial and temporal information. For a survey of methods that evolved from still image techniques, see [27].

Recently, probabilistic approaches have gained popularity in video-based face recognition. For a recent survey, see [7]. In [29], tracked face appearances are modeled as a joint probability distribution of identity and motion using Sequential Importance Sampling (SIS) and the recognition decision is obtained via marginalization. In [15], face appearance manifolds are approximated by a finite number of infinite extent subspaces and temporal information is used to robustly estimate the operating part of the manifold. Some of these techniques are essentially operating on the "still-to-video" scenario i.e. matching a video sequence of a face to a still training image. Furthermore, methods such as [18, 29, 28] formulate both tracking and recognition of faces in the same probabilistic framework — an approach we do not adopt here since we believe that decoupling tracking from recognition may improve the recognition results.

We identify our work as "video-to-video" based i.e. face recognition is performed by matching image sets or video sequences. A similar objective is pursued by [13, 1, 9, 26, 22]. We conjecture that a video-to-video system would be more robust since it takes advantage of all images available for training and recognition. We concentrate on algorithms that match image sets using linear subspaces. In [26], the Mutual Subspace Method (MSM) distinguishes image sets by approximating an image set of the same individual with a linear subspace and using the principal angles between subspaces as a distance measure. An extension of MSM is the Constrained Mutual Subspace Method (CMSM) [9], which is essentially an MSM performed on the projection of the original subspaces onto a *constrained* subspace resulting in more robust classifications. In [13], it was shown how boosting can be used to select the optimal set of principal angles for classification.

Implementations for real-time face recognition from video based on the subspace method were presented in [9, 14]. The CMSM method was implemented on a Visconti image processing LSI chip in [14], yielding excellent speed and recognition rates. Our work differs in that it is implemented entirely on an optimized software platform whilst neither compromising speed nor accuracy. Our work is probably most similar to [9], where a framework for high-speed linear subspace-based face recognition from video was presented. The incremental subspace updating procedure is not described clearly since it is not the main thrust of the paper, but seems to start from the eigen-decomposition of the complete correlation matrix. Our online subspace updating procedure is based on recent advances in incremental Singular Value Decomposition (e.g. see [5]), which is theoretically sound and computationally efficient.

3 Angles and distances between subspaces

We start by extracting the face in a video sequence (we assume only one face is present) by using a high speed face detector. The face sub-images are preprocessed and raster-scanned to form m-dimensional column vectors, m being the number of pixels. We concatenate all n vectors to form a matrix $A \in \mathbb{R}^{m \times n}$ (here, vector ordering is non-crucial). The core concept of the linear subspace method is that if A was constructed using face images, the columns of A can be approximately spanned by a low-dimensional linear subspace $\mathcal{A}$ (see [23, 3, 2, 8, 11]) of dimension r, where $r < n$ and $r \ll m$. Face recognition is done on the premise that different faces generate different subspaces, and distances between the subspaces can be used to distinguish faces.

The notion of angles between two subspaces, called *principal angles*, plays an integral part in quantifying the distance between two subspaces. The principal angles $\theta_1, \theta_2, \cdots, \theta_n \in [0, \pi/2]$ between two n-dimensional subspaces $\mathcal{P}$ and $\mathcal{Q}$, following the definition of [10], are defined by

$$\cos \theta_i = \max_{\mu \in \mathcal{P}} \max_{\nu \in \mathcal{Q}} \mu \cdot \nu = \mu_i \cdot \nu_i \, , \tag{1}$$

for $i = 1, ..., n$, subject to $\mu \cdot \mu = \nu \cdot \nu = 1$, $\mu \cdot \mu_j = 0$, $\nu \cdot \nu_j = 0 \, (1 \leq j \leq i-1)$. The vectors $\{\mu_i\}$ and $\{\nu_j\}$ are *principal vectors* corresponding to the pair $\mathcal{P}$ and $\mathcal{Q}$.

Given n-dimensional subspaces $\mathcal{P}$ and $\mathcal{Q}$ with principal angles $\{\theta_1, \theta_2, \cdots, \theta_n\}$ and $\theta_1 \geq \theta_2 \geq \cdots \geq \theta_n$, the similarity measure used in [26] is

$$S_{MSM}(\mathcal{P}, \mathcal{Q}) = \cos^2 \theta_n \, . \tag{2}$$

We use a distance metric for our classification task instead of a similarity measure. Specifically, we use the *chordal distance* metric which is defined as

$$d_c(\mathcal{P}, \mathcal{Q}) = \sqrt{\sum_{i=1}^{n} \sin^2 \theta_i} \, . \tag{3}$$

Both measures are essentially equivalent in the sense that they induce the same topology.

Let matrices P and Q be orthonormal bases for subspaces $\mathcal{P}$ and $\mathcal{Q}$ respectively. The *projection matrices* of $\mathcal{P}$ and $\mathcal{Q}$ are respectively

$$\mathsf{P} = PP^T \text{ and } \mathsf{Q} = QQ^T \, . \tag{4}$$

Unlike orthonormal bases, the projection matrices define their corresponding subspaces uniquely. The chordal distance can be computed from the projection matrices:

$$d_c(\mathcal{P}, \mathcal{Q}) = \frac{1}{\sqrt{2}} \|\mathsf{P} - \mathsf{Q}\|_F \, , \tag{5}$$

with $\|\cdot\|_F$ representing the matrix Frobenius norm. For a proof of the above, see, for example, [10].

4 Estimating and incrementally updating linear subspaces from image sets

Given the face image set $A \in \mathbb{R}^{m\times n}$, we wish to find an orthonormal basis for an r-dimensional subspace $\mathcal{A}$, with $r < n$ and $r \ll m$, that can span the columns of A. In practice A is almost always full rank, therefore the most we can do is to find $\mathcal{A}$ that spans A closely. An orthonormal basis of $\mathcal{A}$ can be obtained by performing an SVD on A and retaining the first r left singular vectors. The value r can be selected heuristically, e.g. the cumulative sum of the first r singular values of A should cover more than 99% of the sum of all the singular values of A, or analytically such as the methods highlighted in [12].

Given the matrix $A \in \mathbb{R}^{m\times n}$ and its singular value decomposition $A = U\Sigma V^T$, we compute the rank-r *singular value factorization* of A, given by

$$A^r = U^r\Sigma^r\left(V^r\right)^T \ , \tag{6}$$

with $r < n$ and $r \ll m$, $U^r = U\left(:,1:r\right)$, $\Sigma^r = \Sigma\left(1:r,1:r\right)$ and $V^r = V\left(:,1:r\right)$ by using a Matlab notation. If r is large enough, then $A^r \cong A$ and U^r is the orthonormal basis for a subspace that can closely span the columns of A. Hence, only the matrices U^r, Σ^r and V^r are retained, and U^r will be used for subspace distance computations via (5).

Incremental computation of the SVD is motivated by the infeasibility of performing the SVD on large matrices [10]. For our purpose, we are hampered not only by the sheer size of the input matrix (pixel number ranges from 400 to 10,000 depending on the system), we do not have a complete matrix to begin with as new images are acquired on-the-fly. Furthermore, the image set grows unboundedly as new image columns are continually being appended to the image set. Hence, there is a need to incrementally update our linear subspace representation of an image set so that it is always a good representation of all available images of a person. To this end, we follow the approach taken by [5, 17, 6].

Given new columns $C \in \mathbb{R}^{m\times c}$, we would like to find the SVD of $[A^r\ C]$, where $[\cdot]$ signifies matrix concatenations. We proceed by computing the following:

$$\begin{aligned} L &= \left(U^r\right)^T C \ , & (7)\\ H &= C - U^r L \ , & (8)\\ JK &\xleftarrow{\text{QR}} H \ . & (9) \end{aligned}$$

Geometrically, L is the projection of C onto the orthogonal basis U^r, H is the component of C orthogonal to the subspace spanned by U^r, and $H = JK$ is the QR-decomposition of H with J being an orthonormal basis of H and K can be interpreted as the projection of C onto the subspace orthogonal to the subspace spanned by U^r. We can then derive the following identity:

$$\begin{bmatrix} U^r & J \end{bmatrix} \begin{bmatrix} \Sigma^r & L \\ 0_{c\times r} & K \end{bmatrix} \begin{bmatrix} V^r & 0_{n\times c} \\ 0_{c\times r} & I_c \end{bmatrix}^T = \begin{bmatrix} A^r & C \end{bmatrix} \ , \tag{10}$$

with $0_{a\times b}$ denoting a zero matrix of size $a \times b$ and I_a signifying an identity matrix of size $a \times a$.

Let M be defined as the middle matrix of the left-hand-side of (10). We can diagonalize M by invoking the SVD,

$$M = U'\Sigma'\left(V'\right)^T \ , \tag{11}$$

and substituting the resultant into (10). If we define

$$\begin{aligned} U'' &= \begin{bmatrix} U^r & J \end{bmatrix} U' \ , & (12)\\ \Sigma'' &= \Sigma' \ , & (13)\\ V'' &= \begin{bmatrix} V^r & 0_{n\times c} \\ 0_{c\times r} & I_c \end{bmatrix} V' \ , & (14) \end{aligned}$$

then the updated SVD is

$$\begin{aligned} \begin{bmatrix} A^r & C \end{bmatrix} &= U''\Sigma''\left(V''\right)^T \ , & (15)\\ U^r &\longleftarrow U''\left(:,1:r\right) \ , & (16)\\ \Sigma^r &\longleftarrow \Sigma''\left(1:r,1:r\right) \ , & (17)\\ V^r &\longleftarrow V''\left(:,1:r\right) \ , & (18) \end{aligned}$$

with U^r updated to become the orthonormal basis for the r-dimensional subspace that can closely span the current image set $[A^r\ C]$. For potential numerical issues of the procedure and their solutions, refer to [5]. In theory, it is possible to incrementally update the subspace basis using only one or as many new input image vectors as desired (i.e. the size of matrix C). In practice, in order to reduce unnecessary updates due to lack of differences between consecutive frames, we update the subspace basis only after a batch of new image vectors are accumulated.

5 The face recognition from video framework

Figure 2 shows the framework for our face recognition system. The face recognition process starts by capturing a video sequence of a person in front of a fairly uniform background using a high speed FireWire camera. We make use of the Viola & Jones face detector implementation of [4] in our system to detect faces in the video sequence at frame rate. The faces present are subsequently cropped and preprocessed. We noticed that the face detector of [4] returns square sub-images with important facial features consistently in the center region, thus greatly simplifying facial image registration — apart from removing 25% of the

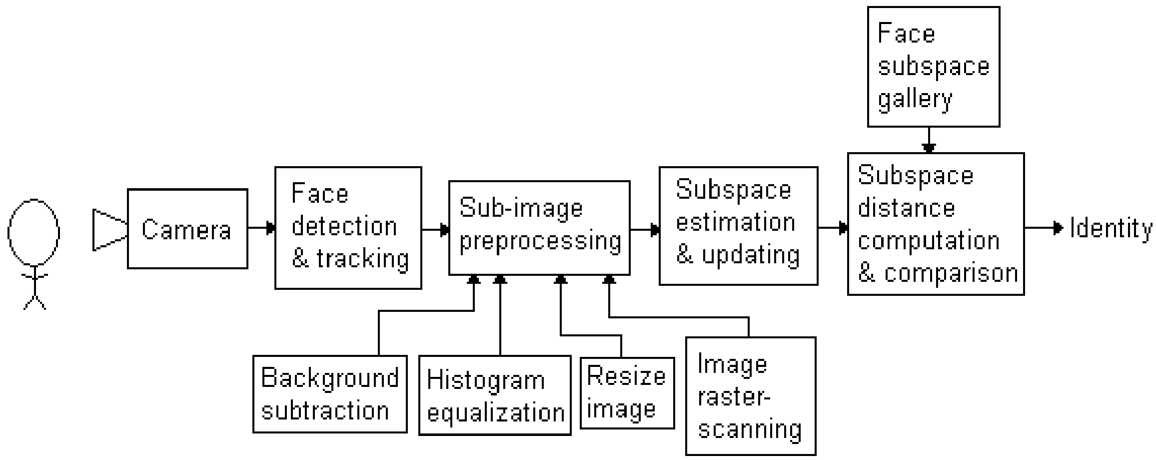

Figure 2. Our face recognition framework.

side columns, we use the face detector outputs *as they are* without performing any complex facial feature alignment. This preprocessing step functions as a crude background subtraction scheme as well. Figure 1 shows examples of the resulting sub-images and illustrates the pose and scaling variability our system can handle. We then equalize the histogram of the remaining pixels only to mitigate illumination effects, and reduce the sub-image to the pre-determined size of 32×24 pixels. The resulting sub-image is raster-scanned to form a high-dimensional vector. This image vector is fed to a subspace estimation/updating module to produce a subspace representation of the face currently being tracked. The next subsection details the algorithmic procedure of this module. Given an updated subspace, face recognition is performed by computing the distance of it to gallery subspaces trained previously from other subjects using the same steps described in this section. The current face is matched to the identity of the subspace with which it has the smallest chordal distance if the value dips below an empirically determined threshold.

Figure 1. Examples of cropped face detector outputs. Rows 1 and 3 show the original face detector outputs, while rows 2 and 4 show their cropped version. Notice that the face detector implementation of [4] manages to locate important features of faces in the center of the output sub-images despite vast changes in pose.

5.1 Pseudo code for subspace estimation and updating

Algorithm 1 shows the pseudo code for our subspace estimation and updating procedure based on the theory in Section 4. The input parameters to the algorithm are as follows:

- *maxFrames*: maximum length of video sequence to process (in terms of image frames).
- r: dimension of linear subspace to use.
- $n, n > r$: number of initial images to accumulate to estimate an initial subspace basis.
- *incr*: batch size of subsequent images to be used for subspace updating.

For our experiments, we used $maxFrames = 400$, $r = 25$, $n = 30$ and $incr = 5$. Our system managed to verify subject identities within 150 frames (approximately 13 seconds), while impostors are rejected at the end of the 400-th frame without having achieved the target proximity with any of the reference subspaces. The value $r = 25$ was determined empirically to be the most suitable value, and the previous work of [26] confirms this to be an optimal value for the case of face images.

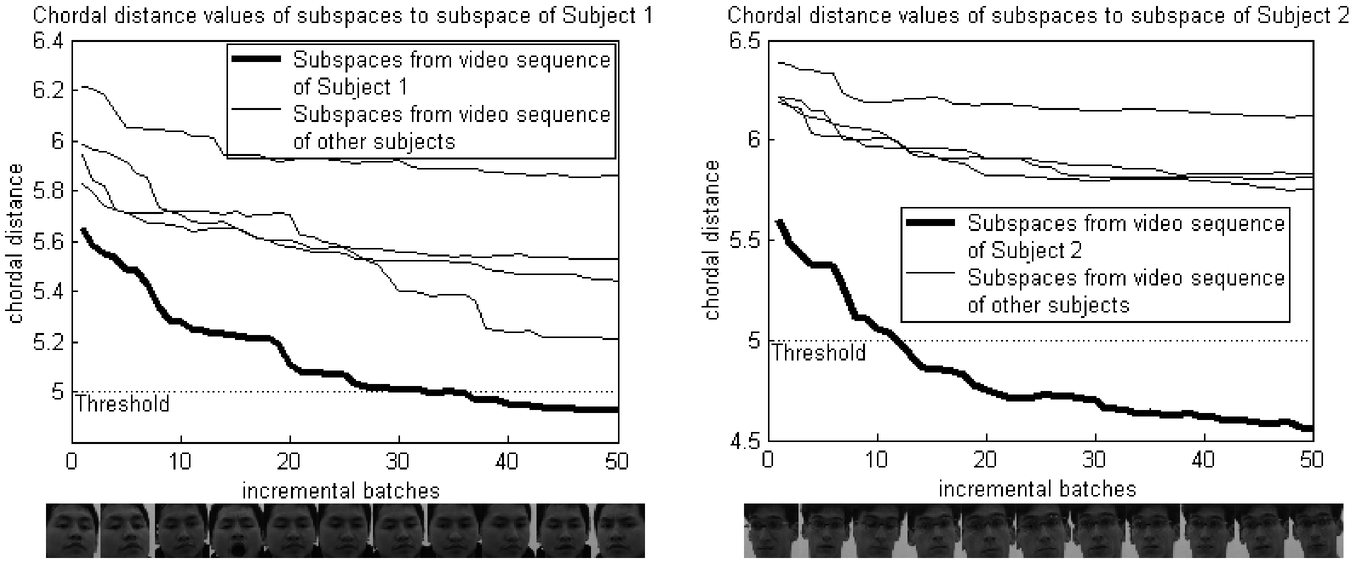

Figure 3. Distance comparisons of test subspaces updated incrementally against two reference subspaces. Images below graphs show key frames of video sequences of reference subjects (not seen before) used to update the test subspaces (corresponding to the bold lines in the graphs).

6 Experimental results

We collected a small *subspace database* of 9 people from our department with 3 sessions per subject. Our method differs from other face recognition-from-video approaches in that once we obtain a subspace representation for a person, we discard his/her video sequence. Furthermore, the subspaces of the subjects' faces were learnt *on-the-fly* as the images are being captured. We use a high-speed FireWire camera in a laboratory environment. Whilst we do not consider illumination effects yet at this stage, we did not actively constrain our illumination conditions. Seated subjects were requested to vary their pose and facial expression in a natural manner during the image acquisition. They were encouraged to talk and interact with other people in the surroundings while remaining seated. For examples of variances of face images produced, see Figures 1 and 3. Table 1 shows chordal distances between subspaces of 5 of the subjects within our database. Each subject has two subspaces in the Table 1, i.e. the reference subspace **Di** (trained earlier) and the subspace learned incrementally **Ei** (all from the same time instance), **i** being the subject number. It can be seen that the distances between subspaces of the same subject are relatively close compared to those of other subjects. From Table 1, we observe that there exists a suitable threshold value for classification. It is shown that our system managed to achieve perfect results for these select individuals. Evaluated on the overall database, we achieved an *Equal Error Rate* (EER) of 10% and an *Overall Accuracy* of 91%. Figure 3 shows how distance measures between subspaces progress with the arrival of new face images. It can be seen that as more information of a particular face is available, the estimated face subspace converges to an accurate position. This is indicated by the fact that the test subspace of a subject "moves" closer to the reference subspace of the same subject as more images are acquired and used to update the subspace. Figure 4 shows the program interface to our face recognition system. Presently we achieve a processing speed of approximately 12 frames per second (this includes *on-the-fly* subspace updating and distance calculation).

	E1	**E2**	**E3**	**E4**	**E5**
D1	**4.62**	5.73	5.77	5.51	5.63
D2	5.89	**4.69**	5.83	5.45	5.89
D3	5.75	5.84	**4.55**	5.81	6.12
D4	5.21	5.44	5.86	**4.93**	5.53
D5	5.89	5.97	6.10	5.71	**4.72**

Table 1. Comparison of chordal distances.

7 Conclusion and future work

We presented a framework for face recognition from video. We treated a video sequence of a face as a set of images, and performed face recognition by using linear subspace methods to match image sets. We applied incremental SVD computation procedures to update our subspace represen-

Algorithm 1 Pseudo code for subspace estimation and updating based on the theory outlined in Section 4.

Find initial n **faces:**
create empty matrix $\mathbf{A}$ of size $pixels \times n$
$face \longleftarrow 0$
while $face < n$ **do**
  perform face detection
  if face detected **then**
    pre-process sub-image
    raster scan and append to $\mathbf{A}$
    $face \longleftarrow face + 1$
  end if
end while
perform SVD on $\mathbf{A}$
retain first r left singular vectors to form $\mathbf{U^r}$
retain first r singular values to form $\mathbf{S^r}$

Find subsequent faces:
for $frame = 0$ to $maxFrames$ **do**
  create empty matrix $\mathbf{C}$ of size $pixels \times incr$
  $face \longleftarrow 0$
  while $face < incr$ **do**
    perform face detection
    if face detected **then**
      pre-process sub-image
      raster scan and append to matrix $\mathbf{C}$
      $face \longleftarrow face + 1$
    end if
  end while
  find $\mathbf{L}, \mathbf{J}$ and $\mathbf{K}$ matrices using $\mathbf{C}$
  construct $\mathbf{M}$ from $\mathbf{S^r}, \mathbf{L}$, and $\mathbf{K}$
  perform SVD on $\mathbf{M}$ i.e. $\mathbf{M} = \mathbf{U_m S_m} (\mathbf{V_m})^{\mathbf{T}}$
  compute $\mathbf{U}'' = [\mathbf{U^r}\ \mathbf{J}]\,\mathbf{U_m}$
  update $\mathbf{U^r}$ by performing $\mathbf{U^r} = \mathbf{U}''(:, \mathbf{1}:\mathbf{r})$
  update $\mathbf{S^r}$ by performing $\mathbf{S^r} = \mathbf{S_m}(\mathbf{1}:\mathbf{r}, \mathbf{1}:\mathbf{r})$
  find chordal distance of $span\{\mathbf{U^r}\}$ to gallery subspaces
end for

tation *on-the-fly* as new images are being acquired. For face classification, we introduced the *chordal distance* metric to quantify the distance between test and reference subspaces. Based on the framework, we implemented a face-recognition-from-video system that is capable of performing online training and classification. By incrementally updating the subspace representations, we are able to perform online training and classification whilst maintaining robustness against pose variation. Our preliminary experimental results demonstrates the effectiveness of our approach. By evaluating the proposed system on our own database, we achieved recognition rates which is on par with previous methods based on the same underlying principles (for example, see [26, 9]), suggesting that the proposed system

Figure 4. Examples of program interface to our system.

maintained the effectiveness of linear subspace based methods while capable of obtaining high speeds.

Future work includes improving the speed of our system and evaluating it more substantially. A more sophisticated background subtraction algorithm can be applied to our system so that it can function on an arbitrary background. Experiments on faces under changing lighting effects should be done in order to gauge the robustness of our method against illumination variations. Though our system performed decently, the accuracy obtained still falls short of expected levels for real-life deployment or commercialization. Kernel-based subspace methods like [25, 21] have been known to be able to improve tremendously the accuracy of face recognition by matching image sets. The challenge is to perform the computations for the kernel-based methods in an online manner so that we can arrive at a system that can learn and distinguish face representations rapidly.

References

[1] O. Arandjelovic, G. Shakhnarovich, J. Fisher, R. Cipolla, and T. Darrell. Face recognition with image sets using manifold density divergence. In *IEEE Conf. on Computer Vision and Pattern Recognition*, 2005.

[2] R. Basri and D. Jacobs. Lambertian reflectance and linear subspaces. *IEEE Transactions of Pattern Analysis and Machine Intelligence*, 25(2):218–223, 2003.

[3] P. Belhumeur and D. Kriegman. What is the set of images of an object under all possible illumination conditions? *International Journal of Computer Vision*, 28(3):245–260, 1998.

[4] G. Bradski, A. Kaehler, and V. Pisarevsky. Learning-based computer vision with Intel's open source computer vision library. *Intel Technology Journal*, 9(2):119–130, 2005.

[5] M. Brand. Incremental singular value decomposition of uncertain data with missing value. In *European Conference on Computer Vision*, 2002.

[6] S. Chandrasekaran, B. Manjunath, Y. Wang, J. Winkeler, and H. Zhang. An eigenspace update algorithm for image processing. *Graphical Models and Image Processing*, 59(5):321–332, 1997.

[7] R. Chellappa and S. Zhou. *Handbook of face recognition*, chapter 9: Face tracking and recognition from video. Springer, 2005.

[8] R. Epstein, P. Hallinan, and A. Yuille. 52 eigenimages suffice: An empirical investigation of low-dimensional lighting models. In *IEEE Workshop on Physics-Based Vision*, 1995.

[9] K. Fukui and O. Yamaguchi. Face recognition using multiviewpoint patterns for robot vision. In *10th International Symposium of Robotics Research*, 2003.

[10] G. H. Golub and C. F. V. Loan. *Matrix Computations*. John Hopkins University Press, 3rd edition, 1996.

[11] P. Hallinan. A low-dimensional representation of human faces for arbitrary lighting conditions. In *IEEE Conference on Computer Vision and Pattern Recognition*, 1994.

[12] S.-W. Kim and B. Oommen. On utilizing search methods to select subspace dimensions for kernel-based nonlinear subspace classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(1):136–141, 2005.

[13] T.-K. Kim, O. Arandjelovic, and R. Cipolla. Learning over sets using boosted manifold principal angles (BoMPA). In *British Machine Vision Conference*, 2005.

[14] T. Kozakaya and H. Nakai. Development of a face recognition system on an image processing LSI chip. In *IEEE Conference on Computer Vision and Pattern Recognition Workshop*, 2004.

[15] K.-C. Lee, J. Ho, M.-H. Yang, and D. Kriegman. Video-based face recognition using probabilistic appearance manifolds. In *IEEE Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 313–320, 2003.

[16] K.-C. Lee and D. Kriegman. Online learning of probabilistic appearance manifolds for video-based recognition and tracking. In *IEEE Conf. on Computer Vision and Pattern Recognition*, 2005.

[17] A. Levy and M. Lindenbaum. Sequential Karhunen-Loeve basis extraction and its application to images. Technical Report CIS9809, Technion, 1998.

[18] B. Li and R. Chellappa. A generic approach to simultaneous tracking and verification in video. *IEEE Trasactions on Image Processing*, 11:530–544, 2002.

[19] S. Li and Z. Zhang. Floatboost learning and statistical face detection. *IEEE Pattern Analysis and Machine Intelligence*, 26(9):1–12, 2004.

[20] R. Lienhart and J. Maydt. An extended set of haar-like features for rapid object detection. In *IEEE International Conference on Image Processing*, volume 1, pages 900–903, 2002.

[21] H. Sakano, N. Mukawa, and T. Nakamura. Kernel mutual subspace method and its applications for object recognition. *Electronics and Communications in Japan*, 88(6), 2005.

[22] G. Shakhnarovich, J. Fisher, and T. Darrell. Face recognition from long term obsevations. In *IEEE European Conf. on Computer Vision*, 2002.

[23] J. Tenenbaum and W. Freeman. Separating style and content with bilinear models. *Neural Computation*, 12:1247–1283, 2000.

[24] 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.

[25] L. Wolf and A. Shashua. Kernel principal angles for classification machines with applications to image sequence interpretation. In *CVPR*, volume 1, pages 635–640, 2003.

[26] O. Yamaguchi, K. Fukui, and K. Maeda. Face recognition using temporal image sequence. In *International Conference on Automatic Face and Gesture Recognition*, pages 318–323, 1998.

[27] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld. Face recognition: a literature survey. *ACM Computing Surveys*, 35(4):399–458, 2003.

[28] S. Zhou, R. Chellappa, and B. Moghaddam. Visual tracking and recognition using appearance-based modeling in particle filters. *IEEE Transactions on Image Processing*, 13(11):1491–1506, 2004.

[29] S. Zhou, V. Krueger, and R. Chellappa. Probabilistic recognition of human faces from video. *Computer Vision and Image Understanding*, 91:214–245, 2003.