PROBABILISTIC FACE RECOGNITION FROM COMPRESSED IMAGERY

Jian Li, and Shaohua Kevin Zhou

Center for Automation Research and Department of ECE
University of Maryland, College Park, MD, 20742
{lij,shaohua}@cfar.umd.edu

ABSTRACT

The effects of image and video compression on face recognition in the still-to-video setting are studied in this paper. We use the probabilistic framework described in [8], which solves tracking and recognition problems simultaneously via sequential importance sampling (SIS)[2]. To account for the illumination and pose variations in test sequences, intrapersonal space (IPS) [5] is constructed from examplar views and used to calculate the likelihood density. Both the gallery images and probe videos are compressed and several experiments are run to study their effects on the recognition rate. Some useful conclusions are drawn from the analysis of the experimental results, which will be helpful for future research on the interaction between recognition and compression. Meanwhile, the experiments also demonstrate the robustness of the proposed methods.

1. INTRODUCTION

As the applications of face recognition system are becoming more pervasive, video-based recognition methods are receiving greater attention in recent years [7]. Most of them deal with the still-to-video setting, meaning still images are used as gallery and video segments are used as probes. Recently Zhou *et. al.* [8] proposed a probabilistic framework to solve the problem of simultaneous tracking and recognition of human faces in video.

While researchers have studied the effects of illumination, pose and expression variations on the recognition rate, to our knowledge, little has been done to investigate the effect of compression in the still-to-video setting. Since most of the videos have to be stored after compression, analysis of the effect of compression becomes very necessary.

Some work has been done in the literature regarding the effects of compression in the still-to-still setting [1], in which case both the gallery and probe sets contain only still images. In FRVT 2000 [1], the effects of JPEG compression on face recognition have been tested and only compression over the probe set has been considered. It was concluded that the recognition rate does not necessarily go down with increasing compression ratio. In their experiment, the recognition rate goes up slightly for compression ratios 10:1 and 20:1. For compression ratio 40:1, the recognition rate goes below that of the uncompressed test.

In this paper, we consider the still-to-video scenario. Unlike FRVT 2000 , we will compress both the gallery and the probe sets to give a more comprehensive study. The rest of the paper is organized as follows. In section 2 we summarize the two state-of-art methods in face recognition literature which are used in our experiments. In section 3 we describe the experimental setting. Section 4 presents the experimental results and analysis. Section 5 concludes the paper and discusses future research directions.

2. THEORETICAL BACKGROUND

2.1. Probabilistic Framework

Before SIS was used in face recognition, most video-based methods relied on a good selection of frames and do not fully exploit the temporal information in the probe video. Zhou *et. al.*'s method [8] successfully used SIS to propagate the posterior density of the identity and motion variables to solve the tracking and recognition problems simultaneously. We choose to use this scheme in our experiments.

To be more specific, a time series state space model is used. The state variable $x_t = \{n_t, \theta_t\}$ includes an identity variable n_t and a motion parameters θ_t, which is assumed to be a 2D affine transformation. The system equation can be written as

$$n_t = n_{t-1}, \quad \theta_t = \theta_{t-1} + u_t, t \geq 1, \tag{1}$$

where we assume that motion variable follows a Markov process with u_t as a white Gaussian noise process.

A simple formulation of the observation equation can be characterized as

$$T_{\theta_t}(z_t) = I_{n_t} + v_t, \tag{2}$$

where, z_t is the observation, and T is an affine transform to normalize the image to the same size of the gallery images. $\{I_1, ..., I_N\}$ is the gallery set with one template per

Partially funded by the ARDA/VACE program under the contract MDA9040-0C-2110.

person and the total number of people is N, and v_t is observation noise. This equation is used to determine the likelihood function $p(z_t|n_t, \theta_t)$.

The recognition is based on a *Maximum A Posteriori* (MAP) decision rule, namely finding n_t that maximizes $p(n_t|z_{1:t})$. SIS is used to approximate and propagate the posterior probability $p(n_t, \theta_t|z_{1:t})$, and marginalization over variable θ_t is carried out before applying the recognition rule. Detailed descriptions can be found in [8].

A simple way to define the likelihood function from (2) is to set it as a 'truncated' Laplacian:

$$p(z_t|n_t, \theta_t) = Laplacian(\|T_{\theta_t}(z_t) - I_{n_t}\|; \sigma_1, \tau_1), \quad (3)$$

where σ_1, τ_1 are some parameters that can be chosen heuristically and the truncated Laplacian function is defined as

$$Laplacian(x; \sigma, \tau) = \begin{cases} \sigma^{-1} exp(-x/\sigma), & if\ x \leq \tau\sigma \\ \sigma^{-1} exp(-\tau), & otherwise \end{cases} \quad (4)$$

The drawback of this simple model is that it fails to capture the variations of the facial imagery due to different variations such as illumination and pose. An updated model is explained in the next subsection.

2.2. Intra-Personal Space and Probabilistic PCA

We applied the methods proposed by Modghaddam [5] to model the variations in illumination and pose etc. in our experiments. The *Intra-Personal Space* refers to the manifold containing the variations of face images belonging to one person. Probabilistic PCA [6] (PPCA) is used in our experiment to construct a probabilistic subspace density upon IPS.

To build IPS, five facial images are cropped from the video for each person. There is no overlap between images from test sequences and the frames from which training face images are cropped. Examples of enlarged gallery are show in Fig. 2.

PPCA assumes that an observation x is generated as follows,

$$x = u + Wy + e, \quad (5)$$

where we have the sample vector x, the hidden variable $y \sim N(0, I)$, the sample mean u, the *loading matrix* W and the measurement noise $e \sim N(0, \sigma^2 I)$. In our problem, we assume that the intrapersonal variations of a face image have the distribution as x in (5). So correspondingly, x is $T_{\theta_t}(z_t)$ and u is I_{n_t} as in (2). From the training images, we can estimate W and σ^2 using a maximum-likelihood (ML) estimating procedure. In practice, we use an approximation to the optimal solution, which will choose W to be the properly weighted principle eigen-vectors of the scatter matrix of the training samples and σ^2 the average of the remaining eigenvalues [6][5]. This provides us the probability of a sample x lying in that subspace as

$$PS(x) = c * exp(-\frac{1}{2}\sum_{i=1}^{d}\frac{y_i^2}{\lambda_i})exp(-\frac{\epsilon^2}{2\rho}), \quad (6)$$

where c is the normalizing constant, λ_i is the eigenvalue of the scatter matrix in the descending order, ρ is the average of the remaining $n - d$ eigenvalues, y_i is i-th component of the representation of x in the d dimensional principle subspace, and ϵ is the representation error.

Thus the likelihood function under this formulation is

$$p(z_t|n_t, \theta_t) = PS(T_{\theta_t}(z_t) - I_{n_t}), \quad (7)$$

where in this case I_{n_t} will be the mean of the samples corresponding to class n_t.

3. EXPERIMENTAL SETTING

We used the outdoor NIST data set as our test sequences. It contains a database of thirty persons and originally each person has one face image in the gallery. We extended the gallery as discussed in the last section to model the variations of the face image and construct an IPS. The sample gallery is shown in Fig. 1 and Fig. 2. The sample of compressed probe video is shown in Fig. 4. This data set is particularly suitable for evaluating the effects of compression because of the difficulty of outdoor sequences and the smaller size of face region.

Fig. 1. Original gallery of face images and its compressed version. The image size is 48 by 42. The compression artifacts are quite obvious.

In order to test the effects of compression, both the gallery and the probe video are compressed and several tests are run to compute the recognition rates in different cases. For the gallery set, we have two cases:

Fig. 2. Extended gallery and its compressed version. These images are cropped from the video frames ensuring no overlap with the test sequences used in recognition.

Fig. 3. The eigenvectors of the Intra-Personal Space. The first row is the first ten principle eigenvector for IPS constructed under UG. The second row is those constructed under CG. The blocking artifacts are also reflected.

1. Uncompressed gallery(UG).

2. Compressed gallery(CG). The images are compressed by choosing JPEG quality 5 with the ultimate compression factor as 1:8. Note that this compression is done over the already down-sampled face image.

For the probe video, the MPEG-2 [4] standard is used to compress the video into different levels [3]. We consider three cases:

1. (UV) Probe video set is uncompressed.

2. (MV) Probe video set is moderately compressed. It is illustrated in Fig. 4. The approximate frame rate is 25KB/s for gray images of size 720x480. The default quantization matrix in MPEG2 is used for compression. The resulting video does not have much visual difference compared to the original video.

3. (HV) Probe video set is highly compressed. The approximate frame rate is 5KB/s for gray images of size 720x480. The target frame rate is set to be very low. The resulting video in size is about 1/5 of MV and has apparent blur and blocking artifacts. It is illustrated in Fig. 5. The video samples are available at *http://www.cfar.umd.edu/∼lij* .

In total, we have 6 different recognition scenarios.

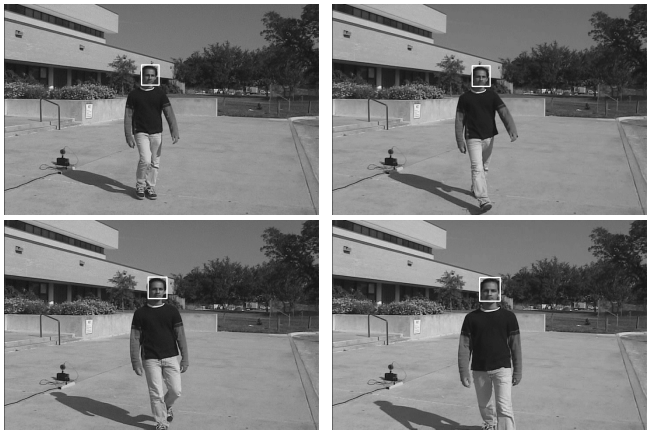

Fig. 4. Sample of the moderately compressed probe video. Tracking results under CG are illustrated by the white rectangle.

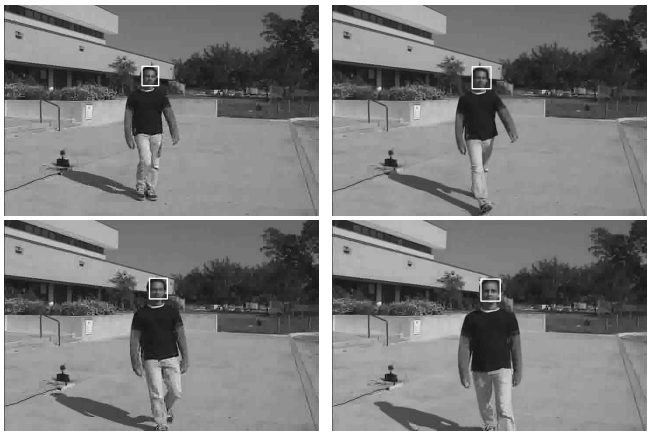

Fig. 5. Sample of highly compressed probe video with tracking results under CG. Size is about 1/5 of the moderately compressed video file, at a frame rate of 5KB/s.

4. EXPERIMENTAL RESULTS AND ANALYSIS

We have tested the tracking and recognition algorithm as described in Section 2 for different degrees of compression. The first ten eigenvectors for IPS constructed under UG and CG are illustrated in Fig. 3. The blocking artifacts under CG are reflected in those eigen-vectors. The tracking results under MV and HV are shown in Fig. 4 and 5. The Cumulative Match Curves (CMC)[7] are shown in Fig. 6. Comparison under each different settings has been illustrated in the figure. The recognition rates based on both the top one match and top three matches are shown in Fig. 7.

From the recognition results, we observe the following:

1. Similar to the conclusion in FRVT 2000 [1], which only considered the still-to-still setting, the recognition rate does not necessarily go down with increasing

compression in the still-to-video setting either. The recognition rates based on the top one match for MV are higher than those for UV when using either UG or CG. It indicates that moderate compression, which can smooth the noise, is helpful for recognition.

2. Using very poor quality gallery set leads to a drop in the recognition rate and its effect on UV is less than its effect on the compressed video. The visual quality of the compressed gallery used in our experiment apparently degrades to a very noticeable level(see Fig. 1 and 2). The recognition results under CG, compared with those under UG, drop in all instances as shown in Fig. 6 and 7 compared with UG. The percentage drops for the top one match for UV, MV and HV because of CG are 3.4, 6.6 and 6.7 respectively. We see that the compressed gallery has less effect on UV than on MV or HV in our experiment. It indicates that video compression and image compression may not necessarily retain the same information for classification even if they have similar compression factor. That will lead us to use an alternate approach to model the variation of compression in video instead of merely learning it from compressed still image gallery.

3. Recognition for HV drops compared to MV. It indicates that for a recognition system using the still-to-video setting, moderate compression without significant degradation of image quality will be appropriate if we need to compress the video.

4. The closeness of all the curves in Fig. 6 shows that the methods we use are fairly robust to compression in all different cases.

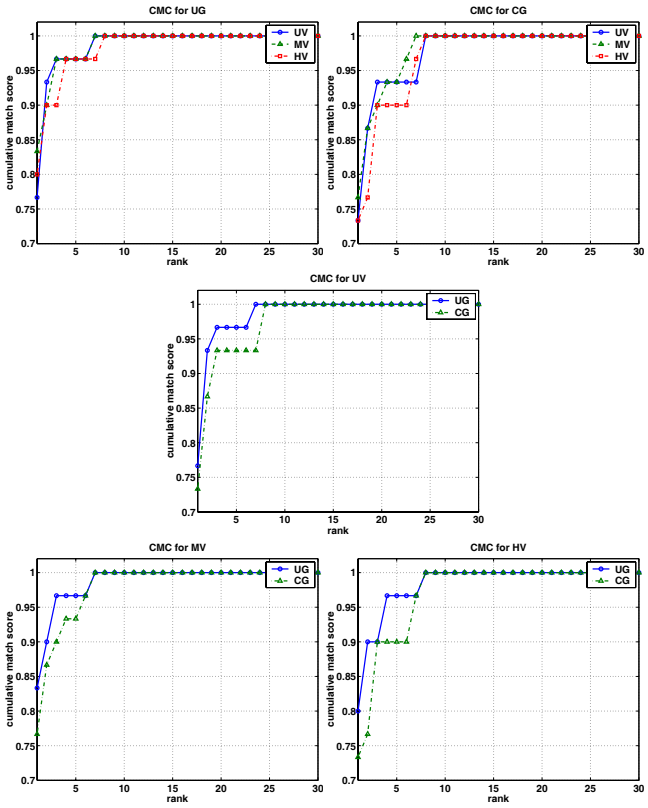

Fig. 6. Comparison of the CMC under different cases.

	Top one Recog. Rate(%)			Top 3 Recog. Rate(%)		
	UG	CG	Avg	UG	CG	Avg
Uncomp. Video	76.7	73.3	75	96.7	93.3	95
Mod. Compr. Video	83.3	76.7	80	96.7	90	93.35
High. Compr. Video	80	73.3	76.65	90	90	90
Average	80	74.4	77.2	94.5	91.1	92.8

Fig. 7. A summary of recognition rates.

5. CONCLUSION AND FUTURE PLAN

The effects of compression of both still images (used in the gallery) and videos (used in the probe sets) are tested by running several experiments under the still-to-video setting. Our future plan will consider utilizing the information inherent in the compressed imagery, such as motion vectors, to make tracking and recognition more efficient and accurate. Larger data set will also be tested to obtain a more general analysis.

6. REFERENCES

[1] D. Blackburn, J. Bone, and P. Phillips. FRVT 2000 evaluation report. *Website: http://www.frvt.org*, 2001.

[2] A. Doucet, N. d. Freitas, and N. Gordon. *Sequential Monte Carlo Methods in Practice*. Springer, 2001.

[3] S. Eckart and C. Fogg. Mpeg2 encoder / decoder, version 1.2. *MPEG Software Simulation Group, website http://www.mpeg.org/MSSG/*, 1996.

[4] J. Mitchell, W. Pennebaker, C. Fogg, and D. LeGall. *MPEG video compression standard*. Chapman and Hall, 1996.

[5] B. Moghaddam. Principal manifolds and probabilistic subspaces for visual recognition. *IEEE Trans. PAMI*, 24(6):780–788, 2002.

[6] M. E. Tipping and C. M. Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society, Series B*, 21(3):611–622, 1999.

[7] W. Y. Zhao, R. Chellappa, A. Rosenfeld, and P. J. Phillips. Face recognition: A literature survey. *Accepted by ACM Computing Surveys*, 2003.

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