PRIVACY PROTECTING, INTELLIGIBILITY PRESERVING VIDEO SURVEILLANCE

Natacha Ruchaud, Jean-Luc Dugelay

PRIVACY PROTECTING, INTELLIGIBILITY PRESERVING VIDEO SURVEILLANCE

Natacha Ruchaud and Jean-Luc Dugelay

Eurecom, Sophia-Antipolis
450 route des Chappes, 06410 Biot
ruchaud@eurecom.fr, dugelay@eurecom.fr

ABSTRACT

Video surveillance is increasingly omni-present in our everyday life and is a key component of many security systems. Not only is the increasing number of cameras, but also the resolution of visual sensors and the performance of video processing algorithms. This evolution generates some important privacy concerns. This article introduces a new visual filter that includes a good trade-off between privacy and intelligibility. It ensures that people are unrecognizable while keeping the scene understandable in terms of events which allows machines to detect abnormal behavior. The algorithm operates in the DCT domain to be compliant with the popular JPEG and MPEG codecs. For each sensitive area of the picture (i.e. area where privacy needs to be protected), the proposed algorithm uses the low-frequency coefficients of the DCT to display a privacy preserved image of the region and the high-frequency coefficients to hide most of the original information. Finally, our process allows authorized users to nearly reverse the process thanks to the hidden information.

***Index Terms*—** Privacy protection, Surveillance, DCT domain

1. INTRODUCTION

There is an impressive growth in the adoption of video surveillance systems in public transport, airport, city center, or residential areas. For instance, the number of cameras in UK is about 5.9 million, including 750,000 in locations such as schools or hospitals. In parallel, we are seeing spectacular progresses in the field of automatic recognition (e.g., identity recognition). For example, in [1], authors report that results from its 2013 test of facial recognition algorithms demonstrate that accuracy has improved up to 30 percent within a period of three years. Moreover, the enhancement in image sensors (e.g., 4k, HD) contributes to increase performances attached to those identification techniques (e.g., a person can be recognized even far away from the camera). The powerful video analytic tools combined with pervasive networks of dense cameras highlight questions about privacy policy. One challenge of this article is to protect privacy with the aim of making compliant the monitoring in video surveillance that appear as two conflicting objectives.

The rest of the paper is organized as follows: in the next section, the literature survey as well as the criteria to assess the effectiveness of an anonymization method, are summarized. Our proposed approach is described in Section 3. Experimental results are presented and discussed in Section 4. Finally, we draw some conclusions and give an outlook for possible future works in Section 5.

2. BACKGROUND AND RELATED WORK

As listed in [13], an ideal privacy filter, in video surveillance, should match five criteria:
i) Privacy protection (e.g., no possibility to recognize people),
ii) Intelligibility, which means keeping a fair visual quality to be able to recognize events in the scene
iii) Reversibility, which is the possibility to re-obtain the original images or videos (available only for authorized people if necessary),
iv) Compression, compliant with the widely adopted JPEG [14] and MPEG [15] standards,
v) Real time processing which is crucial for surveillance.

Naive methods, like blurring, blacking out or pixelization that are already used to anonymize faces (e.g., Google Street View, ObscuraCam [1] on Android) are not reversible.

In [2], authors hide privacy information inside image itself and use an inpainting method to refill the privacy-sensitive regions. Despite the aesthetic of images, actions of people become invisible.

Sweeney [3] applied k-anonymity method and guarantees that each face image appears at least k times in the gallery, therefore confusing face recognition performances. In order to make the k-same obfuscation process reversible, authors in [4] employ a watermarking method. The difference between the original and obfuscated image is compressed, encrypted and embedded within the obfuscated image itself. These algorithms require to use a face library and can only preserve the identification of faces (not the entire body).

[1] https://guardianproject.info/apps/obscuracam/

Table 1: Summary of existing filters according to the five criteria that are summarized at the beginning of the section 2

Filter	Privacy	Intelligibility	Reversibility	Compression	Real Time
Pixelization/Blur	Yes	Yes	No	Yes	Yes
Blacking out	Yes	No	No	No	Yes
Inpainting[2]	Yes	No	Yes	Yes	No
K-anonymity [3, 4]	Yes	Yes	Yes	No	No
PICO[5, 6]	Yes	No	Yes	Yes	Yes
Scrambling [7]	Yes	No	Yes	No	No
Scrambling MPEG [8, 9, 10]	Yes	No	Yes	Yes	Yes
StegoScrambling [11, 12]	Yes	Yes	Yes	No	Yes

In privacy domain, scrambling encrypts data and is employed for its reversibility. The process of PICO (Privacy through Invertible Cryptographic Obscuration) [5, 6] combines cryptographic techniques with image processing to provide a solution to the critical issue of privacy invasion. Melle and Dugelay [7] introduced a method to scramble faces using background self-similarities. Authors in [8] and [9] presented scrambling techniques integrated into MPEG. Dufaux and Ebrahimi [8] proposed to pseudo-randomly invert the sign of selected transform coefficients in the privacy region for MPEG-4. In both, [8] and [9], the insertion of the scrambled step is done in the intra prediction frames of MPEG and in the encoder, unscrambled data are used in the motion-compensated (MC) prediction loop. In [10], authors prove that for all scrambling schemes that they have tested, even if motion vector information was not encrypted, the privacy is still protected. However, to have a very high level of privacy protection they propose to scramble the motion vector. Scrambling is a powerful reversible method to preserve privacy, but does not allow the viewers to recognize actions (especially when the whole body is protected), consequently, it might hamper the surveillance.

Our work is inspired from preliminary works published in [11] [12]. Authors combine scrambling and steganography techniques. This combination is applied in the spatial domain by shifting and saving pixel by pixel the Most Significant Bit (MSB) of a scrambled RoI (Region of Interest) to the Least Significant Bit (LSB). Then, the MSB of the resulting image are substituted by the bits of the edge associated with the RoI in order to keep the scene understandable. This privacy filter is not robust against manipulations such as compression. De facto, many applications can not then use that tool. In our proposed method, we apply the idea of keeping the scene understandable while hiding the most significant information towards the least significant information. Nevertheless, we operate in the DCT domain instead of the spatial domain to be compliant with classical codecs JPEG and MPEG that is a significant advantage over references [11] and [12].

Existing anonymization methods focus on a few criteria only, depending on the exact application and fail to satisfy either intelligibility, reversibility, compression efficiency or real time, as it is shown in Table 1.

Our main concern is to fulfil the five criteria, introduced in the beginning of the section 2 and reported in the Table 1. Hence, the method, described in the next section, ensures privacy while preserving the recognition of events in the scene in real time. Moreover, the proposed system is near lossless reversible for authorized people who own a key and compliant with the classical codecs JPEG and MPEG.

3. PROPOSED METHOD

We apply the proposed approach only on privacy-sensitive regions (e.g., faces or bodies of people), denoted RoI, that are either previously annotated or automatically detected by Viola and Jones [16] or Dalal and Triggs [17] methods.

The Intra (I) frames of MPEG are encoded with the main steps of JPEG process (DCT, Quantization) and the recovery of B, P frames are dependant on the last recovered I frame. Thus, to integrate our process in the MPEG framework, we insert our approach in the I predicted frames. The prediction of P and B frames have to be done with the unscrambled I frame, denoted I', otherwise the prediction error will be compensated by the error compensation (i.e. the difference between the prediction and the original frame). Thus, I' frame has to be kept, temporarily, in memory.

Our process is integrated between the quantization step and before the entropy encoding steps as it is shown in Figure 1. We access to the quantized DCT coefficients either by computing the first steps of JPEG process when we are starting from a raw image: Color transformation / 8*8 blocks / DCT / Quantization or by doing the inverse lossless compression mechanisms when we are starting from a compressed data. We denote, qf, the quality factor of the compression (between 1 and 100).

Each DCT 8*8 block contains 64 coefficients: one DC and 63 AC coefficients. The most significant information is localised in the low-frequency coefficients of the DCT and the least significant information in the high-frequency coefficients.

One key idea of the paper is to create a preserved privacy image (PP) which retains the shapes and the motion of the

body by keeping its DC coefficients and hides the scrambled low-frequency coefficients of the DCT of the RoI in the AC coefficients of the PP.

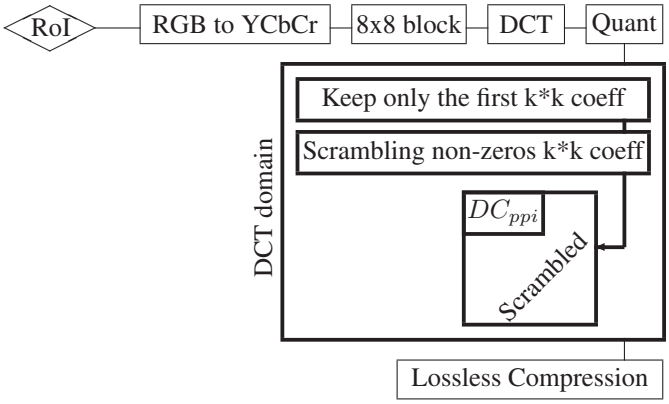


Fig. 1: Workflow of the process

3.1. Design the preserved privacy image (PP)

We compute the PP by keeping only the lowest-frequencies in the quantized DCT coefficients of each block of the RoI. In other word, we retain the quantized DC coefficient for each block, named DC_{pp} in the paper, as it is shown in Figure 1.

This version of PP is shown in Figure 2b, one color is represented for each block therefore it looks like a pixelated image of the RoI. This representation protects privacy while events are still recognizable as we demonstrate in the experimental section, but there is no mechanism to recover the original RoI. In order to have a reversible option, we insert the main information of the RoI in the AC coefficients of the PP DCT blocks, as we explain in the next section. This insertion produces a very limited noise on the final version. This final version is called PP' and is displayed in Figure 2c.

We artificially increase the DC_{pp} by a factor of 5 in order to reduce the visibility of the noise due to the scrambled coefficients during the decompression as explained in the section 3.3. It follows the basic law of Weber Fechner stating that the resolution of perception diminishes for values of greater magnitude.

3.2. Scrambled and embedded information in the Discrete Cosine Transform (DCT) domain

We denoted k, the size of an upper left sub-block (containing low-frequency coefficients). Only the k*k non-zeros low-frequency DCT coefficients of each block (included the DC) are scrambled by computing the operator XOR with random numbers, as it is shown in the formula 1. The 8*8-k*k high-frequency coefficients, which are not used, are definitely lost.

$$c'(i) = c(i) \oplus r(i), \forall i \quad (1)$$

We denoted c a quantized dct coefficient, c' the scrambled c, i the bit position and each bit $\in \{0, 1\}$ and r a random number generated by a pseudo-random number generator (PRG), controlled by a secret key. This secret key is a string known by authorized users only.

The human visual system is much more sensitive to variations in brightness than color, for this reason, we fixed $k = 1$ for Cb and Cr that are the chrominance channels (ac coefficients are removed).

The AC coefficients of the PP block are replaced by the k*k scrambled coefficients associated using the zigzag scanning, as it is shown in Figure 1. The last AC coefficients that are not used, are set to zero.

After embedding the scrambled coefficients in the AC coefficients of the PP blocks, we redo the lossless compression steps in order to measure the impact of our process on the compression rate and we show the trade-off between distortion and compression ratio.

To decompress, two options exist. The first one is to decompress with any interference and the second one is to decompress using the secret key. The result of the first option is the decompressed data where scrambled coefficients are inside the PP whereas the second option retrieves the almost original RoI using the associated secret key.

3.3. Decompression without secret key,

In the default mode, no secret key is provided. The decompression without any secret key leads to visualize the decompressed protected-image which is the PP'. The compressed data is decoded with the inverse process of JPEG.

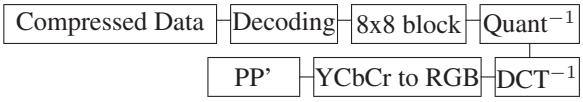
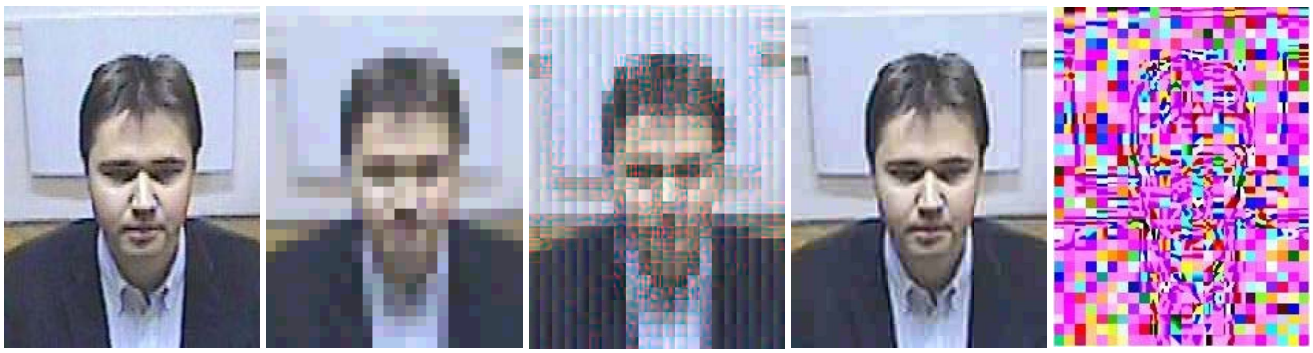


Fig. 4: Decompressed process without the secret key

The decompressed protected-image preserves the global motion and creates a pixelated picture of people as it is shown in Figure 2c. The noise inside blocks is due to the scrambled coefficients.

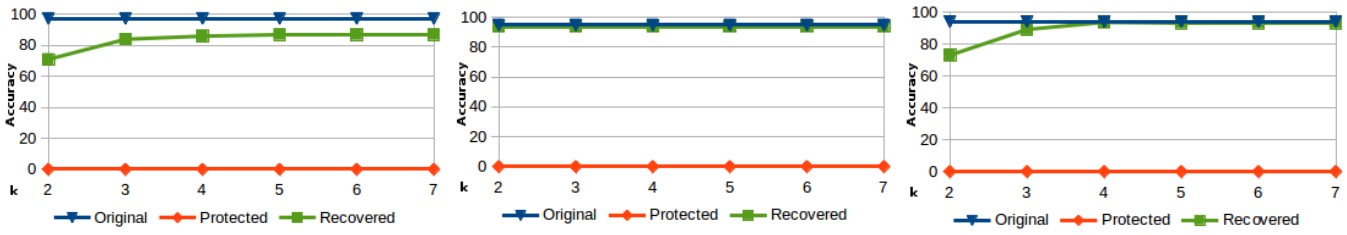
3.4. Decompression with the secret key

With a PRG, controlled by the secret key, numbers are generated with the same pseudo random sequence than the one previously used in 3.2. An XOR is computed between the non-zeros k*k AC coefficients of the 8x8 block and the random numbers to descramble coefficients. Then, the quantization inverse and the DCT inverse are applied on the descrambled coefficients. Finally, all blocks are put together and the recovered image is converted in RGB (Red, Green and Blue channels).



(a) (b) (c) (d) (e)

Fig. 2: With k=4, qf=25: Original RoI (a), PP (b), the PP' (c), recovered image using the correct secret key (d) and using a wrong secret key (e)



(a) (b) (c)

Fig. 3: Accuracy of face recognition for original (blue), protected (red) and recovered (green) faces, depending on k with (a), LBPH (b), Eigen and (c) HoG methods

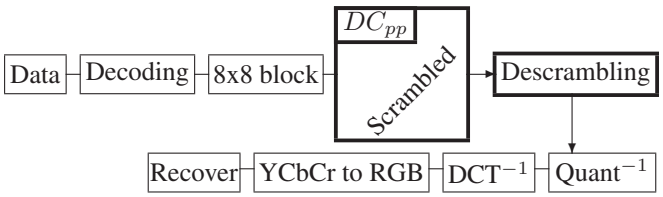


Fig. 5: Decompressed process using a secret key

The recovered image is similar to the original one as it is shown in Figure 2d, contrary to the Figure 2e which is totally scrambled when a user provides a wrong secret key.

4. EXPERIMENTAL RESULTS

We evaluate the efficiency of our proposed approach in terms of privacy protection, intelligibility, reversibility and impact on the compression ratio.

4.1. Face recognition evaluation

Local Binary Pattern Histogram (LBPH) [18], Eigen [19] with the Euclidean distance as well as HoG [20] with SVM, are used as a baseline in face anonymization domain, to compare the impact of anonymization [21]. Thus, in our experiment, we trained these three face recognition algorithms, using a subset of Feret [22] and ScfaceData [23] database.

With $qf = 17$, we apply our proposed privacy filter and the inverse process using the correct secret key, on 300 face images from Feret and ScfaceData, with different values for the parameter k (i.e. the size of an upper left sub-block).

Therefore, face recognition algorithms are tested on original, protected and recovered faces by our proposed approach among different values of k. The performances are shown in Figure 3. The accuracy, for protected faces, is $\sim 10^{-3}$ for all face recognition methods cited previously. The accuracy remain almost the same for Eigen 3b and HoG 3c methods between original and recovered face images and are close for LBPH method 3a.

Thus, the application of our privacy filter foils the identification of people, whereas the reverse application enables to restore almost the initial performances of face recognition.

4.2. Evaluation of sports event classification

To evaluate the impact of the event classification, we chose to test our algorithm on sport events. As classification tool, we utilize Deepdetect[2] which is able to classify 143 sports

[2] http://www.deepdetect.com/

and [24] as dataset. Among the sports available in [24], we select seven categories (Rowing, polo, snowboarding, sailing, rock climbing, croquet, badminton) that can be classified by the tool. Next, we apply our privacy protection approach on the selected images and feed again the classification tool with them. The classification tool gives as results an ordered list of classes from the best to the worst one. Therefore, we have decided to compute the @k accuracy curve with k=10. This curve reports if the proper class is among the ten first results in the ordered list.

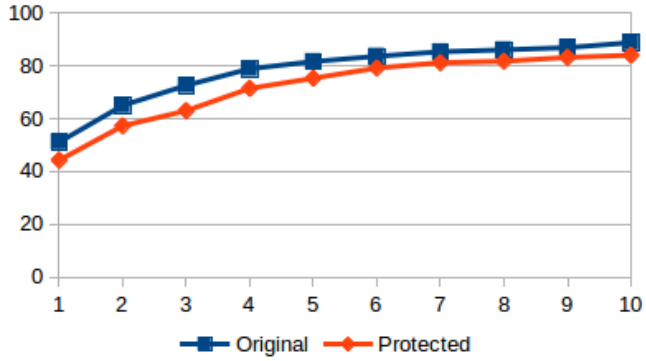


Fig. 6: Accuracy@10 of sports event classification

According to the Figure 6, the accuracy of sport events classification is preserved even if the entire image is protected by our method.

4.3. Reversibility and impact on compression ratio

The peak signal-to-noise ratio (PSNR) measures how much the signal has been corrupted and is commonly used to evaluate the quality of reconstruction. The PSNR and the formula 2, have been selected for, respectively, measuring the visual quality of the recovered images and computing the compression ratio. We have integrated our process with different levels of the strength k on 600 images from the SCFaces Database [23]. Then, we have computed both, the PSNR and the compression ratio between the non-compressed images and the compressed ones in the presence of our privacy process.

$$CompressionRatio = \frac{count(ucd)}{count(cd)} \qquad (2)$$

with *cd* the compressed data, *ucd* the uncompressed data and *count* a function which counts the number of bits of a file.

The quality of the reconstruction and the compression ratio depend on the values of k and are two conflicting objectives. Indeed, according to the Figure 7, the higher is the value of k, the lower is the compression ratio of the image whereas the better is the quality of the reconstruction. The best trade-off is obtained for k=4 or k=5 and qf=2.

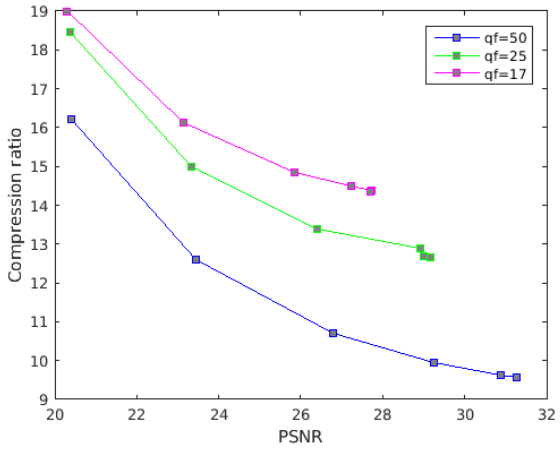


Fig. 7: Compression ratio function of the PSNR for k=2:1:7 (points from the left to the right)

Our process decreases the compression performance due to the scrambling which changes the values of quantized coefficients. The visual quality of the recovered face images is slightly reduced because some original coefficients are definitively lost in the process. Indeed, the original average compression rate and PSNR are, respectively, 22.47 and 30.17 dB, when qf=2.

4.4. Conclusions and Future work

We have presented a novel reversible privacy protection algorithm which fulfils the five significant criteria in video surveillance. The identity of people on faces is protected as we proved it in 4.1. Sport events are still identifiable as we demonstrated it in 4.2. The reversibility is possible for authorized people knowing the key. Our approach is compliant with the compression. Moreover, the parameter k can be tuned depending on the priority between compression and quality of the decompressed image. The integration of the proposed process is not computationally expensive. Indeed, it takes, respectively, 0.001, 0.1, 0.15 seconds longer than the original process of JPEG for a RoI size of 128*128, 256*256, 512*512 (standard code in Matlab).

A preserved privacy version of the image (PP) is designed by keeping only the DC of the RoI, the DC_{pp}. Then, the original information of the RoI is scrambled and hidden within the least significant coefficients of the DCT of the PP. This allows to monitor scenes even if faces and bodies are protected. Thus, motion of people can be followed without being identifiable. Moreover, there is the possibility to change the PP by changing its DC_{pp} values because the DCs of the RoI are included in the k*k non-zeros low frequency coefficients which are scrambled and hidden inside the final image.

Further works would include a subjective evaluation of our approach using human subjects to describe actions recognition as well as privacy protection.

5. REFERENCES

[1] P Grother and M Ngan, "Face recognition vendor test (frvt) performance of face identification algorithms," *NIST Interagency Report*, vol. 8009, pp. 2, 2014.

[2] Jithendra K Paruchuri, Sen-Ching S Cheung, and Michael W Hail, "Video data hiding for managing privacy information in surveillance systems," *EURASIP Journal on Information Security*, vol. 2009, pp. 7, 2009.

[3] Elaine M Newton, Latanya Sweeney, and Bradley Malin, "Preserving privacy by de-identifying face images," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 17, no. 2, pp. 232–243, 2005.

[4] B NAGARJUN SINGH and V DIVYABHARATI, "Reversible de-identification for lossless image compression using reversible watermarking," 2015.

[5] A. Chattopadhyay and Terrance E Boult, "Privacycam: a privacy preserving camera using uclinux on the blackfin dsp," in *Computer Vision and Pattern Recognition. CVPR'07. IEEE Conference on*. IEEE, 2007, pp. 1–8.

[6] Terrance Edward Boult, "Pico: Privacy through invertible cryptographic obscuration," in *Computer Vision for Interactive and Intelligent Environment, 2005*. IEEE, 2005, pp. 27–38.

[7] Andrea Melle and Jean-Luc Dugelay, "Scrambling faces for privacy protection using background self-similarities," in *Image Processing (ICIP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 6046–6050.

[8] Frederic Dufaux and Touradj Ebrahimi, "Scrambling for privacy protection in video surveillance systems," *Circuits and systems for video technology, IEEE Transactions on*, vol. 18, no. 8, pp. 1168–1174, 2008.

[9] Ci Wang, Hong-Bin Yu, and Meng Zheng, "A dct-based mpeg-2 transparent scrambling algorithm," *Consumer Electronics, IEEE Transactions on*, vol. 49, no. 4, pp. 1208–1213, 2003.

[10] W. Zeng and S. Lei, "Efficient frequency domain selective scrambling of digital video," *Multimedia, IEEE Transactions on*, vol. 5, no. 1, pp. 118–129, 2003.

[11] Natacha Ruchaud and Jean-Luc Dugelay, "Privacy protection filter using stegoscrambling in video surveillance," In MediaEval 2015 Workshop, Wurzen, Germany, September, 2015.

[12] Natacha Ruchaud and Jean-Luc Dugelay, "Efficient privacy protection in video surveillance by stegoscrambling," In WIFS, Rome, Italy, November, 2015.

[13] Frederic Dufaux, "Video scrambling for privacy protection in video surveillance: recent results and validation framework," in *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2011, pp. 806302–806302.

[14] William B Pennebaker and Joan L Mitchell, *JPEG: Still image data compression standard*, Springer Science & Business Media, 1993.

[15] Iain E Richardson, *H. 264 and MPEG-4 video compression: video coding for next-generation multimedia*, John Wiley & Sons, 2004.

[16] Paul Viola and Michael J Jones, "Robust real-time face detection," *International journal of computer vision*, vol. 57, no. 2, pp. 137–154, 2004.

[17] Navneet Dalal and Bill Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition. CVPR 2005. IEEE Computer Society Conference on*. IEEE, 2005, vol. 1, pp. 886–893.

[18] Timo Ahonen, Abdenour Hadid, and Matti Pietikainen, "Face description with local binary patterns: Application to face recognition," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 12, pp. 2037–2041, 2006.

[19] Matthew A Turk and Alex P Pentland, "Face recognition using eigenfaces," in *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on*. IEEE, 1991, pp. 586–591.

[20] Oscar Déniz, Gloria Bueno, Jesús Salido, and Fernando De la Torre, "Face recognition using histograms of oriented gradients," *Pattern Recognition Letters*, vol. 32, no. 12, pp. 1598–1603, 2011.

[21] Pavel Korshunov, Andrea Melle, Jean-Luc Dugelay, and Touradj Ebrahimi, "Framework for objective evaluation of privacy filters," in *SPIE Optical Engineering+ Applications*. International Society for Optics and Photonics, 2013, pp. 88560T–88560T.

[22] P Jonathon Phillips, Hyeonjoon Moon, Syed A Rizvi, and Patrick J Rauss, "The feret evaluation methodology for face-recognition algorithms," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 10, pp. 1090–1104, 2000.

[23] M. Grgic, Kresimir D., and S. Grgic, "Scface–surveillance cameras face database," *Multimedia tools and applications*, vol. 51, no. 3, pp. 863–879, 2011.

[24] Li-Jia Li and Li Fei-Fei, "What, where and who? classifying events by scene and object recognition," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*. IEEE, 2007, pp. 1–8.