TUB @ MediaEval 2014 Visual Privacy Task: Reversible Scrambling on Foreground Masks

Sebastian Schmiedeke[1,2], Pascal Kelm[1,2], Lutz Goldmann[2] and Thomas Sikora[1]

[1] Communication Systems Group
Technische Universität Berlin, Germany

[2] Imcube Labs GmbH
Berlin, Germany

ABSTRACT

This paper describes our participation in the Visual Privacy Task of MediaEval 2014, which aims to obscure human occurrence in image sequences. As a result the recorded person should be unrecognisable, but if needed the obscured areas can be recovered. We use an approach which models the background and pseudo-randomly scrambles pixels within disjunct foreground areas. This technique is reversible and preserves the colour characteristic of each area. So, colour-based approaches will still be able to automatically distinguish between differently dressed individuals. The evaluations of our results show that the privacy aspect got a high score in all three evaluation streams. The level of intelligibility and the pleasantness of our approach is below the average, since scrambling results in lower 'aesthetic' images.

1. INTRODUCTION

Video surveillance of public spaces is expanding. Consequently, individuals are increasingly concerned about the 'invasiveness' of such ubiquitous surveillance and fear that their privacy is at risk. The demands of stakeholders to prevent criminal activities are often seen to be in conflict with the privacy requirements of individuals. The main challenge is to preserve the anonymity of the surveyed individuals and also to fulfil the stakeholders needs. The problem of privacy protection in video surveillance is concerned in this year's MediaEval Visual Privacy Task [1]. A typical way to protect privacy in images and videos is to apply techniques such as blurring or masking. Since these techniques are irreversible, scrambling is introduced in [2]: A transform-domain scrambling technique, where pixels in the respective regions are pseudo-randomly scrambled based on a secret key. Our approach is quite similar, but applied on the pixel of disjunct foreground masks to preserve the less invasive image background. An exemplary frame is shown in Fig. 1.

2. METHODOLOGY

Our proposed privacy-protection approach consists of a background modelling module and a scrambling module that obfuscates foreground masks. Since the PEViD videos [4] depict static scenes with a low numbers of occurring and moving people, the scrambled foreground still allows to identify persons' movements and actions. Details such as faces are only recognizable in the recovered (de-scrambled) images.

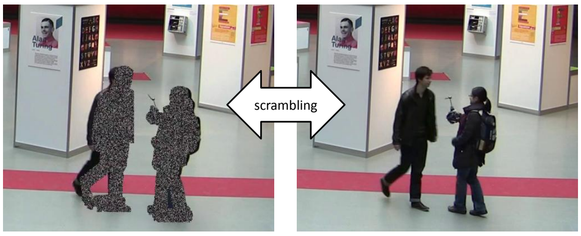

Figure 1: Original frame and its scrambled version.

2.1 Background Modelling

We use background subtraction for generating a foreground mask for each frame. In order to compensate slight camera movements, each frame is subsampled by two and the resulting masks are interpolated properly. Our background modelling module relies on a improved background subtraction scheme [5] based on Gaussian-Mixture models (GMM). This algorithm automatically selects the needed number of Gaussian components per pixel. The mixture of these components tries to reflect the desired background colour by incorporating the recent 300 frames, due to the static video content. The number of components is controlled by a Mahalanobis distance threshold. If the squared Mahalanobis distance of a pixel colour to any existing component exceeds this threshold ($th = 15$) a new Gaussian is generated. Foreground pixels are determined by their belonging to components with small weights. We apply erosion and morphological operations on the foreground masks to eliminate outlier. Our aim was to perfectly expose the silhouettes of persons, but that target was not always achieved (see Fig. 2 for examples of a good foreground estimation and a bad estimation).

2.2 Reversible Scrambling

These foreground areas are then obfuscated by shuffling their pixels. So, an obfuscated area differs from its original version in a changed sequence of their pixels.

The shuffle algorithm is based on a modified variant of the Fisher-Yates method [3] which generates 'random' permutations. The original sequence consists of M disjunct areas to be obfuscated. Each area a is then represented by a vector containing its line-by-line scanned N pixels. These areas are obfuscated by changing the order of its pixels and mapping back the pixels to its original shape. The new pixel order of each area is determined by swapping each i-th pixel with the

MediaEval 2014 Workshop, October 16-17, 2014, Barcelona, Spain

Table 1: Evaluations according different streams (median values of the task are in brackets)

	stream 1	stream 2	stream 3
Intelligibility	73.6 % (74.9 %)	75.2 % (79.3 %)	66.5 % (69.6 %)
Privacy	59.0 % (50.2 %)	62.6 % (46.5 %)	60.7 % (40.7 %)
Pleasantness	21.9 % (24.8 %)	60.8 % (69.6 %)	58.1 % (59.7 %)

j-th pixel, where j is defined by a pseudo-random number generator and the constraint that $j \leq i + 1$.

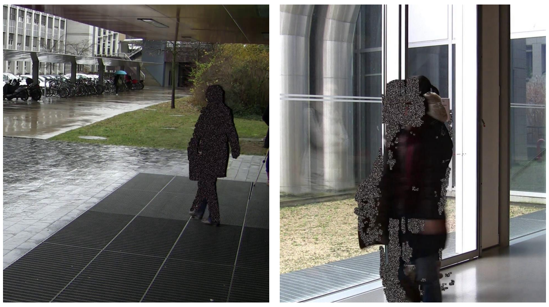

Figure 2: Example for a good foreground mask (left) and a bad mask (right) [image section].

So, the permutation of the pixels of each foreground areas is determined by the order generated by a pseudo-random sequence. The pseudo-random sequence is repeatable due to the characteristics of the pseudo-random number generator (PRNG). The PRNG produces a random, but repeatable sequence of integer numbers by specifying a certain, but fixed seed. This seed is generated from the hash value of a chosen password. This value is fixed for all regions in each frame and video sequence. Since the pseudo-random sequence is repeatable through the given seed, the permutation of pixels is reversible. So, the scrambled image regions can be recovered by knowing the password and the shape of each disjunct scrambled area. We choose for scrambling instead of cryptography to be robust against image compression artefacts and transmission errors. Those errors will also affect the recovered frame in terms of distorted pixels, but these errors will not break the de-scrambling scheme.

3. EXPERIMENTS

The video sequences of the VPT dataset [1] [4] are obscured by scrambling foreground objects within each frame. Since the area of faces are provided with the data set, we include these areas in our foreground masks. So we ensure that the faces are obscured even if it is not part of our foreground mask. We are sure that individuals can be identified not only by their face but also their clothes or accessories. So, the individuals are anonymised at best and a colour-based cluster algorithm may also be able to group areas depicting the same person.

The evaluation of the obscured videos took place using subjective procedures. Three different groups are asked to survey the videos and respond to question concerning the content (number of persons, actions, etc.). Three metrics are generated from these surveys: pleasantness, intelligibility, and privacy. These groups contains of crowdsourced workers and two focus groups, the scores based on their opinions is shown in Table 1.

Pleasantness stands for the influence of the obscuring filter on the human perception of the image distortion. The subjective score is based on the level of user acceptance. Here the score is below the median value resulting from distraction of the users.

Intelligibility stands for the ability of identifying actions and objects within video frames. All three groups evaluate our filter with high scores that are close to the median. Since full masks of person are retained, their action should be recognizable.

The privacy metric concerns about the identification of individuals through their faces, ethics or personal accessories. This score is much higher the average. A high subjective score was excepted, since it is very hard for the human eye to recognise structures within scrambled areas.

We expect higher score in all three categories when applying a more accurate background subtraction algorithm.

4. CONCLUSION

We propose a reversible approach for scrambling foreground masks within images or videos to obscure its content. This approach ensures a high level of privacy, and achieves a standard level in the other aspects, like pleasantness and intelligibility. In future we will investigate the effect of more accurate foreground masks on these privacies scores. The clue is that these areas can be recovered for further analysis, if the foreground mask and the password which generated the seed for the pseudo-random number generation are known.

5. ACKNOWLEDGMENTS

The research leading to these results has received funding from the European Community's FP7 under grant agreement number FP7-261743 (VideoSense).

6. REFERENCES

[1] A. Badii, T. Ebrahimi, C. Fedorczak, P. Korshunov, T. Piatrik, V. Eiselein, and A. Al-Obaidi. Overview of the MediaEval 2014 Visual Privacy Task. In *MediaEval 2014 Workshop*, Barcelona, Spain, October 16-17 2014.

[2] F. Dufaux and T. Ebrahimi. Video surveillance using jpeg 2000. In *Optical Science and Technology, the SPIE 49th Annual Meeting*, pages 268–275. International Society for Optics and Photonics, 2004.

[3] R. Durstenfeld. Algorithm 235: Random permutation. *Commun. ACM*, 7(7):420–, July 1964.

[4] P. Korshunov and T. Ebrahimi. PEViD: privacy evaluation video dataset. *Applications of Digital Image Processing XXXVI*, 25-29 August 2013.

[5] Z. Zivkovic. Improved adaptive gaussian mixture model for background subtraction. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 2, pages 28–31 Vol.2, Aug 2004.