Imcube @ MediaEval 2015 DroneProtect Task: Reversible Masking using Steganography

Sebastian Schmiedeke, Pascal Kelm, and Lutz Goldmann
Imcube Labs GmbH
Berlin, Germany
{schmiedeke, kelm, goldmann}@imcube.de

ABSTRACT

This paper describes Imcube's participation in the DroneProtect Task of MediaEval 2015, which aims to obscure privacy-concerned image regions in videos sequences captured with drones. As a result persons and vehicles should be unrecognisable, but the semantic meaning of the scene should remain understandable to viewer. We use an approach which replaces the privacy-concerned region with an automatically computed composite of inpainted background and foreground contour. Before obfuscation, the image region to be hidden is extracted and steganographically embedded into the processed frame leading to a reversible solution. The evaluation shows that the developed solution achieves good privacy protection while preserving the intelligibility and aesthetic pleasantness of the original video.

1. INTRODUCTION

Since drones become affordable, these devices are increasingly used for security applications. Due to their flexibility videos captures by a drone may contain highly sensitive personal data. 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 DroneProtect Task of MediaEval 2015 deals with the problem of privacy protection in dynamic surveillance videos [1].

A common way to protect privacy in images and videos is to apply techniques such as blurring or masking, as shown in [3]. Since these techniques are irreversible, steganography [4] can be used to preserve this information. A typical steganography algorithm is located in the process chain between quantisation of the DCT coefficients and Huffman coding. The information to be hidden is embedded within the least significant bit of non-zero AC coefficients of each DCT block.

2. APPROACH

Our approach combines both masking and steganography to obtain a visually appealing obfuscated video and to have the possibility to recover the original frame. An example is shown in Figure 1.

MediaEval 2015 Workshop, Sept. 14-15, 2015, Wurzen, Germany

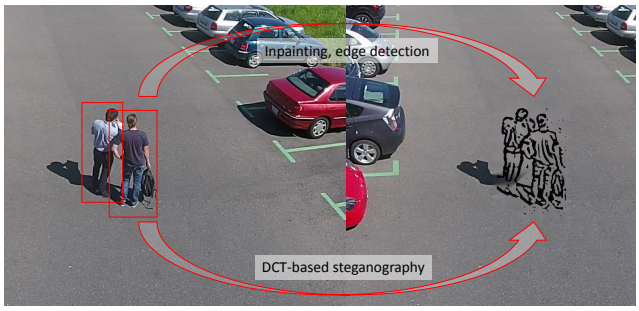

Figure 1: Exemplary frame to visualise our approach: Our masking algorithm obfuscates privacy-concerning regions, but preserving the semantics. The steganography algorithm guarantees that obfuscated areas can be recovered, if needed.

2.1 Masking

Since the videos captured by the mini-drones [3] are highly dynamic compared to static surveillance cameras, traditional background subtraction techniques [5] cannot be applied for foreground object extraction. Instead of that we use edge detection to extract the outline of the foreground objects within the region of interest. Hence, each frame of the sequence is transformed into grey scale and then smoothed by applying a Gaussian kernel (5×5). Edges are detected by applying adaptive thresholding. Therefore, the area around each pixel is cross-correlated with Gaussian windows of a sufficiently large kernel width. Pixels exceeding the weighted sum of cross-correlation become edge pixels which are subsequently used to form the outline. The binary outline is enhanced by applying morphological operations.

In order to remove the original foreground object, the region of interest needs to be filled with reasonable background information. We rely on the inpainting algorithm by Telea [6] which is able to rapidly reconstruct missing image parts and works as follows. Starting with the boundaries the colour information is propagated to the inside of the region by smoothly interpolating along pixel intensity lines. For this purpose, the "image smoothness information" [2], which is estimated by a weighted sum of Laplacians of the known neighbourhood, is propagated along these intensity lines. The direction of the intensity lines, also called "isophotes", is estimated by discretised intensity gradient vectors. Based on the assumption that isophotes have the smallest changes along their direction and the largest changes perpendicular to their direction, it is estimated by finding the largest gra-

Table 1: Subjective evaluation metrics assigned by different subject groups.

Subject group	Privacy	Intelligibility	Pleasantness	Deviation
Category 1 (experts)	0.46	0.65	0.70	0.12
Category 2 (novices)	0.53	0.57	0.63	0.06
Overall	0.50	0.61	0.67	0.09

dient which is orthogonal to the isophote belonging to that pixel. To improve the temporal stability of the inpainting result, the inpainted areas are temporally filtered, e.g. for each pixel the median value with its temporally neighbours is computed.

The obfuscated region of interest is obtained by blending the extracted foreground contour with the reconstructed background texture. Depending on the application the contour may be emphasized with different colours.

2.2 Steganography

Since the masking algorithm itself is irreversible, the original image data contained within the region of interested must be embedded into the obfuscated image. Therefore, we make use of a steganography library, which is based on the F5-algorithm [7]. This algorithm embeds binary information into the DCT coefficients of a JPEG image. The least significant bits of non-zero AC coefficients are replaced by the bits to be embedded in such a way that the statistical distribution of coefficients remains unchanged. Since only non-zero coefficients can carry steganographic values and these coefficients occur less frequent than zero-valued coefficients, only a limited amount of data can be embedded. Depending on the amount and size of regions of interests that shall be hidden, the capacity is often insufficient. Therefore, each region is treated as rectangular region which is JPEG compressed with an adjustable compression parameter. All compressed regions together with their bounding boxes are then concatenated, encrypted and embedded in the obfuscated JPEG encoded image. Since the embedded information maybe destroyed if the image sequence is transcoded, the individual frames are simply combined into a Motion JPEG video.

3. EXPERIMENTS & RESULTS

The video sequences of the DroneProtect dataset [3] are obscured by replacing foreground objects with their outlines. We are sure that individuals can be identified not only by their face but also their clothes or accessories. So, the colour of objects with each regions are replaced by an estimated background. Since the regions of interests are provided with the dataset, we apply our masking algorithm only on the provided areas.

The evaluation of the obscured videos took place using subjective procedures. Two groups of subjects with different experience in surveillance applications were asked to survey the videos and respond to questions concerning the content. Based on the answers to these questions three different metrics (privacy, pleasantness, intelligibility) and the deviation between were computed. The average scores of the different subject groups (experts, novices, overall) for all 38 videos are summarized in Table 1. In the following we will analyse the results for the different metrics and different subject groups.

The *privacy* metric measures how well the identity of the persons and vehicles was protected through the obfuscation or in other words how difficult the obfuscation made the identification of a person or vehicle by hiding relevant visual information. The proposed method achieves a medium overall score (0.50) which suggests that even though only the contour of the object of interest is preserved in some cases it can still be identified. A deeper analysis of these cases is needed to identify potential improvements.

Intelligibility stands for the ability of classifying objects and actions within a video sequence and evaluates how well the activities within a scene are preserved even if the object of interest is obfuscated to prevent its identification. The proposed method achieves a good overall score (0.61) which shows that contour information alone is enough to understand most of the semantics of a scene from a surveillance perspective. A deeper analysis of the individual videos is needed to understand what additional information is needed to improve the intelligibility further.

Pleasantness evaluates the influence of the obfuscation method on the visual quality of the video or by how much the quality of the video is degraded by distortions and artefacts within the region of interest. The subjective score is based on the level of user acceptance. Here, the proposed methods achieves a good overall score (0.67), since the black foreground contours composed over an inpainted background blend well with the original content outside the region of interest. This score may be further improved by using a more sophisticated inpainting algorithm which reconstructs a more plausible background texture.

Since the three metrics mentioned above evaluate quite contrary requirements, the *deviation* evaluates difference between these metrics by computing the standard deviation. As it can be expected from the similar scores for the different metrics, the proposed approach has a very good deviation score (0.09). This shows that it strikes a good balance between the different criteria (privacy, intelligibility and pleasantness).

Comparing the results between the different subject groups shows that the scores of the novices are more equal across the different metrics, while the experts evaluate the intelligibility and pleasantness higher and the privacy lower. This follows the intuition that experts will be able to recognize actions and identities better than novices. The higher pleasantness score suggests that experts value the content of a video higher than its quality.

4. CONCLUSION

A reversible approach for protecting the privacy based on masking and steganographic embedding of the region of interest has been proposed and evaluation on the DroneProtect dataset. The results shows that the approach strikes a good balance between privacy, intelligibility and pleasantness. For developing potential improvements a more detailed analysis of the scores for the individual videos is needed.

5. REFERENCES

[1] A. Badii, P. Koshunov, H. Oudi, T. Ebrahimi, T. Piatrik, V. Eiselein, N. Ruchaud, C. Fedorczak, J.-L. Dugelay, and D. F. Vazquez. Overview of the MediaEval 2015 Drone Protect Task. In *MediaEval 2015 Workshop*, Wurzen, Germany, September 14-15 2015.

[2] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester. Image inpainting. In *Annual Conference on Computer Graphics*, pages 417–424, 2000.

[3] M. Bonetto, P. Korshunov, G. Ramponi, and T. Ebrahimi. Privacy in mini-drone based video surveillance. In *Workshop on De-identification for privacy protection in multimedia*, 2015.

[4] T. Morkel, J. H. P. Eloff, and M. S. Olivier. An Overview of Image Steganography. In H. S. Venter, J. H. P. Eloff, L. Labuschagne, and M. M. Eloff, editors, *Proceedings of the Fifth Annual Information Security South Africa Conference (ISSA2005)*, Sandton, South Africa, 6 2005. Published electronically.

[5] S. Schmiedeke, P. Kelm, L. Goldmann, and T. Sikora. TUB @ MediaEval 2014 Visual Privacy Task: Reversible Scrambling on Foreground Masks. In *Proceedings of the MediaEval 2014 Multimedia Benchmark Workshop*, pages 73–74. CEUR-WS, 2014.

[6] A. Telea. An image inpainting technique based on the fast marching method. *Journal of graphics tools*, 9(1):23–34, 2004.

[7] A. Westfeld. F5-a steganographic algorithm. In I. S. Moskowitz, editor, *Information Hiding*, volume 2137 of *Lecture Notes in Computer Science*, pages 289–302. Springer Berlin Heidelberg, 2001.