The impact of privacy protection filters on gender recognition

Natacha Ruchaud[a], Grigory Antipov[a,c], Pavel Korshunov[b], Jean-Luc Dugelay[a], Touradj Ebrahimi[b] and Sid-Ahmed Berrani[c]

[a]EURECOM, 450 route des Chappes, Biot, France;
[b]EPFL, Station 11, CH-1015 Lausanne, Switzerland;
[c]Orange Labs, 4 rue du Clos Courtel, Cesson Sévigné, France

ABSTRACT

Deep learning-based algorithms have become increasingly efficient in recognition and detection tasks, especially when they are trained on large-scale datasets. Such recent success has led to a speculation that deep learning methods are comparable to or even outperform human visual system in its ability to detect and recognize objects and their features. In this paper, we focus on the specific task of gender recognition in images when they have been processed by privacy protection filters (e.g., blurring, masking, and pixelization) applied at different strengths. Assuming a privacy protection scenario, we compare the performance of state of the art deep learning algorithms with a subjective evaluation obtained via crowdsourcing to understand how privacy protection filters affect both machine and human vision.

Keywords: Gender Recognition, Deep Learning, Crowdsourcing

1. INTRODUCTION

Recent years show a steady growth in the adoption of digital video surveillance systems for monitoring buildings and public spaces. These systems raise many concerns related to the privacy rights of the subjects being monitored.[1–3] At the same time, video analytic tools have created additional challenges, as they can extract privacy sensitive information such as identity, ethnicity, and gender. For instance, face recognition or person re-identification can potentially expose the identity of any individual that appears in the field of view of a wearable camera.

In order to protect privacy of people, several filters have been designed and used, e.g., blurring, masking, or pixelization. The goal of privacy protection filters is to reduce detection and recognition rate of privacy sensitive information. Therefore, privacy filters should prevent the recognition of personal information by humans as well as by computer vision algorithms.

In this paper, we focus on the task of gender recognition in images and evaluate via subjective and objective experiments how the gender recognition abilities of both human vision and computer vision are affected by different privacy protection filters. Previously, humans and computers were compared in their abilities to recognize objects.[4] It was shown that humans can easily recognize the gender (above 95% of recognition accuracy from faces[5]) even in the presence of noise, but such task remains challenging for computer vision. Based on the common knowledge of the world, humans can guess the gender even if an image has been degraded. However, a computer vision model is usually trained only on high-quality images, so the answer remains unclear for the algorithms.

In this paper, we perform subjective experiments using crowdsourcing, which is a viable alternative to conventional laboratory-based subjective assessments.[6] Crowdsourcing also provides access to large groups of subjects (or workers) with varying social and geographical backgrounds for a relatively low cost. To display images of

Further author information:
Natacha Ruchaud, e-mail: ruchaud@eurecom.fr
Grigory Antipov, e-mail: grigory.antipov@orange.com
Pavel Korshunov is now with Idiap Research Institute, e-mail: pavel.korshunov@idiap.ch
Touradj Ebrahimi, e-mail: touradj.ebrahimi@epfl.ch

body silhouettes to different workers and to collect evaluation results, QualityCrowd2* framework was used[7] together with Microworkers† crowdsourcing platform, which provided online workers from around the world.

Objective evaluation is done using Convolutional Neural Networks (CNN), which has recently been employed with success on large-scale datasets. We employ Histogram of Oriented Gradients (HOG) features, as they show to be suitable for pedestrian gender recognition.[8,9] For instance, 76% recognition rate is claimed on MIT dataset,[8] while in Ref. 10, authors announce 80% of the recognition accuracy on the VIPeR dataset.

The images for both crowdsourcing and objective experiments were selected from PETA collection of datasets,[11] which contains 19'000 annotated pedestrian images from various surveillance cameras. Since, in video surveillance, it is difficult to get a clear face of individuals, often due to low quality capture or occlusions, we estimate the gender from a global view of the body silhouette. Methods such as Ellipse fitting[12] and Radon Transform Mean Gait Energy Image plus Zernike moments[13] have been implemented for gender recognition from gait.

The rest of the paper is organized as follows. The previous studies related to our work are introduced in Section 2. The datasets and their modifications are presented in Section 3. Then the used gender recognition approaches are described in details and compared in Section 4. Finally, we conclude and give an outlook on possible future works.

2. BACKGROUND AND RELATED WORK

In this section, we describe privacy protection filters used to generate data for objective and subjective evaluations, as well as, overview the state of the art on gender recognition and crowdsourcing approaches commonly used for subjective assessment.

2.1 Privacy filters

To protect privacy, some surveillance systems apply traditional techniques such as masking, Gaussian blurring, and pixelization (e.g., Google Street View, FacePixelizer ‡, and ObscuraCam § on Android).

More advanced privacy filters have been proposed to hide identity, such as morphing[14] which merges two images and K-means[15] which groups colors inside a picture.

Every privacy filter contains a varying parameter which controls the level of privacy protection. The experiments in Ref. 16 demonstrate that, in general, an increase in strength of the filters leads to an increase in privacy (i.e., reduction in recognition accuracy) and at the same time results in a reduced intelligibility (i.e., reduction in detection accuracy).

Masking filter is obtained by applying the following formula:

$$ImgMasking = originalImg * (1 - \alpha), \tag{1}$$

with α representing the opacity between *0.5* and *0.9* (the bigger α, the stronger is the filter).

The idea behind the morphing filter[17] is to find an average face image between the source and the target faces according to a given interpolation level. The source face corresponds to the face of the individual whose identity must be preserved. The target face is any generic human face. The method divides both images into Delaunay triangles[18] and transforms the vertices of the source image to the vertices of the target image. The pixel intensities are also interpolated with respect to a second parameter. Morphing is compression and format independent. It is also reversible. Its security can be ensured by encrypting the key points and randomizing the interpolation level and the pixel interpolation values for each triangle. However, as the algorithm begins with triangulating the face images, it may fail to work in cases where the faces are not captured from 'ideal' angles.

*https://github.com/ldvpublic/QualityCrowd2

†http://microworkers.com/

‡http://www.facepixelizer.com/

§https://guardianproject.info/apps/obscuracam/

Pixelization can be perceived as a downsampling of the images, reducing the overall image quality to protect privacy. The strength of pixelization is controlled by the size of the squares, i.e., ensembles of adjacent pixels (see Figure 1).

Gaussian blurring uses Gaussian kernel to filter an original image:

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}, \tag{2}$$

where x and y are the pixel coordinates and σ is the standard deviation of the Gaussian distribution.

The K-means filter works with pixel values. K-means minimizes the Euclidian distance between a pixel and a centroid of the closest color. First, a number of clusters (denoted here by N) is chosen. Then K-means clustering is applied on the color space (r, g, b) and N centroids are obtained. Finally, original pixel values are replaced by the corresponding centroid values. Strength of K-means is controlled by the number of clusters.

Figure 1. Privacy filters. From left to right: original body, masking, morphing, pixelization, Gaussian blur, K-means.

The goal of the paper is to estimate the robustness of a gender recognition algorithm against privacy filters and to provide a subjective evaluation based on crowdsourcing. This will also allow us to compare the impact of privacy filters on machine vision and on human vision respectively. The following five filters are considered: black masking with opacities equal to 0.5, 0.7, and 0.9 using Eq. 1, morphing with opacities equal to 0.4, 0.7, and 0.9, pixelization with size of the squares equal to 3, 5, and 7, Gaussian blurring with standard deviations equal to 2, 4, and 6, and K-means with number of clusters equal to 2, 4, and 6. An illustration of these filters is shown in Figure 1 and the used strengths are summed up in Table 5.

2.2 Gender recognition in computer vision

Automatic gender recognition might be useful in many applications such as human-computer interaction systems, surveillance systems, content-based indexing and search, biometrics, demographic collection and targeted advertising .[19]

The characteristics of a person (age, ethnicity, facial expressions or accessories) and the image capture process (pose, lighting, illumination, image quality) affect the ability of a computer vision system to recognize the gender.

Guo et al.[20] report that gender classification accuracy is significantly affected by age (adult faces are recognized more easily than young or senior faces). In our paper, we demonstrate that image quality controlled by a filter (blurring, pixelization, masking, K-means and morphing) can affect the recognition by humans as well as by automatic recognition systems.

Even if our work focuses only on gender recognition based on body silhouettes, in this section, we provide a short overview of gender recognition from other modalities as well.

First Author, Year	Feature extraction	Classifier	Average accuracy %
Gutta, 2000[21]	Pixel values	RBF + Decision tree	96
Jain, 2005[22]	ICA	SVM	95.67
Lian, 2006[23]	LBP histogram	SVM-polynomial	94.08
Fok, 2006[24]	Pixel values	Convolutional neural network	97.2
Yang, 2007[25]	LBP histogram	Real Adaboost	96.32
Xia, 2008[26]	LGBMP hist	SVM-RBF	94.96
Wang, 2010[27]	SIFT, Gabor	Adaboost	97
Wu, 2011[28]	LGBP	SVM-RBF	97
Zheng, 2011[29]	LGBP-LDA	SVMACe	99.8

Table 1. Gender recognition from face.

2.2.1 Gender recognition from face

Table 1 summarizes a list of representative works on face gender recognition.

In many approaches, an image preprocessing (e.g. contrast and brightness normalization or geometric alignment) is applied after a face is detected from an image. Feature extraction methods for face gender classification can be categorized into geometric-based and appearance-based methods.[30,31] Zheng et al.[29] obtain the best result for face images from FERET dataset. However, only frontal faces from the dataset were used.

2.2.2 Gender recognition from gait

Works, that are listed in Table 2, are focused on gender recognition from gait (i.e walking, running, jogging and climbing stairs).

First Author, Year	Feature extraction	Classifier	Average accuracy %
Chen, 2009[32]	Radon transform + Relevant component analysis	Mahalanobis dist	95.7
Chang, 2010[33]	DCT	EHMM	94
Felez, 2010[12]	Ellipse fittings	SVM-linear	94.7
Hu, 2010[34]	Gabor + MMI	GMM-HMM	96.77
Oskuie, 2011[13]	RTMGEI + Zernike moments	SVM	98.9

Table 2. Gender recognition from gait.

Oskuie et al.[13] report the best results for the CASIA gait dataset.

2.2.3 Gender recognition from body silhouettes

One of the first attempts to recognize gender from images of body silhouettes was performed by Cao et al.[8] who used HOG features with Adaboost classifier. Then Collins et al.[10] proposed their descriptor called PixelHOG using dense HOG features computed from an edge map. Biologically-inspired features (BIF) were also used for human body gender recognition by Guo et al.[35] Existing methods on gender recognition from body are summarized in Table 3.

First Author, Year	Feature extraction	Classifier	Average accuracy %
Cao, 2008[8]	HOG	Adaboost variant	75
Collins, 2009[10]	PiHOG, colour	SVM-linear	80
Guo, 2009[35]	BIF+PCA/LSDA	SVM-linear	80.6
Bourdev, 2011[9]	HOG, colour histogram, skin pixels	SVM	82.4
Ng, 2013[36]	CNN	CNN	80
Antipov, 2015[37]	CNN	SVM	85 (standrad protocol) 80 (cross-dataset protocol)

Table 3. Gender recognition from body silhouettes.

In Ref. 36 authors train a CNN for gender recognition from body silhouettes on the MIT pedestrian dataset having only 900 images and reach 80 % recognition rate. In Ref. 37 authors obtain the state-of-the-art performance using a CNN trained on PETA collection of datasets.

2.3 Convolutional neural networks (CNN)

Basic notions in visual neuroscience[38] led to the creation of the convolutional and pooling layers in CNN. From 1990s, CNN are used for object detection in natural images, including faces and hands[39] and for face recognition.[40]

In images, deep neural networks exploit the property of local combinations of edge motifs which can be regrouped into parts. Firstly, local groups of pixels are often highly correlated in images, forming distinctive local motifs that are easily detected. Secondly, the local statistics of images are invariant to location.

The architecture of a typical CNN is structured as follows: the first few stages are composed of two types of layers: convolutional layers and pooling layers. Units in a convolutional layer are organized in feature maps and each unit is connected to local patches in the feature maps of the previous layer through a set of weights called a filter bank. The result of this filter bank is then passed through a non-linearity such as a ReLU. The same filter bank is connected with all units in a feature map. Mathematically, a discrete convolution is the filtering operation performed by a feature map.

The role of the convolutional layer is to detect local correlations of features from the previous layer and the role of the pooling layer is to merge semantically similar features into one. The maximum of a local patch of units in one feature map is computed in a typical pooling unit. Usually two or three stages of convolution, non-linearity and pooling are cumulated. Finally, back-propagation gradients allow all the weights in all the filter banks to be trained.

2.4 Privacy assessment based on crowdsourcing

Until recently, little was done to better understand privacy issues in practical multimedia applications. But lately the impact of privacy protection tools has been analyzed in video surveillance and effective evaluation methodologies have been developed to take into account both the context and the content. The objective evaluation of several primitive privacy filters has been reported by Newton *et al.*,[41] where the authors showed that such filters cannot adequately protect from successful face recognition, because recognition algorithms are robust. The robustness of face recognition and detection algorithms to primitive distortions is also reported in Ref. 42. Further, in a work by Dufaux *et al.*,[43] a framework is defined to evaluate the performance of face recognition algorithms applied to images altered by various obfuscation methods.

Crowdsourcing has shown to be a viable alternative to conventional laboratory-based subjective assessments, especially for cognitive tasks.[6] Crowdsourcing-based evaluation of privacy tradeoff in video surveillance has shown good consistency with laboratory-based studies.[44] The crowdsourcing methodology benefits from a large number of participants and can be performed efficiently and at a relatively low cost without requiring a significant commitment from subjects, which are called workers in the crowdsourcing terminology. Workers accept to undertake a task (usually a short 5–20 minutes task) and are grouped in larger units, called batches. When the evaluation experiment is over, workers submit their answers. Unlike laboratory-based experiments, crowdsourcing cannot impose specific displays or controlled illumination of surroundings in which assessments take place. However, since standard environment and equipment conditions for surveillance operators have not been established, typical monitors even with different resolutions and color settings are considered as appropriate in this study.

To display video sequences to different workers and to collect evaluation results, we selected QualityCrowd2 framework[7] and the Microworkers crowdsourcing platform in order to access online workers from around the world. QualityCrowd2 is an open-source framework designed for QoE evaluation with crowdsourcing. This framework was selected because it is easy to modify for our gender recognition task using the provided simple scripting language for batch creation and training sessions.

3. DATASET DESCRIPTION

Contrary to previous studies,[8,10] we do not focus on a single dataset, but on a collection of pedestrians datasets PETA.[11] An extensive evaluation of a recognition algorithm requires a dataset containing various people in diverse environments and situations. Therefore, a large and diverse database of pedestrians has been selected. PETA is a large open-access collection of pedestrian images with several annotations including age, backpack, hat, jacket, jeans, logo, long hair, gender, muffler, no accessory, plastic bag, sandals, shorts, skirt, sunglasses, trousers, t-shirt, etc. Examples of PETA images are presented in Figure 2.

Figure 2. PETA dataset.

Originally, the PETA collection consists of 10 datasets of different sizes with a total of 19'000 images. Appearances of images significantly vary between different datasets of PETA in terms of image resolutions (from 17x39 to 169x365 pixels), camera angles (pictures are taken either by ground-based cameras or by surveillance cameras which are set at a certain height), and environments (indoors or outdoors), as shown in Figure 3.

Datasets	#Images	Camera angle	View point	Illumination	Resolution	Scene
3DPeS	1012	high	varying	varying	from 31x100 to 236x178	outdoor
CAVIAR4REID	1220	ground	varying	low	from 17x39 to 72x141	outdoor
CUHK	4563	high	varying	varying	80x160	outdoor
GRID	1275	varying	frontal&back	low	from 29x67 to 169x365	indoor
i-LIDS	477	medium	back	high	from 32x76 to 115x294	indoor
MIT	888	ground	back	high	64x128	outdoor
PRID	1134	high	profile	low	64x128	outdoor
SARC3D	200	medium	varying	varying	from 54x187 to 150x307	outdoor
TownCentre	6967	medium	varying	medium	from 44x109 to 148x332	outdoor
VIPeR	1264	ground	varying	varying	48x128	outdoor
Total = PETA	**19000**	**varying**	**varying**	**varying**	**varying**	**varying**

Figure 3. PETA dataset.

Thanks to the proposed annotations, we can evaluate gender recognition accuracy on the PETA collection of datasets.

Unfortunately, PETA contains many images of the same persons which can considerably bias the resulting prediction rate. This is why these repetitions are removed from PETA. Moreover, images with very low resolutions (when height is less than 120 pixels or width is less than 40 pixels), images with more than one person, and

images of babies in strollers are also removed. Finally, we are left with 8'365 images which is less than half of the initial size of PETA.

4. EVALUATION EXPERIMENTS

4.1 CNN-based gender recognition

Following the work in Ref. 37, we employ a convolutional neural network for gender recognition. In particular, we adopt an architecture proposed by Krizhevsky *et al.*[45] This architecture is presented in Figure 4. It consists of five convolutional layers and three fully-connected layers. We train a variant of this architecture on PETA dataset. The variation is in the last fully-connected layer, where we use two neurones instead of 1'000, since we only have two target classes (male and female). The network was trained using the Tesla K20c graphical processor.

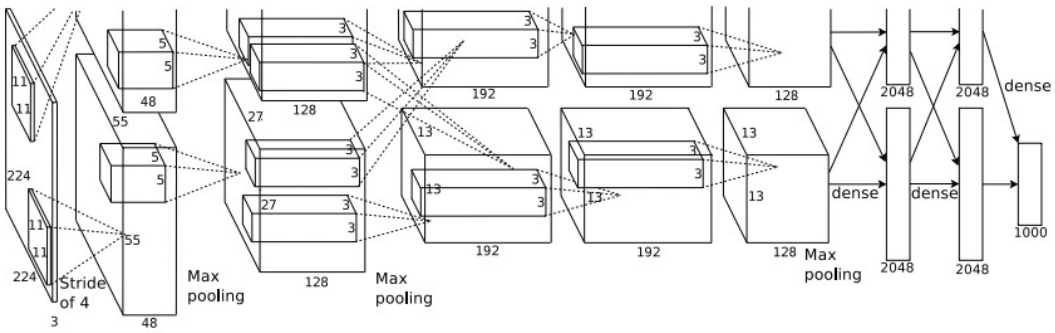

Figure 4. CNN model.

In order to generalize our CNN on heterogeneous and even completely unseen datasets, we firstly train CNN on the dataset which is composed of training parts of CUHK, PRID, GRID, MIT, and VIPeR taken together (4488 Male and 2744 Female pedestrian images). Then, our model is tested on a single big dataset composed of images from the 10 datasets which have not been used in the training set (160 Male and 161 Female pedestrian images). Table 4 summarizes the exact number of training and testing images for each dataset.

Privacy protection filters at different strengths, as described in Table 5, have been applied only on the testing part.

Dataset	**Train size (male + female)**	**Test size (male + female)**
CUHK	$3432 = (2420 + 1012)$	$73 = (34 + 39)$
PRID	$942 = (449 + 493)$	$36 = (16 + 20)$
GRID	$928 = (531 + 397)$	$23 = (10 + 13)$
MIT	$792 = (532 + 260)$	$69 = (41 + 28)$
VIPeR	$1138 = (556 + 582)$	$48 = (24 + 24)$
3DPeS	0	$15 = (10 + 5)$
CAVIAR	0	$9 = (3 + 6)$
i-LIDS	0	$6 = (3 + 3)$
SARC3D	0	$5 = (3 + 2)$
TownCentre	0	$37 = (16 + 21)$

Table 4. Split between training and testing parts per dataset.

In order to evaluate performance of the model, the mean average precision (MAP) is used. Figure 5 shows that the trained CNN is robust to K-means and pixelization (more than 70% of recognition accuracy even if the filter impact is strong), while it is sensible to morphing (less than 60% of recognition accuracy).

Filter	Parameter	Parameter values
Black Masking	opacity	0.5, 0.7, 0.9
Morphing	opacity	0.4, 0.7, 0.9
Pixelization	size of squares	3, 5, 7
Gaussian Blur	standard deviation	2, 4, 6
Kmeans	number of cluster	6, 4, 2

Table 5. Privacy filters and the corresponding varying parameters.

4.2 Crowdsourcing Evaluation

The crowdsourcing assessment aims at checking if a person in a given image can be correctly identified as female or male by an individual, even after the privacy protection filters are applied. For this purpose, each crowdsourcing worker was asked to look at the image of a person and answer the question "What is the gender of the person?" with the following options: "male", "female", and "I don't know".

In total, 300 random images from the PETA dataset were used in the crowdsourcing experiment. They were protected by five different privacy filters at three different strength levels, resulting in $300 \times 5 \times 3 = 4500$ images evaluated in this crowdsourcing study.

To ensure a statistically significant number of evaluations for each image, also taking into account the presence of unreliable subjects (about 50% in a typical crowdsourcing evaluation), 40 subjects were assigned to each image, with a total of 2652 subjects participating in the evaluations.

All versions of the images were randomly distributed among the batches; special care was devoted to guarantee that a particular content was used only once in each batch, i.e., that each subject assessed only one version of a given content. Each batch starts with a message about what is required from the subject, followed by a training session describing the evaluation procedure. A display brightness test is performed using a method similar to that described in Ref.[46] and permits to estimate the subjects' display settings. Subjects are not allowed to skip any content or to avoid answering any question.

Unlike laboratory-based subjective experiments where all subjects can be observed by operators and its test environment also can be controlled, the major shortcoming of the crowdsourcing-based subjective experiments is the inability to supervise participants behavior and to restrict their test conditions. When using crowdsourcing for evaluation, there is a risk of including untrusted data into analysis due to the wrong test conditions or unreliable behavior of some workers who try to submit low quality work in order to reduce their effort while maximizing their received payment.[46] For this reason, unreliable workers detection is an inevitable process in crowdsourcing-based subjective experiments. To identify a worker as 'trustworthy', the following four factors were taken into account in our experiments:

- Task completion time;
- Mean observation time per question;
- Observation duration deviation;
- Number of minority answers.

The objective of the first three factors is to filter out the workers who have strange behaviors in the middle of their tasks, because they are either not serious or have poor concentration. The observation time per question is measured as the time from when the question is displayed until the time the answer is given by the worker. The task completion time, mean observation time and observation duration deviation can be calculated using this data. If the task completion time or mean observation time per question is too long compared to the average of all workers, it can be deduced that they did not take the test seriously or were distracted during their tasks. Unreliable workers were also identified using an approach similar to a typical outlier detection method, commonly used in most subjective quality evaluations. However, typical subjective tests use scoring methods like five-grade evaluation, and outlier detection is performed on mean opinion score.[47] Our experiments do not have opinion scores because of the specific privacy-oriented questions we used. Therefore the number of minority answers

has been used for the outlier detection instead. The assumption is that a participant who has a lot of different answers when compared to the majority of workers is unreliable. The threshold was determined in the same way as for the other three factors.

The above unreliable worker detection methods filtered out 198 workers out of total 2652, resulting in 2454 scores used in the analysis.

4.3 Results

An output of the CNN is binary ("male" or "female") while the output of the crowdsourcing is ternary ("male", "female" and "I don't know"). Therefore, in order to fairly compare the results of the gender recognition based on CNN and crowdsourcing, we assume that 50% of answers "I don't know" registered during the crowdsourcing had been correct if the response would have been selected randomly. Figure 5 illustrates the results of the CNN-based gender recognition as well as crowdsourcing results. Results are given for filtered images (with different parameter values) and for original images (i.e. no filter). The CNN shows results close to those by crowdsourcing for Gaussian blurring, K-means and morphing. CNN is more robust to pixelization ($\sim$ 10% better than human vision). In the case of masking, the human vision is $\sim$ 15% better than the CNN. Protection of the masking filter really depends on the brightness of the display and its environment. Indeed, some crowdsourcing participants might have increased the luminosity of their computer screens making the test images more visible, whereas the CNN works with values of pixels regardless of the display method.

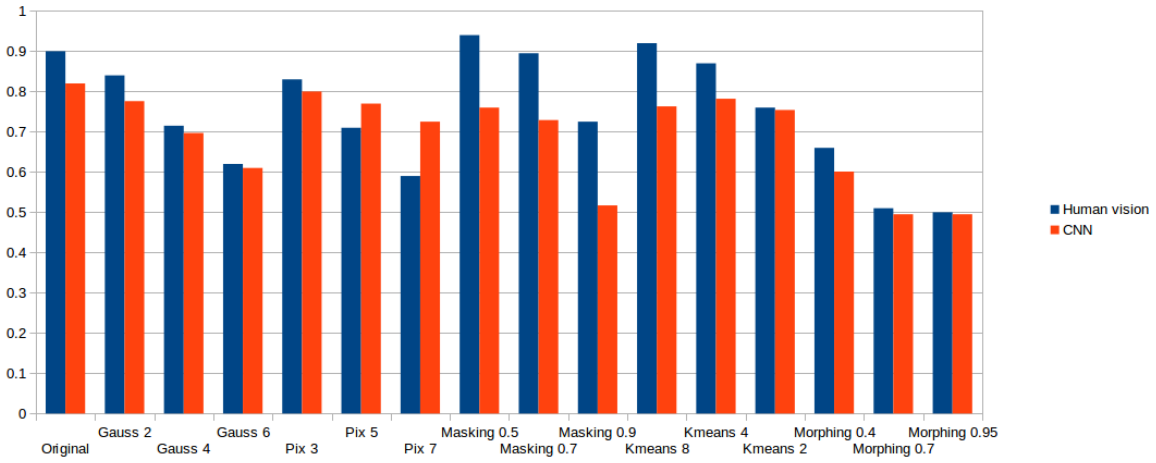

Figure 5. Accuracy results of human vision and CNN.

5. CONCLUSION AND FUTURE WORK

This paper demonstrates the impact of privacy filters on gender recognition by machine vision algorithms and by human vision. We assessed gender objectively (using a CNN) and subjectively (using a crowdsourcing approach). One might expect human vision to be more robust to privacy filters than computer vision. Nevertheless, our results show that humans and automatic gender recognition systems perform almost equally.

In our future work, we would like to explore and compare the nature and causes of errors committed by humans and automatic gender recognition systems.

6. ACKNOWLEDGEMENTS

This work has been partially performed under the European Network of Excellence VideoSense under FP7 ¶.

¶http://videosense.eu/

REFERENCES

[1] Dufaux, F. and Ebrahimi, T., "A framework for the validation of privacy protection solutions in video surveillance.," in [*ICME*], 66–71 (2010).

[2] Fradi, H., Melle, A., and Dugelay, J.-L., "Contextualized privacy filters in video surveillance using crowd density maps," in [*Multimedia (ISM), 2013 IEEE International Symposium on*], 92–99, IEEE (2013).

[3] Moon, H.-M. and Pan, S. B., "Implementation of the privacy protection in video surveillance system," in [*Secure Software Integration and Reliability Improvement, 2009. SSIRI 2009. Third IEEE International Conference on*], 291–292, IEEE (2009).

[4] Borji, A. and Itti, L., "Human vs. computer in scene and object recognition," in [*Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*], 113–120, IEEE (2014).

[5] Bruce, V., Burton, A. M., Hanna, E., Healey, P., Mason, O., Coombes, A., Fright, R., and Linney, A., "Sex discrimination: how do we tell the difference between male and female faces?," *Perception* (1993).

[6] Korshunov, P., Nemoto, H., Skodras, A., and Ebrahimi, T., "Crowdsourcing-based evaluation of privacy in hdr images," in [*SPIE Photonics Europe*], 913802–913802, International Society for Optics and Photonics (Apr. 2014).

[7] Keimel, C., Habigt, J., Horch, C., and Diepold, K., "Qualitycrowd a framework for crowd-based quality evaluation," in [*Picture Coding Symposium (PCS), 2012*], 245–248, IEEE (2012).

[8] Cao, L., Dikmen, M., Fu, Y., and Huang, T. S., "Gender recognition from body," in [*Proceedings of the 16th ACM international conference on Multimedia*], 725–728, ACM (2008).

[9] Bourdev, L., Maji, S., and Malik, J., "Describing people: A poselet-based approach to attribute classification," in [*Computer Vision (ICCV), 2011 IEEE International Conference on*], 1543–1550, IEEE (2011).

[10] Collins, M., Zhang, J., Miller, P., and Wang, H., "Full body image feature representations for gender profiling," in [*Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*], 1235–1242, IEEE (2009).

[11] Deng, Y., Luo, P., Loy, C. C., and Tang, X., "Pedestrian attribute recognition at far distance," in [*Proceedings of the ACM International Conference on Multimedia*], 789–792, ACM (2014).

[12] Martin-Felez, R., Mollineda, R. A., and Sánchez, J. S., "Towards a more realistic appearance-based gait representation for gender recognition," in [*Pattern Recognition (ICPR), 2010 20th International Conference on*], 3810–3813, IEEE (2010).

[13] Oskuie, F. B. and Faez, K., "Gender classification using a novel gait template: radon transform of mean gait energy image," in [*Image Analysis and Recognition*], 161–169, Springer (2011).

[14] Korshunov, P. and Ebrahimi, T., "Towards optimal distortion-based visual privacy filters," in [*2014 IEEE International Conference on Image Processing (ICIP)*], 6051–6055 (Oct 2014).

[15] Fradi, H., Yan, Y., and Dugelay, J.-L., "Privacy protection filter using shape and color cues," (2014).

[16] Korshunov, P., Melle, A., Dugelay, J.-L., and Ebrahimi, T., "Framework for objective evaluation of privacy filters," in [*SPIE Optical Engineering+ Applications*], 88560T–88560T, International Society for Optics and Photonics (2013).

[17] Korshunov, P. and Ebrahimi, T., "Using Face Morphing to Protect Privacy," in [*IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS)*], 208 – 213 (Aug. 2013).

[18] Benson, P. J., "Morph transformation of the facial image," *Image and Vision Computing* **12**(10), 691 – 696 (1994).

[19] Ng, C. B., Tay, Y. H., and Goi, B. M., "Vision-based human gender recognition: A survey," *arXiv preprint arXiv:1204.1611* (2012).

[20] Guo, G., Dyer, C. R., Fu, Y., and Huang, T. S., "Is gender recognition affected by age?," in [*Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*], 2032–2039, IEEE (2009).

[21] Gutta, S., Huang, J. R., Jonathon, P., and Wechsler, H., "Mixture of experts for classification of gender, ethnic origin, and pose of human faces," *Neural Networks, IEEE Transactions on* **11**(4), 948–960 (2000).

[22] Jain, A., Huang, J., and Fang, S., "Gender identification using frontal facial images," in [*Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*], 4–pp, IEEE (2005).

[23] Lian, H.-C. and Lu, B.-L., "Multi-view gender classification using local binary patterns and support vector machines," in [*Advances in Neural Networks-ISNN 2006*], 202–209, Springer (2006).

[24] Phung, S. L. and Bouzerdoum, A., "A pyramidal neural network for visual pattern recognition," *Neural Networks, IEEE Transactions on* **18**(2), 329–343 (2007).

[25] Yang, Z. and Ai, H., "Demographic classification with local binary patterns," in [*Advances in Biometrics*], 464–473, Springer (2007).

[26] Xia, B., Sun, H., and Lu, B.-L., "Multi-view gender classification based on local gabor binary mapping pattern and support vector machines," in [*Neural Networks, 2008. IJCNN 2008.(IEEE World Congress on Computational Intelligence). IEEE International Joint Conference on*], 3388–3395, IEEE (2008).

[27] Wang, J.-G., Li, J., Lee, C. Y., and Yau, W.-Y., "Dense sift and gabor descriptors-based face representation with applications to gender recognition," in [*Control Automation Robotics & Vision (ICARCV), 2010 11th International Conference on*], 1860–1864, IEEE (2010).

[28] Wu, T.-X., Lian, X.-C., and Lu, B.-L., "Multi-view gender classification using symmetry of facial images," *Neural Computing and Applications* **21**(4), 661–669 (2012).

[29] Zheng, J. and Lu, B.-L., "A support vector machine classifier with automatic confidence and its application to gender classification," *Neurocomputing* **74**(11), 1926–1935 (2011).

[30] BenAbdelkader, C. and Griffin, P., "A local region-based approach to gender classi. cation from face images," in [*Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*], 52–52, IEEE (2005).

[31] Li, B., Lian, X.-C., and Lu, B.-L., "Gender classification by combining clothing, hair and facial component classifiers," *Neurocomputing* **76**(1), 18–27 (2012).

[32] Chen, L., Wang, Y., Wang, Y., and Zhang, D., "Gender recognition from gait using radon transform and relevant component analysis," in [*Emerging Intelligent Computing Technology and Applications*], 92–101, Springer (2009).

[33] Chang, C.-Y. and Wu, T.-H., "Using gait information for gender recognition," in [*Intelligent Systems Design and Applications (ISDA), 2010 10th International Conference on*], 1388–1393, IEEE (2010).

[34] Hu, M., Wang, Y., Zhang, Z., and Wang, Y., "Combining spatial and temporal information for gait based gender classification," in [*Pattern Recognition (ICPR), 2010 20th International Conference on*], 3679–3682, IEEE (2010).

[35] Guo, G., Mu, G., and Fu, Y., "Gender from body: A biologically-inspired approach with manifold learning," in [*Computer Vision–ACCV 2009*], 236–245, Springer (2010).

[36] Ng, C.-B., Tay, Y.-H., and Goi, B.-M., "A convolutional neural network for pedestrian gender recognition," in [*Advances in Neural Networks–ISNN 2013*], 558–564, Springer (2013).

[37] Antipov, G., Berrani, S.-A., Ruchaud, N., and Dugelay, J.-L., "Learned vs. hand-crafted features for pedestrian gender recognition," in [*Proceedings of the ACM International Conference on Multimedia*], ACM (2015).

[38] Hubel, D. H. and Wiesel, T. N., "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *The Journal of physiology* **160**(1), 106 (1962).

[39] Vaillant, R., Monrocq, C., and Le Cun, Y., "Original approach for the localisation of objects in images," *IEE Proceedings-Vision, Image and Signal Processing* **141**(4), 245–250 (1994).

[40] Lawrence, S., Giles, C. L., Tsoi, A. C., and Back, A. D., "Face recognition: A convolutional neural-network approach," *Neural Networks, IEEE Transactions on* **8**(1), 98–113 (1997).

[41] Newton, E., Sweeney, L., and Malin, B., "Preserving privacy by de-identifying face images," *IEEE Trans. on Knowledge and Data Engineering* **17**, 232–243 (Feb. 2005).

[42] Korshunov, P. and Ooi, W. T., "Video quality for face detection, recognition, and tracking," *ACM Trans. Multimedia Comput. Commun. Appl.* **7**, 14:1–14:21 (Sept. 2011).

[43] Dufaux, F. and Ebrahimi, T., "A framework for the validation of privacy protection solutions in video surveillance," in [*Proceedings of IEEE International Conference on Multimedia & Expo (ICME 2010)*], (July 2010).

[44] Korshunov, P., Cai, S., and Ebrahimi, T., "Crowdsourcing approach for evaluation of privacy filters in video surveillance," in [*Proceedings of the ACM Multimedia 2012 Workshop on Crowdsourcing for Multimedia*], *CrowdMM'12*, 35–40 (Oct. 2012).

[45] Krizhevsky, A., Sutskever, I., and Hinton, G. E., "Imagenet classification with deep convolutional neural networks," in [*Advances in neural information processing systems*], 1097–1105 (2012).

[46] Hossfeld, T., Keimel, C., Hirth, M., Gardlo, B., Habigt, J., Diepold, K., and Tran-Gia, P., "Best practices for QoE crowdtesting: QoE assessment with crowdsourcing," *IEEE Transactions on Multimedia* **PP**(99), 1–1 (2013).

[47] ITU-R BT.500-13, R., [*Methodology for the subjective assessment of the quality of television pictures*], International Telecommunication Union, Geneva, Switzerland (2012).