FOREGROUND AND SCENE STRUCTURE PRESERVED VISUAL PRIVACY PROTECTION USING DEPTH INFORMATION

Semir Elezovikj[1] *Haibin Ling*[1] *Xiufang Chen*[2]

[1]Department of Computer and Information Science, Temple University, Philadelphia, PA USA
[2]College of Education, Rowan University, Glassboro, NJ USA
{semir.elezovikj, hbling}@temple.edu, chenx@rowan.edu

ABSTRACT

In this paper, we propose the use of depth-information to protect privacy in person-aware visual systems while preserving important foreground subjects and scene structures. We aim to preserve the identity of foreground subjects while hiding superfluous details in the background that may contain sensitive information. We achieve this goal by using depth information and relevant human detection mechanisms provided by the Kinect sensor. In particular, for an input color and depth image pair, we first create a *sensitivity map* which favors background regions (where privacy should be preserved) and low depth-gradient pixels (which often relates a lot to scene structure but little to identity). We then combine this per-pixel sensitivity map with an inhomogeneous image obscuration process for privacy protection. We tested the proposed method using data involving different scenarios including various illumination conditions, various number of subjects, different context, etc. The experiments demonstrate the quality of preserving the identity of humans and edges obtained from the depth information while obscuring privacy-intrusive information in the background.

Index Terms— Visual privacy protection, sensitivity map, inhomogeneous diffusion

1. INTRODUCTION

Privacy protection in person-aware visual systems is an ill-posed problem: the attitude towards privacy is different with different people. With the increasing technological capability of cameras in recent years, video has become commonplace. Pervasive video surveillance systems inherently eliminate video anonymity and allow for jeopardizing identities and activities of people [8]. Google's video service YouTube has become a popular destination for videos of protest and civil disobedience as Google has made an active step towards "helping activists sidestep autocratic regimes"[12]. With the "Blur All Faces" option, users are able to automatically obscure the identity of all people in the video and they also have the option to permanently delete the original, unblurred version of the video.

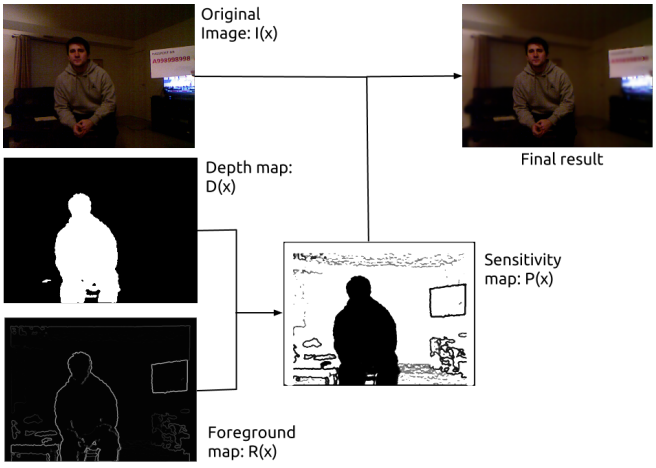

Fig. 1. The flowchart of the proposed method.

Due to its importance, privacy protection in visual data has been attracting a large amount of research, especially from the multimedia and computer vision communities. A lot of systems and algorithms have been developed in the context of visual surveillance [11], such as [4, 3, 10, 6] etc. Another group of studies focus on designing privacy obscuring algorithms that at the same time preserve data utility as much as possible, such as de-identification for human body [1], car license plate [5], faces [7, 2], etc.

In this paper we focus on achieving privacy protection in person-aware visual systems. The specific task can vary depending on what information from the original image content we want to privatize. We can obscure the background of a scene and maintain the identity of a person in order to protect the privacy of the surroundings. An example of this is a video conference where the video participants can see plenty of the surroundings with the large field of view offered by modern video conferencing cameras. Privacy in this scenario would mean blurring the background. The reason we would want to do this is to protect potentially sensitive information or simply not wanting to disclose the environment. Another type of privacy would be the privacy where the identity of the people is protected.

In particular, we aim to privatize potential privacy-

intrusive information in the surroundings of the foreground subject of interest. Intuitively, we need to distinguish two kinds of information from private ones: (1) humans in the foreground region who need to remain clear in the visual communication tasks (e.g., video conferences) and (2) geometric structure information (e.g. depth edges) that relates a lot to scene understanding but little to identity. Towards the above goal, we make two main contributions in this paper:

- *Sensitive information inference via depth information.* We propose to use depth information to estimate a *sensitivity map* to measure per-pixel privacy sensitivity. With the help of modern depth sensor, we can identify the image regions for foreground subjects. In addition, we can also derive the depth edges from the gradient magnitude of the depth map. These two strategies are integrated into the estimation of the sensitivity map.
- *Preservative privacy protection using inhomogeneous obscuration.* We model the privatization with an inhomogeneous diffusion process. Specifically, we adopt the Perona-Malik diffusion [9] framework and associate the sensitivity map with the rate of diffusion factor. Intuitively, the more sensitive a pixel is, the larger diffusion kernel is applied to obscure it.

The flowchart of the proposed method is summarized in Figure 1. Note that, while using person-aware visual communication as the application context, our technology can be easily extended to other application scenarios, since both the sensitivity map and the inhomogeneous obscuration techniques are general and easy to accommodate different requirements of different privacy definitions.

2. VISUAL PRIVACY PROTECTION USING DEPTH INFORMATION

2.1. Data Preparation

To get simultaneously depth and color images, the Microsoft Kinect Sensor is used in our study. Kinect provides a depth image camera and a functionality for real-time human-background disambiguation, allowing us to separate humans from the background in a cluttered environment with varying lightning conditions and camera angles. We start by retrieving a color image frame and a corresponding depth frame from the Kinect Sensor, both in the default resolution of 640×480 at 30 frames per second. The SDK provides methods for aligning the color frame and the depth frame. The SDK also performs human detection and returns the result as a player index map, where each pixel has an associated player index or 0 for non-player. Using the player indices, we are able to separate the background objects from human objects.

After some simple calibration steps using information provided by the sensor, we finally have three inputs from the Kinect sensor: color image $I(x)$, depth map $D(x)$ and player index map $Q(x)$.

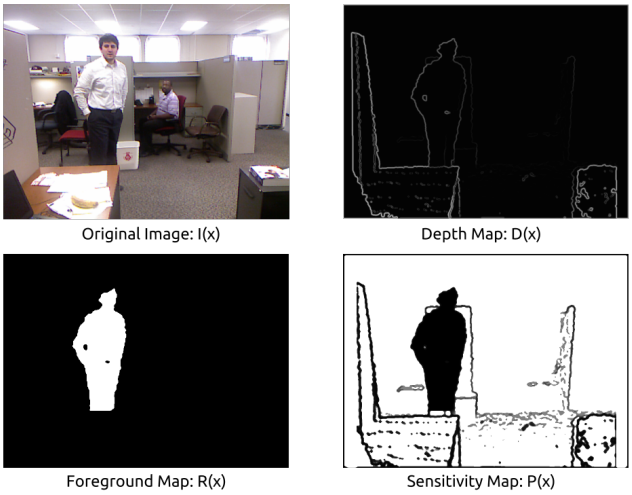

Fig. 2. Example of a sensitivity map

2.2. Sensitivity map

The sensitivity map, as described previously, should capture information from two sources: the foreground information and the privacy-irrelevant structure information.

We derive the foreground information from the player index map by treating all players (within a certain depth threshold) as foreground. This way, we create a binary region-of-interest (ROI) map $R(x)$ defined as:

$$R(x) = \begin{cases} 1 & \text{if } Q(x) > 0 \text{ and } D(x) < \tau \max_x(D(x)) \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where τ is a threshold set to 0.5. For scene structure information, we measure it by measuring the depth change, which is captured by using the magnitude of the depth gradient, i.e., $\|\nabla D(x)\|$. Note that, theoretically, second order derivatives also reflect depth discontinuity. In practice, we find that the first order gradient performs very well already and therefore discard the second order statistics for computational efficiency.

Combining the two components, we define the sensitivity map, $P(x)$, as follows:

$$P(x) \propto f\left(\frac{1 - R(x)}{\|\nabla D(x)\| + \epsilon}\right),$$

where ϵ is a small positive number to avoid underflow, $f(.)$ is a sigmoid-like function to avoid extreme values. Figure 2 shows an example of a sensitivity map.

2.3. Inhomogeneous Privacy Obscuration

The idea to use sensitivity map for privacy protection is to encourage large obscurity in the areas of large sensitivity, while keep non-sensitive regions as un-altered as possible. We implement the idea using the inhomogeneous diffusion framework, and use the sensitivity map to guide the diffusive factor. In particular, we formulate the problem as the following

diffusion process:

$$\partial_t I(x,t) = \text{div}\left(g(I(x,t), P(x))\, \nabla I(x,t)\right)\ ,\ t > 0, \quad (2)$$

where $g(.)$ defines the diffusivity and t is the "artificial" time step for the diffusion process.

We base our implementation on the classical Perona-Malik diffusion [9] on each of the three channels of the color image (red, green, blue) in combination with the sensitivity map. In [9], two flux functions which will help limit the diffusion process to contiguous homogeneous regions and not cross region boundaries are introduced:

$$c(\|\nabla I\|) = \exp\left(-\|\nabla I\|^2/K\right), \quad (3)$$

$$c(\|\nabla I\|) = \left(1 + \|\nabla I\|^2/K\right)^{-1}. \quad (4)$$

The first flux equation prefers high-contrast edges over low-contrast ones. The second flux functions prefers wide regions over narrow ones. In our paper, we are using the flux function in (3) since it empirically demonstrates better performance. To combine the flux function with the sensitivity map $P(x)$, we define our diffusivity function as

$$g(I(x,t), P(x)) = c(\|\nabla I\|)P(x). \quad (5)$$

For implementation, the original P-M diffusion iteratively updates the image and at iteration $t+1$ the numerical solution is:

$$I_{i,j}^{t+1} = I_{i,j}^{t} + \lambda[c_N \cdot \nabla_N I + c_S \cdot \nabla_S I + c_E \cdot \nabla_E I + c_W \cdot \nabla_W I]_{i,j}^{t},$$

where, $N, S, W,$ and E correspond to the pixel above, below, to the left and to the right of the pixel that we are currently processing. The update parameter λ affects how much smoothing happens in each iteration. The algorithm can be stable if λ is in the range of $[0, 0.25]$ as suggested in [9].

To combine with the sensitivity map $P(x)$, we revise the iteration as following:

$$I_{i,j}^{t+1} = I_{i,j}^{t} + \lambda \cdot P_{i,j}^{t}[c_N \cdot \nabla_N I + c_S \cdot \nabla_S I + c_E \cdot \nabla_E I + c_W \cdot \nabla_W I]_{i,j}^{t}.$$

The resulting Perona-Malik diffusion on the original color image will have a smoothing factor of 0 when the pixel has a player index other than 0 associated with it, and a small value at the edges calculated based on the gradient of the distance map $D(x)$. Looking at the sensitivity map, we can infer knowledge about the granularity of regions. This way we preserve semantically meaningful boundaries, and we use those boundaries to achieve intra-region smoothing in homogeneous regions, succeeding in obscuring potentially sensitive information in the background. In order to achieve high level of privacy and completely obscure the background we use 30 iterations and the smoothing factor in each iteration, λ is set to 0.25, which when combined with $P(x)$ will be in the range between 0 and 0.25.

3. EXPERIMENTAL RESULTS

We use different scenarios to evaluate our approach to visual privacy protection. The proposed algorithm in this paper has been tested in various conditions, e.g., having multiple people on the scene, the person of interest being turned with the back towards the camera, scenarios where the people of interest are seating or standing as well as testing scenarios under different angles. A number of examples are shown in Figure 3.

In Figure 3, example (a) shows a scenario with multiple people. The people of interest are not necessarily in the "foreground" depth-wise since there are objects that are closer. We can see from the sensitivity map that the region that we want to maintain $R(x)$ is precisely identified, and that the edges based on depth gradient are properly identified as well. The last image in example (a) is the result of Perona-Malik diffusion in combination with the sensitivity map. Compared to Perona-Malik diffusion without using our sensitivity map, we can see that our results preserve humans and maintain the strength of the edges, while smoothing heavily inside the homogeneous regions. The information on the whiteboard is successfully obscured as well.

Example (b) shows the performance of our algorithm in a scenario with varying lightning conditions. Given that the semantic boundaries are calculated from the depth information, the boundaries are successfully retrieved regardless of the small variety of color in the scene. The "player" region is successfully retrieved as well and the image that is protected using the sensitivity map demonstrates how potential sensitive information in the scene is successfully protected. In this case, we have a paper with information about a passport number. The sensitivity map demonstrates that the edges of the paper are preserved, while there is an equal level of intra-region smoothing as the image that has Perona-Malik diffusion performed on the image directly, with no information about the privacy level.

Examples (c) and (d) show a user standing and sitting sideways, and we can see that the sensitivity map is properly calculated, helping achieve good privacy protection when combined with Perona-Malik diffusion.

Examples (c) and (d) also demonstrate the performance of our algorithm from different distances. In (c), the player is near the edge of the image. We can see from the sensitivity map that the edges are successfully detected and the player region as well. The resulting privacy protected image is a good example of an image that is successful at protecting the privacy, but also maintaining the visual semantics of the scene. Example (c) demonstrates that the player is successfully retained even after being further away in the scene with his back towards the sensor.

4. CONCLUSION

In this paper we presented a new privacy protection method for visual data using depth information acquired from depth

Fig. 3. Results for privacy protection based on depth information in various scenarios. From right to left: original image, Perona-Malik diffusion on original image, sensitivity map, Perona-Malik diffusion in combination with sensitivity map

sensors like the Kinect. The depth information helps to distinguish non-private foreground regions and scene structures, though human detection and depth gradients. Such information is used to create a sensitivity map to define per-pixel privacy levels. The map is then combined with an inhomogeneous diffusion process to achieve privacy protection. We evaluated the proposed approach using a set of RGBD images containing various image conditions and scenarios. The results show that our method successfully obscures potentially privacy-intrusive information, while at the same time maintaining the non-private visual features.

Acknowledgement. This work is supported in part by NSF Grant IIS-1218156.

5. REFERENCES

[1] P. Agrawal and P. Narayanan, "Person de-identification in videos," *IEEE T. CSVT*, 21(3):299–310, 2011.

[2] D. Bitouk, N. Kumar, S. Dhillon, P. N. Belhumeur, and S. K. Nayar, "Face Swapping: Automatically Replacing Faces in Photographs," *ACM SIGGRAPH*, 2008.

[3] T. E. Boult, "PICO: Privacy through invertible cryptographic obscuration," in *Computer Vision for Interactive and Intelligent Environment*, 2005.

[4] Y. Chang, R. Yan, D. Chen, and J. Yang, "People identification with limited labels in privacy-protected video," *ICME*, 2006.

[5] L. Du and H. Ling, "Preservative license plate de-identification for privacy protection," *ICDAR*, 2011.

[6] 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," *ICCV*, 2009.

[7] R. Gross, L. Sweeney, F. D. la Torre, and S. Baker, "Semi-supervised learning of multi-factor models for face de-identification," *CVPR*, 2008.

[8] Martnez-Ballest, Antoni, "Towards a trustworthy privacy in pervasive video surveillance systems", *Pervasive Computing and Communications Workshops,* 2012.

[9] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion", *IEEE T. PAMI*, 12(7):629-639, 1990.

[10] J. Schiff, M. Meingast, D. K. Mulligan, S. Sastry, and K. Y. Goldberg, "Respectful cameras: detecting visual markers in real-time to address privacy concerns," *IROS*, 2007.

[11] A. Senior, *Protecting Privacy in Video Surveillance*, in *Privacy Protection in a Video Surveillance System*, 2009.

[12] Guardian Unlimited, "Google introduces face-blurring to protect protesters on YouTube", 2012.