Privacy-Protected Camera for the Sensing Web

Ikuhisa Mitsugami[1], Masayuki Mukunoki[2], Yasutomo Kawanishi[2], Hironori Hattori[2], and Michihiko Minoh[2]

[1] Osaka University, 8-1, Mihogaoka, Ibaraki, Osaka 567-0047, Japan, {mitsugami}@am.sanken.osaka-u.ac.jp,
[2] Kyoto University, Yoshida-Nihonmatsu, Sakyo, Kyoto 606-8501, Japan, {mukunoki,ykawani,hattori,minoh}@mm.media.kyoto-u.ac.jp,
WWW homepage: http://mm.media.kyoto-u.ac.jp/sweb/

Abstract. We propose a novel concept of a camera which outputs only privacy-protected information; this camera does not output captured images themselves but outputs images where all people are replaced by symbols. Since the people from this output images cannot be identified, the images can be opened to the Internet so that we could observe and utilize the images freely. In this paper, we discuss why the new concept of the camera is needed, and technical issues that are necessary for implementing it.

1 Introduction

In these days, many surveillance cameras are installed in our daily living space for several purposes; traffic surveillance, security, weather forecast, etc. Each of these cameras and its captured video are used only for its own purpose; traffic surveillance cameras are just for observing congestion degree of cars, and security cameras are just for finding suspicious people. The video may include various information other than that for the original purpose. If the video is shared among many persons through the Internet, the camera will become more convenient and effective. For example, we could get weather information from a traffic surveillance camera, congestion degree of shoppers in a shopping mall from a security camera, and so on. Considering these usages, we notice the usefulness of opening and sharing real-time sensory information on the Internet. The Sensing Web project[1, 2], which was launched in the fall of 2007, proposes to connect all available sensors including the cameras to the Internet, and to open the sensory data to many persons in order that anyone can use the real-time sensory data for various purposes from anywhere.

On opening and sharing the sensory data, the most serious problem is privacy invasion of observed people. As long as the sensory data is closed in a certain system operated by an institution in the same way as most existing systems, the privacy information can be managed and controlled by the corresponding institution. We thus do not need to take care of the problem. On the other hand, in the case of the Sensing Web, the sensory data is opened to the public so that anyone can access any sensory data without any access managements.

Fig. 1. The output of a traditional security camera and privacy-protected camera.

Especially, the video, which is the sensory data obtained by the cameras, contains rich information of the people and may cause the privacy invasion. In fact, a person in a video can be easily identified by his/her appearance features (face, motion, colors of cloths, etc.). The privacy invasion problem, therefore, has been a main obstacle against opening the sensory data.

In the Sensing Web project, we tackle the problem to realize an infrastructure where any sensory data are opened and shared. To overcome the problem, the privacy information has to be erased from the image before it is opened to the Internet. One of the ways to realize this privacy elimination is to mask the appearances of the people on the image. In fact, the Google Street View (GSV)[3] adopts this approach. Though this service offers not real-time sensory data but just the snapshots at a past moment, it faces the same problem as mentioned above. To overcome this problem, each person in the captured image is detected and masked automatically in the case of the GSV. This operation can be executed using a human detection technique. However, as the technique does not works perfectly, some people cannot masked correctly and their privacy accordingly cannot be protected; when a person is detected in a wrong position or not detected, the mask is overlaid on a wrong position or is not overlaid at all, and as a result the person is left unmasked and clearly appeared in the output image. We thus propose another approach to overcome the privacy invasion problem based on a novel idea; the image of the camera is reconstructed by generating a background image without any people, and overlaying symbols at the positions of the corresponding people on the generated image. This idea is implemented as a new concept of the camera that we call a "Henshin" camera , which means a privacy-protected camera (Fig.1). In the case of this camera, even if the human detection does not work well, it just causes the rendering of the symbol on the different position or the lack of the character, but never causes the privacy invasion.

For realizing this privacy-protected camera, we need two techniques; a human detection and a background image generation. The former one has been studied for a long time, and there are a lot of existing studies. They are mainly categorized into two types of approaches; background subtraction methods[4, 5] and human detection methods[6, 7]. In this paper, we use a HOG-based human detection[7], which is known as a method that works robustly even when the luminosity of the scene changes frequently. On the other hand, the latter one has to be well considered. Although it looks just a conventional issue at a glance, it is indeed much different from the many existing methods for the background generations. Considering the concept of the privacy-protected camera, we have to design the background generation method ensuring that people would never appear in the output image even if a person stops for quite a long time in the scene, which is treated as not the foreground person but the background by the most methods. Besides, our method has to generate the background image as verisimilar to the truth as possible, because we would like to know various information of the observed area from this background image. Especially, lighting condition by the sun is helpful to know the weather. Therefore, such kind of information has to be well reconstructed. Considering the above discussions, this paper proposes a novel background generation technique which preserves the shadow accurately in outdoor scene while ensuring that a person never appears in the image. This technique is realized by collecting the images for super long term, categorizing them by time, and analyzing them using the eigenspace method.

2 Background Generation Using Long-Term Observation

2.1 Traditional Background Generation Methods

If a human detection performs perfectly and all the people in the image thus can be erased, the whole image except the people regions can be used as the accurate background image. However, when the people exist, the corresponding regions would be left as blanks so that the background image cannot be always fully generated. In addition, there has been no ideal method for the human detection, which is apparent to see that there are still many challenges for this topic. Therefore, in terms of the privacy protection, we must not directly use each image which is observed by the camera. We have to take an analytic approach by collecting many images for a certain period of time.

Calculating median or average of each pixel of the image sequence is a simple approach to generate the background image. In order to follow the background changing, the term of the image sequence is usually not very long. However, people who stop for the term appear in the generated image, which causes the privacy invasion. For generating the background avoiding privacy invasion, we have to analyze images collected during much longer term than people might stop. On the other hand, in terms of reconstructing the background image as similar to the truth as possible especially from the viewpoint of the lighting condition by the sun, such analytic approaches with long term image sequence

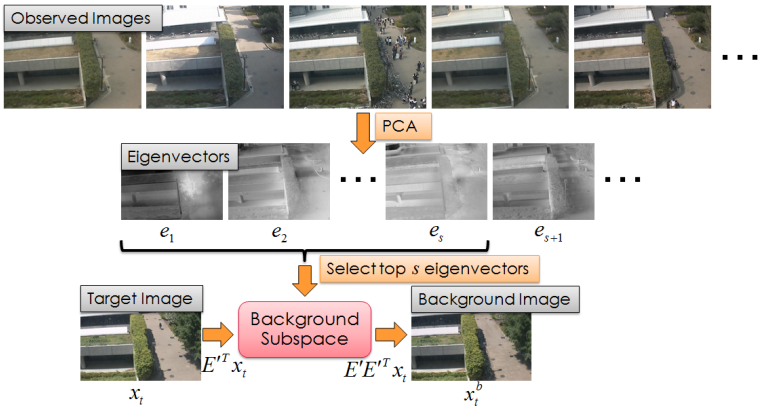


Fig. 2. The output of a traditional security camera and privacy-protected camera.

do not perform well; they cannot follow immediately the sudden and frequent changes of the strength of the sunlight, because the generated image is influenced by the many images in the past. Such approaches cannot fulfill the demand for applying to the privacy-protected camera.

The eigenspace method is often used to analyze huge amount of data. We apply this method to the images collected by long term surveillance. Using the eigenspace method[8], we can analytically reconstruct the background image from the current image captured by the camera that may contain some people. This is achieved by the following process.

First, the eigenvectors $e_1, e_2, \cdots, e_k$ (sorted in descending order by their contributions) are calculated from a number of images by the principal component analysis (PCA). As the eigenspace defined by these eigenvectors indicates the variation of the image sequence, the background image x_t^b can be estimated from observed image x_t using this eigenspace; x_t^b is calculated by the following equation:

$$x_t^b \approx Ep = EE^T x_t \tag{1}$$

where p describes the corresponding point in the eigenspace and $E = |e_1, e_2, \cdots, e_k|$ is an orthonormal matrix. We have to use the images each of which may contain people. Note that the appearance of the people should have less influence than the variation of the background, as the people are usually much smaller than the size of the image and each of them moves randomly and is observed for just a short term. Thus, even if we use such images, we can get the eigenspace which includes no influence of people by using only s ($s << k$) eigenvectors to reconstruct the background. We use the orthonormal matrix $E' = |e_1, e_2, \cdots, e_s|$ instead of E to estimate the background image x_t^b (Fig.2). x_t^b is calculated by the following equation:

$$x_t^b \approx E'p' = E'E'^T x_t \tag{2}$$

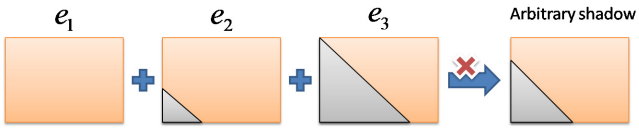


Fig. 3. Various shadow edges cannot be described as linear sum of the small number of the eigenspace.

where p' describes the corresponding point in the eigenspace. It means that even when we use the image which may contain people, we expect to get the similar result to the case using the images without any people in them when we choose only such the small number of the eigenvectors.

Nevertheless, another problem still exist; the lighting condition may not be able to reconstructed by such the small (s) dimensional eigenspace.

2.2 Eigenspace Method with Classification by Observed Time

In outdoor scene, there are sharp shadow edges in the observed images and the position of the shadow edges are shifted gradually caused by the solar position. When we generate the background which includes various shadows in the scene, we need eigenvectors in the very high dimensional eigenspace which correspond to each position of shadow edges. When the dimension s is small as discussed in the previous section, such the eigenvectors corresponding to the moving shadow edges may be neglected. We, therefore, could not generate the background image keeping such various shadow edges by linear sum of top eigenvectors (Fig.3).

Our method relies on the fact that the shadows appear in the similar position in the same time even in another day. We collect huge amount of images by super long term surveillance and classify them into image sets according to their capture time. The images of each set are expected to have similar spatial appearance of shadows. Fig.4 shows the shadow appearances. Looking at the images observed in a day, there are shadows in the different position. On the other hand, we can see that the shadows appear in the similar position when we look at the images observed in 15 o'clock of different days.

Thus, the procedure of the proposed method is as follow. First, at a time t of a day d, we get a target image $x_{d,t}$. We then classify the observed images $x_{d,t}$ according to their observed time t, and we can get an image set $x_{d-1,t}, x_{d-2,t}, \cdots$ which were observed at the time t of the different days $d-1, d-2, \cdots$. We refer each of the image sets as I_t. Finally, we apply the eigenspace method to the image set I_t, and then we can get the background keeping lighting conditions for each target image $x_{d,t}$ which is a raw image that may contain some people in it.

To show the effectiveness of our method, we experimented in some outdoor scenes. We generated the background images by the simple eigenspace method and the proposed method. We used the following two scenes:

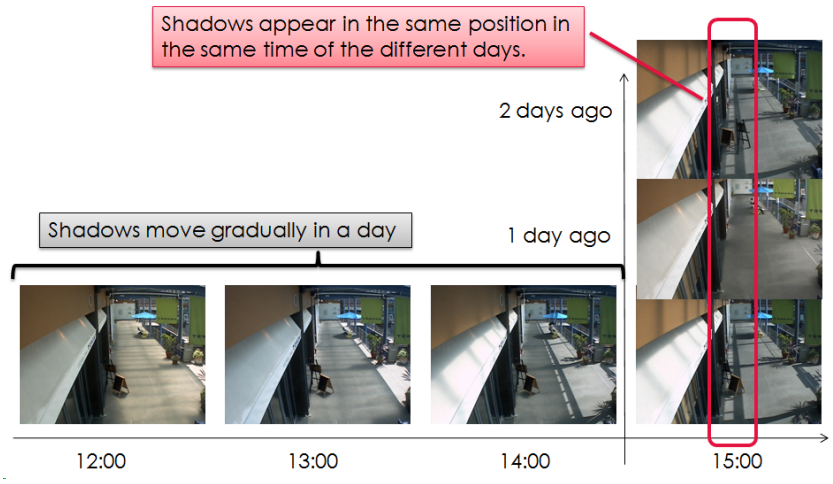


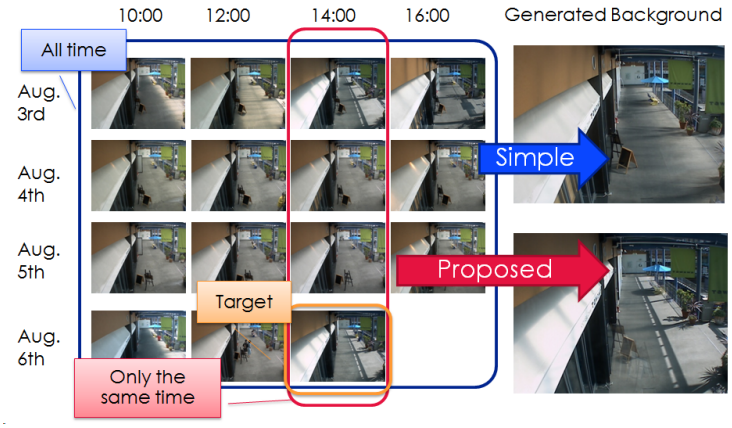
Fig. 4. Examples of the shadow positions. Shadow moves gradually in a day, and appears in the similar position in the images which were observed at the same time of the different days.

- Scene 1: The input images and the results are shown in Fig.5(a). The images are collected from 1st to 31st in August at a shopping mall.
- Scene 2: The input images and the results are shown in Fig.5(b). The images are collected from 1st to 31st in August at our university.

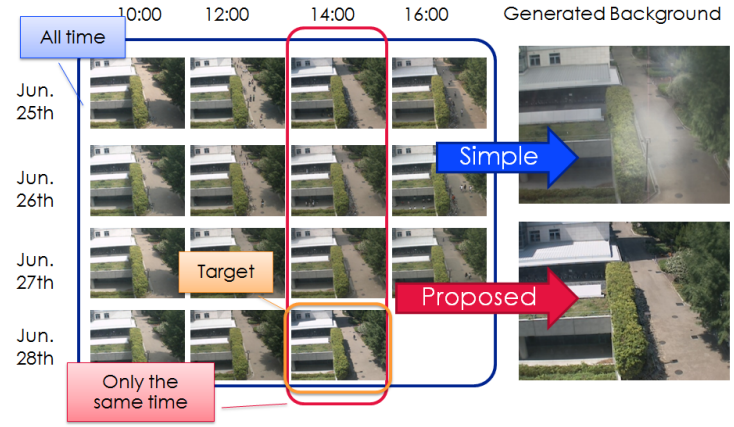
For each scene, to generate background images of a certain day we applied the two methods: the simple eigenspace method which uses all the images for calculating the eigenspace, and the proposed method which first classifies the images into the sets according to the time and then applies the eigenspace method to the set. For the proposed method, we used images observed in 15 seconds around 14:00 of everyday. Comparing those two results, it is visually confirmed that the proposed method can generate better background image than the traditional one from the viewpoint of keeping the lighting condition.

3 Scene-Adaptive Human Detection

For realizing this privacy-protected camera, a human detection method as well as the background image generation mentioned in the previous section is important. For this purpose, the HOG-based method[7] is currently known to show good performance. This paper therefore fundamentally relies on this method, but proposes a scene adaptation framework to modify its performance.



(a) Scene in a shopping mall.



(b) Scene in our university.

Fig. 5. Examples of the background image generation.

3.1 HOG-based Human Detection Method

First, we introduce the HOG descriptor defined in [7]. For a human image, local appearance and shape within the image are described by the distribution of intensity gradients. The descriptors can be calculated by dividing the image into small connected regions, which are called cells, and compiling a histogram of gradient for the pixels within each cell. The local histograms are then combined and normalized across a larger region of the image, which is called a block. The HOG descriptor denotes the combination of the histograms in the whole image. This descriptor is known to show good invariance to changes in illumination or shadowing.

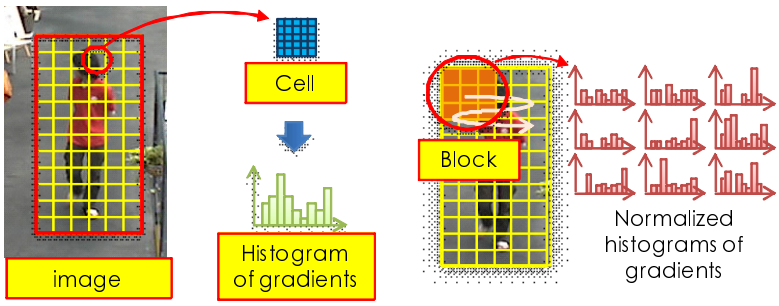


Fig. 6. The HOG descriptor.

The method then applies supervised learning technique to judge whether an image is of a human or not. Binary classifiers such as the Support Vector Machine(SVM) are often used. Once trained on images containing people, the classifier become able to make decisions regarding the appearance of a human.

3.2 Scene Adaptation of HOG-Based Human Detector

Though in fact the HOG-based human detector mentioned in Section 3.1 runs quite well, it often causes the following errors; (i) it occasionally detects a false area since it might mistakenly detect a background area whose pattern is incidentally similar to human, and (ii) it sometimes spoils a human region when it fails to acquire enough features in that region. These failures are caused by the fact that the detector does not use information of the observed scene; the detector uses knowledge about human/non-human patterns contained in the images captured by other cameras and of other scenes. That is why it may mistakenly detect a pattern which is of the background but accidentally looks similar to a human, or it may sometimes spoil a human in the image.

The proposed method improves the performance by using additional information specific to each camera, which can be obtained by judging true or false detections from these results. This judgment cannot be achieved by using information only about each frame. However, it can be achieved by analyzing time series of detected results, because it is different between true and false detection. By this analysis we judge true or false detections and acquire additional information specific to each camera, and they are then used additionally in the supervised learning to update the detector, as shown in Fig.7.

The proposed method also improves how to search people in the image. The normal detector usually searches people through the whole image with all possible sizes and orientations, since it does not use any scene-specific information. On the other hand, the proposed method is designed to obtain relation between the sizes and orientations and positions of people specific in the scene by automatic camera calibration using the true detections. This modification is effective not only for the processing cost but also for the detection accuracy, because the

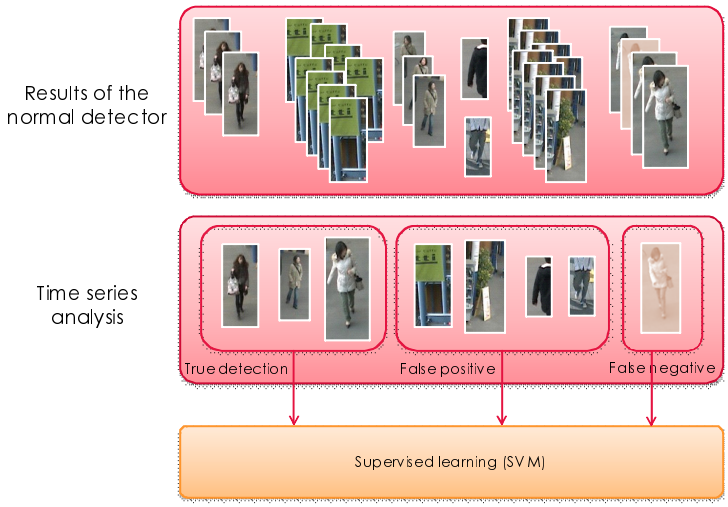


Fig. 7. Time series analysis for additional learning.

Results of the existing method.

Results of the proposed method.

Fig. 8. Experimental results.

patterns which are of the background but accidentally look similar to people can be eliminated from the candidate of the judgement.

Fig.8 shows the results of the existing and proposed methods. It is confirmed that the proposed method works more effectively; people are accurately detected throughout the image sequences while fail samples are much reduced.

4 Conclusion

In this paper, we proposed a novel concept of a camera, named the "Henshin" camera, which outputs privacy-protected information. This camera outputs only images where all people are replaced by symbols. The images thus can be opened and shared on the Internet so that we could observe and utilize the images freely.

For realizing this privacy-protected camera, we proposed a new background generation method and a human detection method. The former one was designed so as to ensure that people would never appear in the image even if the people stop for quite a long time in the scene. It was realized by collecting the images for super long term, categorizing them by time, and analyzing them using the eigenspace method. The latter one was based on the HOG human detection method but extended to adapt the detector to each scene so as to improve the accuracy of the detection. This is also realized by collecting the detected results from the past images.

Our future work contains pervasiveness of the concept of the privacy-protected and social investigation of its acceptability from the viewpoint of the privacy. The improvement of the proposed method is also one of the important topics.

References

1. Minoh, M., Kakusho, K., Babaguchi, N., Ajisaka, T.: "Sensing Web Project - How to handle privacy information in sensor data," Proceedings of International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU), 863-869 (2008)
2. Minoh, M.: "Sensing Web: Concept and Problems (¡Special Issue¿Sensing Web)" [In Japanese], Journal of Japanese Society for Artificial Intelligence (2009)
3. Google: "Google Street View," http://maps.google.com (2008)
4. Haritaoglu, I., Harwood, D., Davis, L.S.: "Hydra: Multiple people detection and tracking using silhouettes," IEEE Workshop on Visual Surveillance, 0, 6 (1999)
5. Jabri, S., Duric, Z., Wechsler, H., Rosenfeld, A.: "Detection and location of people in video images using adaptive fusion of color and edge information," International Conference on Pattern Recognition, 4, 4627 (2000)
6. Chen, Y.T., Chen, C.S.: "Fast human detection using a novel boosted cascading structure with meta stages," IEEE Transactions on Image Processing, 17, 8, 1452-1464 (2008)
7. Dalal, N., Triggs, B.: "Histograms of oriented gradients for human detection," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, 1, 886-893 (2005)
8. Oliver, N., Rosario, B., Pentland, A.: "A Bayesian computer vision system for modelling human interactions," In Proceedings of ICVS99, Gran Canaria, Spain (1999)