Assessment of H.264 Video Compression on Automated Face Recognition Performance in Surveillance and Mobile Video Scenarios

Brendan Klare[a] and Mark Burge[b]

[a]Department of Computer Science and Engineering
Michigan State University
East Lansing, MI
[b] Noblis, Falls Church, VA

ABSTRACT

We assess the impact of the H.264 video codec on the match performance of automated face recognition in surveillance and mobile video applications. A set of two hundred access control (90 pixel inter-pupilary distance) and distance surveillance (45 pixel inter-pupilary distance) videos taken under non-ideal imaging and facial recognition (e.g., pose, illumination, and expression) conditions were matched using two commercial face recognition engines in the studies. The first study evaluated automated face recognition performance on access control and distance surveillance videos at CIF and VGA resolutions using the H.264 baseline profile at nine bitrates rates ranging from 8kbs to 2048kbs. In our experiments, video signals were able to be compressed up to 128kbs before a significant drop face recognition performance occurred. The second study evaluated automated face recognition on mobile devices at QCIF, iPhone, and Android resolutions for each of the H.264 PDA profiles. Rank one match performance, cumulative match scores, and failure to enroll rates are reported.

Keywords: Face recognition, video compression, surveillance, mobile

1. INTRODUCTION

Video, be it from surveillance cameras or other people's video (OPV), is often the ultimate source of the images used in both automated and forensic facial recognition. In many scenarios, such as 24-hour surveillance, video is constantly being acquired and must be compressed for transmission and storage. While the impact of video compression on human visual perception has been well researched,[1] to our knowledge the impact of modern video codecs on automated face recognition has not been analyzed. We present the results of an analysis of the impact of the H.264 video compression algorithm on face recognition performance.

The design of an effective video surveillance system for facial recognition must consider both system (e.g., camera location, illumination) and subject (e.g., pose, expression, cooperation) factors. While systematic studies evaluating the impact of factors such as pose, illumination, and expression on automated facial recognition systems have been in place for over a decade,[2,3] the impact of video transmission and storage (e.g., compression codec, image resolution, acquisition frame rate, and encoding bitrate) has not been as thoroughly assessed.

In a typical video surveillance system, face recognition does not take place on the camera. Instead, the video is encoded and transmitted to a remote location for either automated processing and storage or human examination. The bandwidth requirements for raw video transmission necessitate its compression before transmission. This paper addresses the trade-offs between compression and face recognition performance when using standard video codecs,[4] camera resolutions, and commercial off-the-shelf facial recognition software.

Further author information:
B.K.: E-mail: klarebre@msu.edu,
M.B.: E-mail: mburge@mitre.org
This research was conducted while the authors were at the Center for National Security and Intelligence, Noblis, Falls Church, VA.

Figure 1. Gallery contained a single image for each of the 100 subjects. Images were acquired under non-ideal imaging and facial recognition (e.g., pose, illumination, and expression) conditions.[10]

Previous research has examined the impact of JPEG image compression on the performance of face recognition algorithms. The impact of image compression on template size is an important design factor, especially in respect to the management of large galleries[5] and the embedding of digital templates within travel documents[8] such as smartcards. A comprehensive evaluation of the impact of JPEG and JPEG2000 image compression[5,6] claims that facial images can be compressed by at least an order of magnitude without significant degradation in automated face recognition performance. In addition, research into facial recognition algorithms specifically designed for compressed images[7] has been conducted.

The remainder of the paper is organized as follows: Section 2 discusses the H.264 video codec, Section 3 describes our experimental design, Section 4 presents the results, and Section 5 summarizes our analysis and future research directions.

2. VIDEO COMPRESSION

The Common Intermediate Format (CIF)[11] specifies standard pixel-based resolutions for encoding video signals. Of these formats, the mostly widely adopted for video surveillance are: QCIF (176x144), CIF (352x288), 4CIF (704x576), and VGA (640x480) formats. At a frame rate of approximately 30 Frames Per Second (fps), one hour of uncompressed VGA resolution video requires either on the order of 25 Gigabytes of storage or more than 50 megabit per second (Mbit/s) to transmit[*] For this reason video signals are commonly compressed before transmission or storage.

Legacy surveillance video systems typically encode video for transmission and storage using either MJPEG or the MPEG-4 Part 2 Advanced Simple Profile. MJPEG provides high quality video at the expense of high bitrate requirements by using JPEG to encode each frame independently (i.e., only intra-frame encoding which does not take advantage of the temporal redundancy between frames). MPEG-4 Part 2 Advanced Simple Profile is an inter-frame encoder which builds upon the earlier H.263 codec that was designed to compress CIF and QCIF video signals for low-bitrate transmission.

H.264 was designed to support a broad gamut of video compression scenarios ranging from real-time, low-bitrate surveillance to studio-quality broadcasts. In order to support this wide range of scenarios a set of standard profiles were developed defining the minimum subset of the codec that compliant encoder and decoder implementations would need to support. Each profile was further refined by a set of levels which defined additional constraints (e.g., bitrate, resolution) that a profile compliant to a given level must support.

Previous to the development of H.264, most video compression codecs could be modelled by four steps: prediction, transformation, quantization, and entropy coding. In addition to offering improvements on each of these steps, H.264 adds an additional filtering step to the decoding stage. The filtering step mitigates the jagged edges that result from image and video codecs (e.g., JPEG, MPEG) that independently transform and quantize 16x16 luminance and 8x8 chroma macroblocks.

[*] This does not account for the format specific encoding overhead and the 4:2:0 quantization of the YC_bC_r encoded color information.

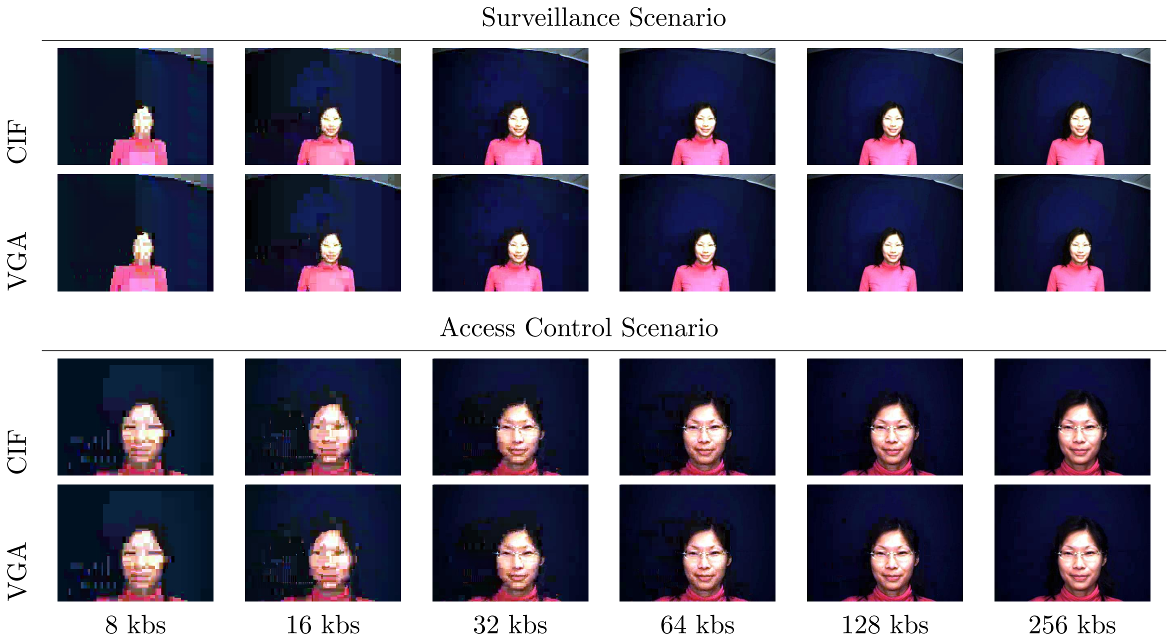


Figure 2. Examples of the impact of bitrate on the visual quality of H.264 Baseline profile compressed video for each scenario and resolution.

The impact these compression artifacts can have on on the match performance of various image-based biometrics[5,6] has been evaluated and attempts at mitigation ranging from simple post-processing blurring to customized codecs[12] have been developed. In the case of H.264, such mitigation is now provided automatically by the decode-time filtering stage of the codec.

Specifically, the H.264 algorithm[13] uses an n-tap discrete wavelet transform, where n ranges from 3 to 5, to smooth the boundaries between 4x4 pixel sub-blocks of the macroblocks. The decision to smooth as well as the degree of smoothing is based on a combination of spatial and temporal factors (e.g., quantization levels and gradient directions of adjacent blocks) as well as user-selected codec quality parameters. It is important to note that H.264's filtering algorithm was designed to improve human perceptual quality, which does not necessarily correlate with an improvement in automated facial recognition performance. As this study does not isolate the impact of this filtering stage, we will assess its impact on perceptual video quality and automated face recognition performance in a future study.

3. EXPERIMENTAL DESIGN

The following experiments utilized two COTS (Commercial of the Shelf) face recognition systems, an open source implementation of the H.264[14] codec, and 100 subjects selected at random from the CMU Face In Action (FIA) database.[10] FIA is a collection of videos captured at VGA resolution of subjects performing actions similar to those at passport checkpoints.

Because no designated gallery images are provided in the FIA database, for each subject, we manually selected a single image to use in the gallery from a separate acquisition session (i.e., FIA, indoor session 2, camera 3, 640x480, no compression) then those used for testing. Figure 1 shows examples of the gallery images.

Face recognition performance was computed using probe images from cameras 3 and 5 of the FIA indoor session 1. Camera 3 is in close proximity to the user with an average of 90 pixel inter-pupilary distance (IPD) and forms the basis for our *Access Control Scenario*. Camera 5 captures the user at a further distance for an average 45 pixel IPD and forms the basis of our *Surveillance Scenario*.

Probe frames were selected by generating match scores between each uncompressed image frame in each test video and the their corresponding gallery image. For each subject, 10 probe frames were selected in decreasing match score order with the restriction that each frame must be at least 5 frames from a previously selected frame. This resulted in 1000 probe images, with 10 from each of the 100 subjects. These individual frames represent a selection of the best candidate frames for face recognition for each subject prior to image compression. These same frames were subsequently used to evaluate the impact of video compression on face recognition performance.

Using the selected probe images and enrolled gallery we ran two separate experiments. The first experiment evaluated the face recognition performance with respect to the kilobits per second (kbs) bandwidth. This experiments was designed to offer insight into configuration trade-offs for (1) surveillance systems used to store video data for later forensic analysis and (2) those which transmit video for live face recognition. The second experiment used automated face recognition as an indicator of how video compression for transmission to PDA devices can impact the ability of a user to identify individuals in the video transmitted to their handheld device.

Two separate face recognition engines were used in our experiments, however we are unable to disclose which face recognition engines we used. This should not diminish the findings in this paper because: (1) both face recognition engines are commercial face recognition engines used by leading law enforcement and government organizations, (2) both engines were highly competitive participants in the most recent Face Recognition Vendor Test,[2] (3) two separate matchers are used to corroborate the results.

3.1 H.264 Compression for Surveillance

Using the Baseline H.264 compression profile, we varied the compression bitrate of the videos containing the selected probe frames. We considered each video in the standard CIF and VGA resolution formats. After compressing each video at various bandwidths the probe frames were then be extracted and used for face recognition. The intent of this experiment was to determine the minimal bandwidth requirements that still permit face recognition results comparable to those using frames from the full uncompressed signal. Figure 2 shows the visual impact of decreasing bitrates.

Finding the proper format and compression rate for surveillance videos with respect to face recognition is of critical importance. This is illustrated with an example of a video surveillance system that would be employed to record anything from a convenience store to a section of an airport. Financial constraints will limit the amount of storage capacity of the DVRs (smaller systems) or file servers (larger systems) used to record the captured video. As the bandwidth of the video increases, the number of hours of video that can be stored decreases, negatively impacting the intent of the system. Conversely, as the bandwidth of the video decreases the quality of the video also decreases, which in turn lowers the ability to identify an individual using face recognition.

This results in a classic trade-off of quantity versus quality. If the relationship between the two were linear then one would simply choose a bandwidth based on a cost value analysis. However, if a non-linear relationship between bandwidth and face recognition performance exists then an optimal bandwidth range can be identified that offers both high levels of compression and similar levels of face recognition with respect to face recognition using the uncompressed images. Previous research on single image compression identified the presence of just such a non-linear relationship, resulting in decreased storage requirements for large face recognition systems.[5, 6]

3.2 H.264 Compression in Mobile Video Applications

Our second experiment examined the impact of the H.264 PDA compression profiles on face recognition performance. We tested the four H.264 PDA profiles in their designated QCIF format (192x144), as well as the Iphone and Android H.264 profiles using a their native screen resolution (424x318).

The motivation of this experimentation stems from the growing use of PDA and cell phone devices for transmission of videos. Using automated face recognition as a proxy for human face recognition, this experiment seeks to imply how successful a compressed surveillance video stream could be sent to a client on a portable device with the intent of the user being able to identify someone in the video. Though limited cellphone data bandwidth and smaller device screens increase the difficulty of face recognition, this experiment provides insight as to what degree face recognition is feasible on video streamed to portable devices.

3.3 Evaluation

A robust video compression algorithm with respect to face recognition is an algorithm that can significantly lower the amount of transmittable data without a noticeable degradation in face recognition performance. Thus, the distribution of true match scores and imposter match scores should maintain a similar level of separation between uncompressed and compressed probe data.

In order to compare two different compression levels a measure for the amount of compression used and the facial recognition performance is needed. Using these two measures a plot of the performance of facial recognition against the level of compression will offer a visual means of deciding which algorithm offers the best trade-off. Ideally this plot will indicate some non-linear relationship between the two criteria. A non-linear relationship will result in a "knee" in the curve that indicates an ideal operating range.

The choice of a compression criteria is kilobits per second (kbs), which is a standard choice when evaluating the impact of video compression. Face recognition performance is reported as: (1) the average rank one performance, and (2) the failure to enroll (FTE) rate. The average rank one performance indicates how the performance of face recognition changes with respect to the compression bandwidth. The FTE rate indicates when the image quality drops to a level that causes the face recognition engine to no longer be able to recognize the face or find eyes.

4. RESULTS AND ANALYSIS

Figure 3 contains the results for the H.264 surveillance compression experiment. For both cameras, the rank one performance demonstrates a clearly non-linear relationship between the bitrate and face recognition rates, resulting in a knee in the curve around the range of 128kbs to 512kbs. To put this level of compression into context: the FIA uncompressed 600 frame VGA videos are of size 540MB, and at 128kbs the video size are roughly 320KB, a compression ratio of roughly 1/1600. Single image compression studies indicated that levels of compression around one order of magnitude had little impact on face recognition performance,[6] and our studies indicate the same findings for an even higher degree of compression.

It is important to note that the results of Figure 3, should not be interpreted as the lower resolution CIF format offers improved recognition performance over the higher resolution VGA format – as face recognition performance is known to improve at higher resolutions when more pixels are available across the face.[2] Instead, the correct interpretation is that Since compressing VGA to 128kbs requires four times as much compression as when compressing CIF to 128kbs, the correct interpretation is that more information that is pertinent to face recognition performance is lost by increasing the amount of compression than is lost by lowering the image resolution.

In Figure 4 the failure to enroll rates for each matcher are shown as a function of the bitrate. These results indicate that video compression can be increased further if the surveillance application being used is person counting instead of face recognition. The higher FTE of Matcher 1 in the surveillance scenario resulted impacted the matcher's face recognition performance as shown in Figure 3.

Figure 5 presents the face recognition performance and failure to enroll rates for the H.264 PDA compression profiles in tabular form. For the access control scenario both the PDA profiles as well advanced iPhone and Android compression formats had similar performance, however this was not the case for the surveillance scenario. In the surveillance scenario both the iPhone and Android compression formats performed at the same order as the original uncompressed videos, while the PDA profiles had far less success. We previously observed that the lower resolution CIF format led to higher face recognition performance than the VGA format, but the resolution of the PDA profiles appear to move into a range that is too low for successful face recognition. Because modern mobile devices have larger screen resolutions and faster computational capabilities, we believe that future needs to transmit videos to remote users for successful identification will be met by improved devices.

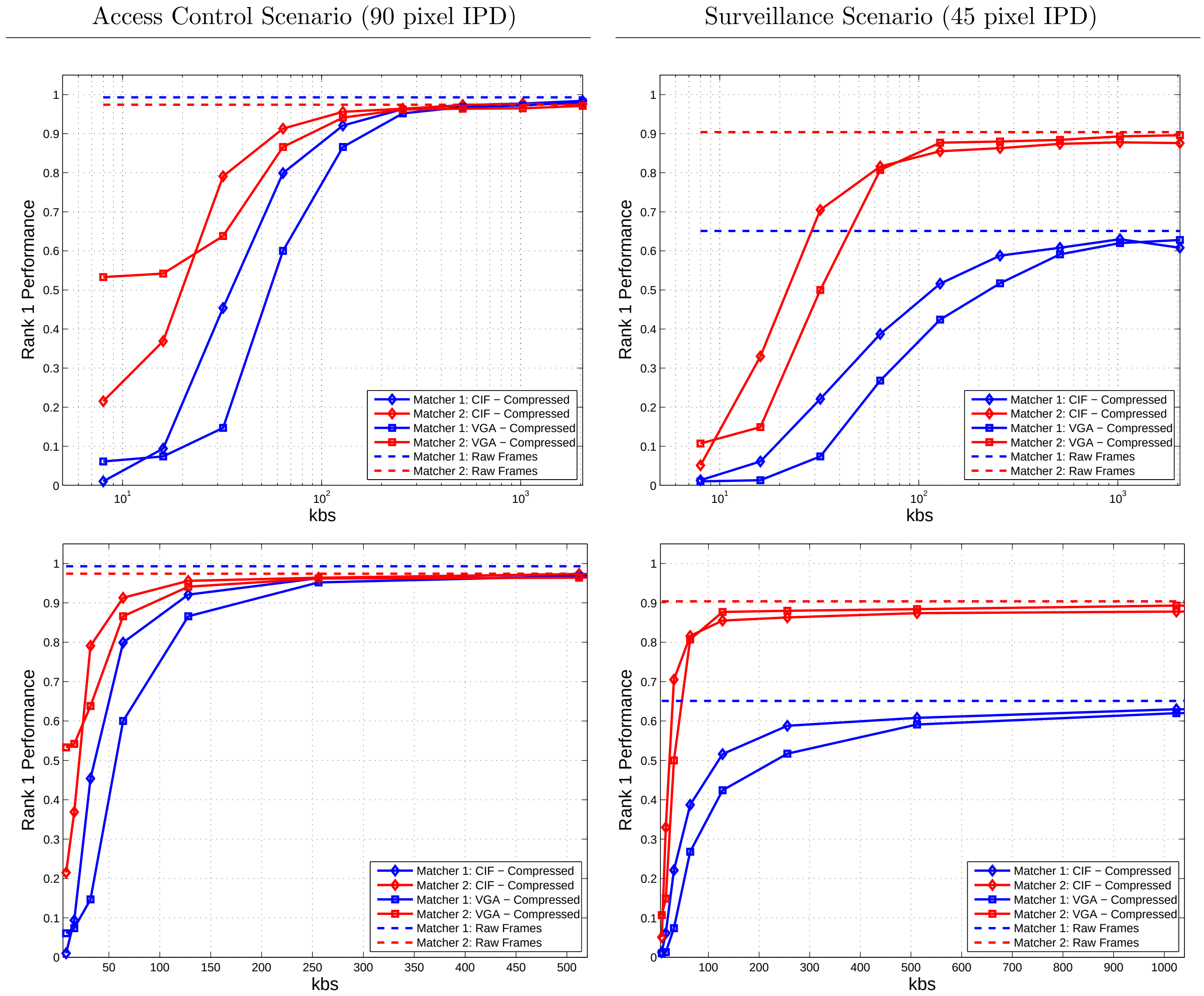


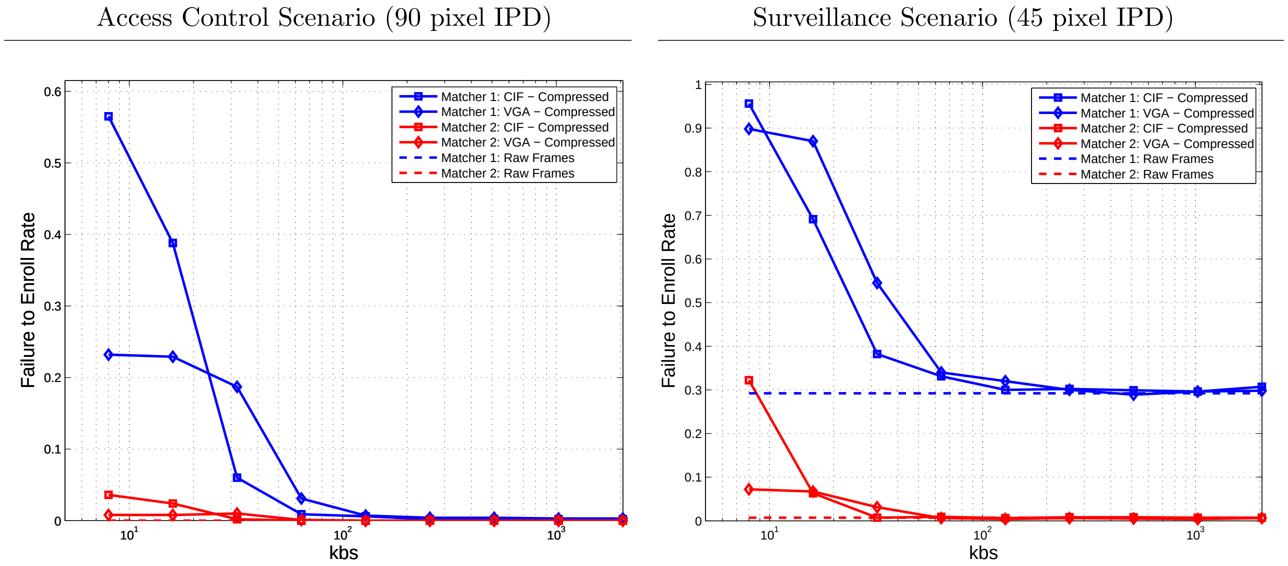
Figure 3. Impact of H.264 compression on face recognition match performance in access control (left) and surveillance (right) scenarios. The plots in row one show the Rank 1 recognition rates as a function of bitrate with a log axis. Row two shows the same results using a linear axis. The linear axis more clearly demonstrates the knee of the curve around the 128kbs bitrate.

5. CONCLUSIONS

We assessed the impact of the H.264 video codec on automated face recognition match performance in typical surveillance and mobile video applications. Non-ideal videos sequences from the FIA database[10] were used to construct experiments simulating typical access control and surveillance scenarios. Face recognition performance was measured using two of the top performing face matchers from the Face Recognition Grand Challenge.[2]

The first study evaluated automated face recognition performance on access control and distance surveillance videos at CIF and VGA resolutions using the H.264 baseline profile at nine bitrates rates ranging from 8kbs to 2048kbs. The study indicated that automated face recognition performance comparable to that of uncompressed video could be obtained at bitrates as low as 128kbs using the H.264 baseline profile at CIF and VGA resolutions. We concluded that a bitrate of 128kbs can provide an improvement of three orders of magnitude over the transmission and storage requirements of identically formatted, uncompressed video.

The second study evaluated automated face recognition performance on video streamed to mobile devices with QCIF, iPhone, and Android resolutions using each of the H.264 PDA profiles. The study found (1) that all


Access Control Scenario (90 pixel IPD)

Surveillance Scenario (45 pixel IPD)

Figure 4. Failure to Enroll (FTE) rates for each matcher in the access control (left) and surveillance (right) scenarios.

mobile device resolutions had acceptable face recognition performance across all of the PDA profiles for scenarios like access control, where the subject's face fills on the order of 2/3 of the frame; (2) that in scenarios where the subject is at a greater distance the performance rapidly degraded when using the profiles designed for older mobile devices.

For uncompressed VGA videos with subjects at a distance from the camera between 45 and 90 IPD, key findings and results can be further summarized as:

1. H.264 compression of up to 128kbs (1/1600th the original video size) offered face recognition performance on the same order as the raw, uncompressed videos
2. Lowering the image resolution to the CIF standard prior to compression offered higher face recognition performance on both matchers tested
3. Based on the failure to enroll rates, the surveillance application of person counting allows for far more video compression (down to 32kbs or 64kbs)
4. The improved screen resolutions and processors in modern mobile devices (iPhone and Android) permit videos to be successfully transmitted to such devices for face recognition

In our on-going research we are (1) using this dataset to evaluate the predictive power of video quality metrics[1] on face recognition performance; (2) developing novel no-reference and partial-reference video quality metrics that are predictive of face recognition performance; and (3) expanding this study to include additional video codecs and classes of face recognition algorithms.

REFERENCES

[1] Winkler, S. and Mohandas, P., "The evolution of video quality measurement: from PSNR to hybrid metrics," *Broadcasting, IEEE Transactions on* **54**(3), 660–668 (2008).

[2] Phillips, P. J., Scruggs, W. T., O'Toole, A. J., Flynn, P. J., Bowyer, K. W., Schott, C. L., and Sharpe, M., "Face Recognition Vendor Test 2006: FRVT 2006 and ICE 2006 large-scale results," in [*Tech. Report NISTIR 7408, NIST*], (2007).

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

	Access Control		Surveillance	
	Rank-1	FER	Rank-1	FER
Raw	0.993	0.000	0.650	0.292
PDA 1.0	0.877	0.006	0.215	0.446
PDA 1.1	0.953	0.006	0.262	0.443
PDA 1.2	0.954	0.004	0.261	0.447
PDA 1.3	0.953	0.005	0.264	0.444
Android	0.972	0.001	0.574	0.320
iPhone	0.973	0.004	0.586	0.288

Figure 5. Impact of H.264 PDA Profile compression on face recognition performance reported as rank one results and failure to enroll percentages.

[4] Hanzo, L., Cherriman, P., and Streit, J., [*Video compression and communications: from basics to H.261, H.263, H.264, MPEG4 for DVB and HSDPA-style adaptive turbo-transceivers*], ch. 1, Wiley-IEEE (2007).

[5] McGarry, D., Arndt, C., McCabe, S., and D'Amato, D., "Effects of compression and individual variability on face recognition performance," *Proceedings of SPIE* **5404**, 362–372 (2004).

[6] Delac, K., Grgic, M., and Grgic, S., [*Image compression effects in face recognition systems*], ch. 5, I-Tech (2007).

[7] Eickeler, S., Mueller, S., and Rigoll, G., "Recognition of JPEG compressed face images based on statistical methods," *Image and Vision Computing* **18**(4), 279–287 (2000).

[8] Abiantun, R., Savvides, M., and Vijaya Kumar, B., "How low can you go? Low resolution face recognition study using kernel correlation feature analysis on the FRGCv2 dataset," in [*Biometric Consortium Conference*], (2006).

[9] Bourlai, T., Kittler, J., and Messer, K., "JPEG compression effects on a smart card face verification system," in [*IAPR Conference on Machine Vision Applications*], (2005).

[10] Goh, R., Liu, L., Liu, X., and Chen, T., "The CMU Face In Action (FIA) database," in [*Proc. of Analysis and Modeling of Face and Gestures*], (2005).

[11] Rijkse, K., "H.263: video coding for low-bit-rate communication," *Communications Magazine, IEEE* **34**(12), 42–45 (1996).

[12] Bradley, J. N., Brislawn, C. M., and Hopper, T., "FBI wavelet/scalar quantization standard for gray-scale fingerprint image compression," *Visual Information Processing II* **1961**, 293–304 (1993).

[13] Richardson, I. E., [*H.264 and MPEG-4 video compression: video coding for next generation multimedia*], Wiley, 1 ed. (August 2003).

[14] Wiegand, T., Sullivan, G. J., Bjntegaard, G., and Luthra, A., "Overview of the H.264/AVC video coding standard," *Circuits and Systems for Video Technology, IEEE Transactions on* **13**(7), 560–576 (2003).