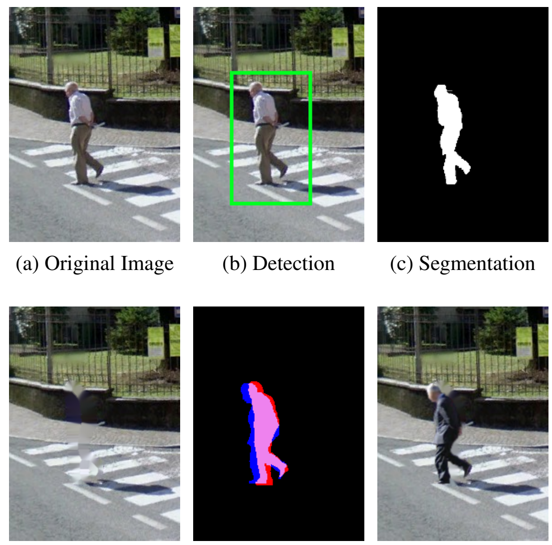
Digital Privacy: Replacing Pedestrians from Google Street View Images

Angelo Nodari, Marco Vanetti, Ignazio Gallo
Dipartimento di Scienze Teoriche e Applicate, University of Insubria, Italy
{*angelo.nodari,marco.vanetti,ignazio.gallo*}@*uninsubria.it*

Abstract

Given the lack of modern techniques to ensure the digital privacy of individuals, we want to pave the way for a new approach to make pedestrians in cityscape images anonymous. To address these concerns, we propose an automated method to replace any unknown pedestrian with another one which is extracted from a controlled and authorized dataset. The techniques used up to now to make people anonymous are based mainly on the blurring of people's faces, but even so it is possible to trace the identity of the subject starting from his clothing, personal items, hairstyle, the place and time where the photo was taken. The proposed method aims to make the pedestrians completely anonymous, and consists of four phases: firstly we identify the area where the pedestrian is located, we separate the pedestrian from the background, we select the most similar pedestrian from a controlled dataset and subsequently we substitute it. Our case study is Google Street View because it is one of the online services which suffers most from this kind of privacy issues. The experimental results show how this technique can overcome the problems of digital privacy with promising results.



(a) Original Image (b) Detection (c) Segmentation

(d) Inpainting (e) Replacing (f) Final Result

Figure 1: Summarization of the steps explained in Section 2 used to automatically make a pedestrian anonymous within a generic cityscape image.

1. Introduction

Nowadays we have access to an increasing amount of multimedia content and for this reason the digital privacy is becoming a fundamental phenomenon on the world stage. First of all Google Inc. with its Street View service, has collected from 2007 to now a huge amount of pictures of the major cities and suburbs of more than 37 countries around the world [2], and it seems to be a growing phenomenon. This process of gathering public images has inevitably collided with privacy issues. Privacy advocates have objected to these Google services, which show men leaving strip clubs, protesters at an abortion clinic, sunbathers in bikinis and in general people engaging in activities visible from public property in which they do not wish to be seen publicly [16]. This point corresponds to the concept of Informational Privacy [15] which deals with the fact that a person wants to be in control of his personal information. Often the publication of this kind of images becomes a privacy intrusion which corresponds to a situation in which personal information is collected or disseminated without consent of the person involved or represented in it [11]. Google over the years has addressed this issue by blurring the people's faces and removing or modifying pictures on request. Besides the fact that not all the faces have been correctly identified, it is easy to trace a person's identity basing on his clothing, personal items, hairstyle, the place and time where the photo was taken. Consequently, nowadays, we are not yet able to provide a reliable service to ensure the privacy in this type of images.

The goal of this study consists in opening the way to a technique, to the best of our knowledge, not yet explored, concerning the anonymization of pedestrians within public images by replacing them with images of pedestrians taken from a controlled and authorized dataset composed of pedestrians in different poses. This dataset can be constructed artificially or built using people who have given their permission for the public usage of their images. For simplicity, in this study we built a dataset composed of pedestrians taken from Google Street View images.

Using the proposed method it is possible to perform a complete replacement of the pedestrian keeping a life-like picture and at the same time ensuring a complete anonymity of the photographed subjects. An example of the application of this technique is shown in Figure 1.

In order to make a pedestrian anonymous, the preliminary step consists in its detection. The main used techniques were developed for Collision Avoidance Systems and Driver Assistance Applications and in recent years they are growing very quickly due to industrial interests [8, 5]. The main approaches are based on dedicated hardware capable of collecting infrared, stereo or multiple view images [8]. In this study we decided to work with single 2D images as a case study, with the possibility to extend it to other types of images in the future. The use of 2D images makes the detection task very competitive because there is not any additional spatial information about where the objects are and what their general size is. Instead, for the techniques which focus only on the anonymization of the face it is sufficient to identify the faces within images [19]. In both cases, once detected the pedestrian or his face, these techniques blur the pedestrian's face or remove him completely reconstructing the area that the pedestrian had occluded.

The blurring of the pedestrians' face is the main technique used to make the Google Street View images anonymous, it is very fast to compute and false positives make artifacts that are not too disturbing for the user [7]. As we mentioned above, despite the false positives, this technique does not guarantee a full anonymization, because it is possible to trace the identity of a person from other information.

A very promising way to make a pedestrian anonymous, which is showing good results, is the Inpainting technique [13], which consist in to remove foreground objects from a scene replacing their covered area with an estimation of the surrounding background. This technique was initially created for the restoration of pictures and recently has been used for the removal of pedestrians with interesting results [6, 12]. This technique, however, has severe limitations concerning the area to be rebuilt and it is sometimes impossible to properly reconstruct all the scene occluded by a pedestrian. Therefore, in order to reconstruct this information, it is possible to use stereo images or multiple view images, but also with this type of images it may occurs unfortunate occlusions that prevent to retrieve the entire occluded area. Moreover, using this technique, the removal of the shadows that often are on the sidewalk remains a challenging problem, an example is showed in Figure 4. So even the way of the inpainting suffers of major limitations.

2. Proposed Method

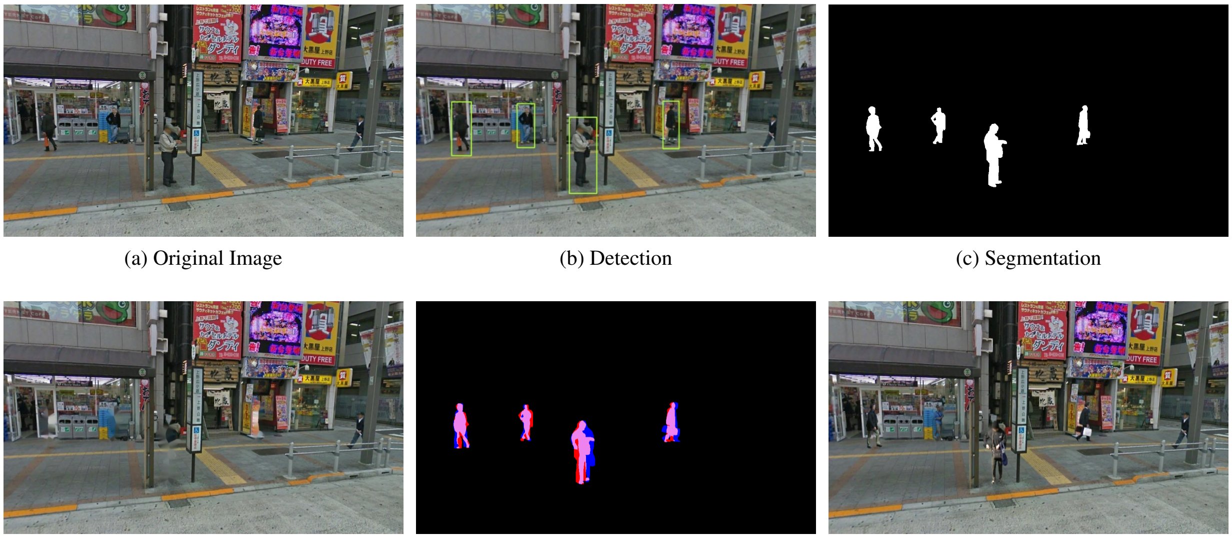
For the aforementioned reasons we propose a new method which is based upon four steps: firstly we use a people detection algorithm based on Histogram of Oriented Gradients (HOG) [4], in the second step we segment the pedestrian from the background using a multi-neural network approach [1], in the third step for each segmented pedestrian we find the most similar one in the dataset of controlled pedestrians and subsequently we replace it using a set of techniques of Image Matting and Composition [18].

We chose the HOG as a feature descriptor because it is considered the state of the art in pedestrian detection and its performance and its computational speed on our datasets have given very good results, an example is showed in Figure 1b.

After have detected the pedestrians, a crucial step consists in the segmentation of the subjects from the background. To perform this step we used our recently published algorithm, called MNOS [1], based on a Multi-Neural Networks approach, which has not yet been applied to the segmentation of people, but it is highly configurable and is giving very good results. Due to the many poses of the pedestrians, their personal items which may be of different shapes and the occlusions with other pedestrians, this task is considered very difficult. For this reason we have restricted the problem considering not crowded scenes and pedestrians carrying only small objects like bags or backpacks and not for example luggage, bicycles or strollers, cases which require a deep analysis on the human pose estimation.

The output of the MNOS is a segmentation mask, showed in Figure 1c, which we used to perform the inpainting process, Figure 1d. In this way we can subsequently replace the pedestrian with another one, without worrying about any area which remains uncovered into the replacement process, because it will be filled with an estimation of the surrounding background by the inpainting phase.

Our purpose at this step is to find a pedestrian from the controlled dataset, which best overlaps with the



(a) Original Image (b) Detection (c) Segmentation

(d) Inpainting (e) Replacing (f) Final Result

Figure 2: A successful pedestrians anonymization in a complex scene using the method proposed in this study. In the original image (a) the pedestrians are detected and these regions of interest (b) are processed by a segmentation step (c). This resulting information is used in the inpainting phase in which the pedestrian to replace is removed and the area compensated with the background (d). Consequently the best overlapping pedestrians are chosen (e) to obtain the final result (f).

pedestrian to be replaced and at the same time we want that it has a similar pose, but not necessarily the same, because our goal is to fill at the best the area which was covered by the removed pedestrian. To do that we used the Chamfer distance [10], which has already been used as a distance between pedestrians images, showing that it is fast to compute and robust to minor occlusions. Given $E(a, b)$ the Euclidean distance between two points a, b and two images I and J the Chamfer Index C is computed as

$$C(I, J) = \sum_{i \in I} E(i, \arg\min_{j \in J}(E(i, j)))$$

where i is an edge point in I and j in J.

In order to replace a pedestrian with another one, we have to deal with problems of scale, artifacts derived from the fusion between foreground and background, lighting, color balance, etc. To address these issues we used a sampling-based approach of Image Matting [17] which performs well with medium resolution images, as in our study case, to extract a subject from an image. The MNOS algorithm produces a map that we used as a Trimap to tune the α value and to estimate the size and the position of the replacing pedestrian.

To insert the chosen pedestrian without the introduction of artifacts due to the difference between its background and the background of the replaced pedestrian, we used a Compositing approach [3]. We applied a linear interpolation using

$$C = (1 - \alpha)B + \alpha F$$

where α is provided by the segmentation algorithm, C is the final color, B is the background color and F the foreground color. The result of this process is showed in Figure 1f.

3. Experiments

As additional contribution in this study, we built two datasets used in the experimental phase and uploaded to our homepage in order to be used by future methods for comparison [1]. The first one is called Google Street View Scenes (GSVS) and consists of 200 cityscape images, the second is the Controlled Pedestrians Dictionary Dataset (CPDD) consisting of 100 pedestrian images used for the replacement and we assume, for simplicity, that every pedestrian has given his authorization to use his images in public scenes. All the images, from both the datasets, are taken from Google Street View and manually segmented.

In order to evaluate the performance of the detection algorithm, we tested it on the GSVS dataset obtaining the result showed in Table 1.

[1] `http://www.dicom.uninsubria.it/arteLab/`

We tested the segmentation performance of the MNOS algorithm using the evaluation metrics presented in [1] showing the results in Table 1.

We propose a supervised evaluation method in order to compare shape matching techniques on this particular problem focused on the overlapping area, from a quantitative point of view. We ranked the results returned by the matching algorithm in according to the Chamfer distance and to properly evaluate the matching phase we used a well-know method in Information Retrieval called Normalized Discounted Cumulative Gain ($nDCG$) [9], Equation 2. This is computed from the Discounted Cumulative Gain (DCG), Equation 1.

$$DCG_p = \sum_{i=1}^{p} \frac{2^{rel_i} - 1}{\log_2{(1+i)}} \tag{1}$$

given P the dataset of pedestrian, p is the number of retrieved results to evaluate, i is the rank position of a specific pedestrian in the retrieved results, rel_i is its relevance value (in our case $rel_i \in \{1, |P|\}$ such that $rel_i = 1$ is the least relevant pedestrian and $rel_i = |P|$ is the most relevant one).

The $nDDG$ takes also in consideration the Ideal Discounted Cumulative Gain ($IDCG$) which corresponds to the ideal ranking. Given that our goal is to replace a pedestrian with another one that best overlaps, our ideal ranking is in function of the covered area estimated using the Jaccard Index [14] applied to the truth segmentation masks as a measure defining the value between the intersection on the union of two segmentation areas, in our case the area of the pedestrian that we have to replace and the area of the pedestrian extracted from the controlled dictionary. An example is showed in Figure 1e, where the shape of two pedestrians are represented in red and blue and their overlapping area in violet.

$$nDCG_p = \frac{DCG_p}{IDCG_p} \tag{2}$$

Using the Chamfer distance, the $nDCG$, whose value is in $[0, 1]$ where 1 is the best result, is reported in Table 1. We used the Chamfer distance, in the matching phase, because it needs only the image and not necessarily the mask, which we used only for the evaluation, whose generation is time consuming, so in a real application it is possible to build arbitrarily large datasets. The inpainting phase is closely related to the segmentation step and therefore can only be evaluated considering the segmentation results.

After having identified the most similar pedestrian to the one to be replaced, it is separated from the background using the segmentation mask and, after a resizing according to the pedestrian to replace and a registration on the center of mass of the two pedestrians, overwritten on the scene. We used a sampling-based approach to overcome the introduction of artifacts due to the different backgrounds where were being the two pedestrians [18]. To evaluate the replacing step we used the Jaccard Index reporting the results in Table 1.

(a) Original (b) Processed (c) Original (d) Processed

Figure 3: Some lighting errors due to the different light conditions of the original scene and the scene belonging to the replacing pedestrian.

For each phase we have provided an evaluation method and in order to evaluate the overall accuracy of the proposed method we used the Jaccard Index on the output images of all the steps, we also report the computational time for each step using a single C# thread code, on a Intel®Core ™i5 CPU at 2.30Ghz, these results are showed in Table 1.

A successful example of the whole process in a scene with many pedestrians is shown in Figure 2. Nonetheless, the experiments show the limits of this application, which consist in the management of crowded and complex scenes, which can be the starting point for more detailed investigations into the issue presented in this study. For example the most noticeable errors are due to the pedestrians occlusions with the scene, such as a fire hydrant, a street light, an advertising billboard, etc. which can be overcome automatically selecting only the appropriate pedestrian parts to be replaced. An unsolved problem consists in managing of light effects and illumination consistency, in other words, the replacing pedestrian must be subjected to the same light conditions of the replaced one in order to maintain the appearance of the scene realistic, artifact that can be noticed in Figure 3. In addition, a last aspect is to replace a pedestrian with another one that has no only a similar shape, but who is moving also in the same direction, so as to avoid paradoxical situations, such as a pedestrian who is walking towards a wall, a street pole or another pedestrian.

(a) Original (b) Processed

Figure 4: An example of the successful management of shadows by the proposed method.

Table 1: "Segmentation", "Matching" and "Replacing" steps are evaluated on a single pedestrian of the CPDD, "Detection" and "All Phases" on the GSVS composed by an average of 1.6 pedestrians for each scene.

Phase	Evaluation Method	Result	Time
Detection	F-measure	0.76	5.5s
Segmentation	Single Class Accuracy	0.75	3.8s
Matching	$nDCG_5$	0.71	4.2s
Replacing	Average Jaccard Index	0.68	0.3s
All phases	Average Jaccard Index	0.47	21.4s

4. Conclusions

In this study we have presented an innovative method to ensure digital privacy in cityscape pictures based on the replacement of pedestrians, in particular we worked on Google Street View images. The proposed method produces a completely anonymous image and at the same time does not present the artifacts, like blurring areas, typical of other techniques used today to manage this privacy issue. As a contribution in this study, we provided two datasets, a baseline for future works and an evaluation method to measure the goodness of this type of replacing approach. Given the promising results in this field and the growing interest in digital privacy, we are working on the application of this technique in images of crowded scenes, occluded pedestrians, cyclists etc. in order to overcome the constrains of the proposed method.

References

[1] S. Albertini, I. Gallo, M. Vanetti, and A. Nodari. Learning object segmentation using a multi network segment classification approach. In *International Conference on Computer Vision Theory and Applications*, 2012.

[2] D. Anguelov, C. Dulong, D. Filip, C. Frueh, S. Lafon, R. Lyon, A. S. Ogale, L. Vincent, and J. Weaver. Google Street View: Capturing the World at Street Level. *IEEE Computer*, 43:32–38, 2010.

[3] J. F. Blinn. J. Blinn"s Corner: Compositing Part 1: Theory. *IEEE Computer Graphics and Applications*, 1994.

[4] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition*, volume 1, pages 886–893, 2005.

[5] M. Enzweiler and D. M. Gavrila. Monocular Pedestrian Detection: Survey and Experiments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31:2179–2195, 2009.

[6] A. Flores and S. Belongie. Removing pedestrians from Google street view images. In *Computer Vision and Pattern Recognition*, 2010.

[7] A. Frome, G. Cheung, A. Abdulkader, M. Zennaro, B. Wu, A. Bissacco, H. Adam, H. Neven, and L. Vincent. Large-scale privacy protection in Google Street View. In *International Conference on Computer Vision*, pages 2373–2380, 2009.

[8] D. Gernimo, A. M. Lpez, A. D. Sappa, and T. Graf. Survey of Pedestrian Detection for Advanced Driver Assistance Systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32:1239–1258, 2010.

[9] K. Jarvelin and J. Kekalainen. Cumulated gain-based evaluation of ir techniques. *ACM Transactions on Information Systems*, 20:422–446, 2002.

[10] I. Katz and H. K. Aghajan. Multiple camera-based chamfer matching for pedestrian detection. In *ICDSC*, pages 1–5, 2008.

[11] B. Rossler and R. Glasgow. *The value of privacy*. Polity Press: Cambridge, 2004.

[12] T. Shibata, A. Iketani, and S. Senda. Fast and structure-preserving inpainting based on probabilistic structure estimation. In *IAPR Conference on Machine Vision Applications*. ACTA Press, 2011.

[13] T. K. Shih and R. C. Chang. Digital inpainting survey and multilayer image inpainting algorithms. In *International Conference in IT Application*, 2005.

[14] P.-N. Tan, M. Steinbach, and V. Kumar. *Introduction to Data Mining*. Addison Wesley, 2005.

[15] H. Tavani. *Ethics and Technology: Controversies, Questions, and Strategies for Ethical Computing*. John Wiley & Sons, 2010.

[16] J. Thornton. Individual privacy rights with respect to services such as Google Street View. pages 70–76, 2010.

[17] J. Wang and M. F. Cohen. Image and Video Matting: A Survey. *Foundations and Trends in Computer Graphics and Vision*, 3:97–175, 2007.

[18] J. Wang and M. F. Cohen. Simultaneous Matting and Compositing. In *Computer Vision and Pattern Recognition*, 2007.

[19] C. Zhang and Z. Zhang. A survey of recent advances in face detection, 2010.