Object Detection under Illumination Variation with Privacy Protection

Hidenori TANAKA†, Hiroyuki ARAI†, Hitoshi NAKAZAWA†, Takayuki YASUNO†† (Member), Hideki KOIKE†
† NTT Cyber Space Laboratories, NTT Corporation
†† NTT Intellectual Property Center, NTT Corporation

Summary We propose an object detection method that aims to protect privacy while providing video surveillance even when the illumination changes. Recently, many surveillance cameras have been installed in public spaces for monitoring activities. Unfortunately, there are too few authorized people to allow all of the resulting video streams to be perused continuously. Therefore, it is necessary to allow concerned citizens to watch the video streams. To permit this, we must address the privacy issue. In the proposed method, we first extract the object regions (which include private information) using the background estimated by taking account of illumination variations. Next, we filter the extracted regions to protect privacy. Experiments reveal that our method can successfully catch objects in surveillance videos, even when the objects stop for a long time under varying illumination conditions, while concealing the private information.

Key words: Object detection, Illumination variation, Background estimation, Privacy protection, Surveillance cameras

1. Introduction

Video surveillance is increasingly common in public spaces for monitoring activities and preventing crimes. However, there are too few authorized people (e.g. guards) to be able to peruse all video streams continuously. Therefore, the streams will have to be opened up to other people. This raises the serious problem of a loss of privacy. We are currently researching a form of privacy protection technology suitable for static surveillance cameras.

We explain our proposed approach with a scenario in a residential area, see Fig. 1. We assume that several static surveillance cameras are capturing images of objects like people and cars in a local neighborhood. The video data contains private information such as faces and license plates. The proposed method first extracts object regions using the background image; this image is stable regardless of illumination changes. Next, it filters the extracted regions to protect privacy. As a consequence, scenes captured on video can be understood even though

Fig. 1 Privacy protection with surveillance cameras in a residential area
the objects cannot be identified from the filtered data. This enables monitoring by residents, which will lead to safer residential areas.

The illumination conditions are always changing, especially outdoors, and existing object extraction techniques do not handle these changes reliably. Our approach allows the background image to be estimated accurately from the input images. Our method can robustly extract the objects present even if they are stationary for long periods. This is the key advance of our research.

The next section introduces related works and our motivation. Section 3 details the key components of our method. Experiments that demonstrate the effectiveness of the proposed method are described in Section 4.

2. Related Works

The widespread use of many sensors yields serious issues with regard to privacy, and research on privacy protection is becoming popular\(^1\),\(^2\). Research has been done on preserving privacy in the event that video surveillance images are widely/publicly available\(^3\)\(^5\). The main point of these studies is to extract the object regions\(^6\)\(^8\). However, when the illumination conditions change and/or the objects are stationary for long periods, object extraction becomes unreliable. Moreover, privacy protection is not assured.

Frame differencing is the most direct method of extracting moving objects. However, this method does not detect stationary or slow-moving objects, and it also fails when the objects are not sufficiently textured relative to the background. Frame differencing succeeds in detecting motion when the change between two successive frames is strong enough.

Background differencing is one of the most common methods of extracting objects. In this method, objects are detected by subtracting the pixel value of the input image from the pixel value of the background image and applying a threshold. However, the entire image, including the background, may change with the shadows cast by buildings and clouds and lighting conditions can change widely. This method depends heavily on the threshold value.

Averaged-background differencing is one of the most effective methods of extracting moving objects in real world scenes. This method focuses on the changes at each pixel in successive frames, and creates the background image by averaging. This method is robust against changes in the background image due to shadow or lighting changes. However, it is necessary to update the whole background image, and the method fails if an object is stationary for a long period.

To overcome the above problems, we propose a new object detection method that more accurately estimates the background image.

3. Proposed Method

The proposed method consists of two phases. As illustrated in Fig. 2, the first phase is the object detection phase, and the second one is the filtering phase.

3.1 Object Detection Phase

In this phase, the system estimates the background image. The object regions are extracted by subtracting the pixel value of the background image estimated in the previous frame from the pixel value of the current input image. The pixel value is the gray-scale intensity of the pixel and for a color image the intensity is calculated using Eq.1. Here, \( I \) is gray-scale intensity and \( R/G/B \) is the \( R/G/B \) value of a color image. We assume that the first input image is captured without objects and is stored as the first estimated background image.

\[
I = 0.212671 \times R + 0.715160 \times G + 0.072169 \times B
\]  

(1)

The system always linearly updates the non-object regions of the estimated background image using those of the latest input image in Eq. 2 because it
is thought that the illumination is always changing. Here, \( \alpha \) is the training coefficient, \( 0 \leq \alpha \leq 1 \), \( I_{\text{cur}} \) is the pixel value of the current estimated background image, \( I_{\text{cur}} \) is the pixel value of the current input image, and \( I_{\text{prev}} \) is the pixel value of the estimated background image in the previous frame.

\[
I_{\text{cur}} = \alpha I_{\text{cur}} + (1 - \alpha) I_{\text{prev}}
\]

Note that the object regions of the estimated background image cannot be updated because we cannot directly obtain the pixel value of the object regions of the current background image. Therefore, we estimate the object regions of the current background image using the current input image and the base background image as illustrated in Fig. 3. The system uses the regions surrounding the object as a measure of estimation because the object’s surrounding regions are adjacent to the object regions, and the object regions of the estimated background image and the object’s surrounding regions in the current input image are strongly correlated. The surrounding regions are obtained by subtracting the pixel value of the object regions from the pixel value of the regions created by slightly expanding the object regions.

If the difference between the surrounding regions of the base background image and those of the input image are under the threshold, the system moves to the next frame. We estimate the object regions of the current background image from the object regions of the base background image using the transformation coefficients in Eq. 4. Here, \( I_{\text{cur}} \) is the pixel value of the current estimated background image and \( I_{\text{base}} \) is the pixel value of the base background image.

\[
I_{\text{cur}} = a I_{\text{base}} + b
\]

3.2 Filtering Phase

In this phase, the system filters the current input image using the object regions extracted in the first phase to protect privacy. We first should decide on the kind of filter employed. Several image filters (e.g. median, gaussian, blur, bilateral, or fill with a specific color) can be used to protect private information. Filter selection is based on, in large part, the size of the objects relative to the background; objects that occupy a large (small) percentage of the field of view need a strong (light) filter. Here, the size of the objects is the number of pixels in the object regions. Two filter parameters (filter level and filter size) are set at the same time. The system filters the current input image using the object regions extracted in the object detection phase. We also store the current input image as the original image because none of the filters should be reversible. This makes it possible to examine the scene later if there was a crime.

4. Experiments

4.1 Simple Lighting Condition

In this experiment, images were captured with a Sony digital video camera at 30 frames/sec. We chose a simple lighting indoor scene for testing. When the trial started, the scene with no object was captured for 1 frame. We put a block in the scene at 208 frames.
and dimmed the light from 209 frames and removed the block at 813 frames.

We compared the proposed method with the Background Differencing and Averaged-background Differencing (90 frames were used for averaging). In each method, the threshold value of the background subtraction was set at 20. In the proposed method, we set the training coefficient $\alpha$ to 0.25 (updating the background by about 100 millisecond) considering the time range of illumination changes in this experiment.

Fig. 4 shows the input image, and the background as estimated by the proposed method, and the object detection results (the regions occupied by the block are shown in white.) obtained by the proposed method, Background Differencing and Averaged-background Differencing. As can be seen, the estimated background image was sufficient to accurately extract the object regions and the proposed method could detect the object most robustly under illumination variations in the simple lighting condition.

4.2 Outdoor Scene

In this experiment, images were captured with a Panasonic network camera at one frame/sec. We chose a parking lot as the site for testing and the weather was cloudy. First, we recorded the lot with no cars in it at 6:00 am. People started to arrive and four parked cars were present from 6:00 am to 12:00 am. Here, we set the training coefficient $\alpha$ to 0.01 (updating the background by about 1 minute) considering the expected time range of illumination changes in outdoor scene.

Fig. 5 shows the input image, the background image as estimated by the proposed method, the object detection image obtained by the proposed method, and the filtered image. As can be seen, the estimated background image was sufficient to extract the object regions and the proposed method could detect the objects robustly in an outdoor scene. As a result, our method could successfully conceal the private information of the objects by filtering, even when the objects stopped for a long time under varying illumination conditions in an outdoor scene. Note that the window regions of the cars were not extracted because the difference between the reflectance property of the window (a part of the object regions) and of the asphalt (the surrounding regions) was too great. This is due to the linear transformation adopted by the proposed method.

4.3 Implementation of Our System

This section introduces our privacy protection scheme. In the basic scheme, the raw images are continually received and stored. At the same time, the objects in the raw images are extracted and then filtered by the processing device. The filtered images are output for public viewing. In some applications, the locations at which the filtered images are displayed may be restricted (for example an apartment building). In such cases, it is possible to weaken the filter to allow the public viewer to be able to discriminate between their neighbors and strangers while still protecting the privacy of the strangers.

As shown in Fig. 6, our system is implemented using two PCs (one for image processing and the other for viewing), a storage device, and a fixed surveillance camera; all were connected by Ethernet. The camera was fixed on a pole and captured 640×480 (pixels) video frames at about three frames/sec. The two PCs and the storage device are described in...
Table 1 System specifications

<table>
<thead>
<tr>
<th></th>
<th>CPU</th>
<th>Memory</th>
<th>HDD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Storage</td>
<td>ARM 600 Mhz</td>
<td>256 MByte</td>
<td>400 GByte (RAID5)</td>
</tr>
<tr>
<td>Processing PC</td>
<td>Pentium W 3.6 Ghz</td>
<td>2 GByte</td>
<td>-</td>
</tr>
<tr>
<td>Viewing PC</td>
<td>Pentium W 2 Ghz</td>
<td>1 GByte</td>
<td>-</td>
</tr>
</tbody>
</table>

Table 1. All operations were processed in real-time. As can be seen, the private information of the standing person was well concealed by our system.

Test confirmed that a high end PC running our system could handle the output of eight cameras at the same time.

5. Conclusion

We proposed an object detection method that protects privacy to allow the public release of surveillance camera data. Our proposed method consists of two phases: object detection and filtering. We demonstrated that objects are robustly extracted since the system accurately estimates the background image regardless of illumination changes and object speed; the private information captured on video is concealed by filtering.

We implemented the proposed method in a system consisting of two PCs, a storage device, and a fixed surveillance camera. In tests, all operations were processed in real-time for 640×480 (pixels) video frames at about three frames/sec and with a high-end PC, our system could handle the output of eight cameras at the same time.

Future work includes considering errors caused by the linear transformation. For example, sometimes, object regions that reflected sunlight are not extracted and dark shadow regions are extracted (in sunny days), regions of reflection from the surface of bodies of water are extracted (in rainy days). To solve these errors, we will consider other transformation methods.

References


(Received December 28, 2007 ; Revised March 1, 2008)

Hidenori Tanaka He received the B.E. and M.E. degrees in science and technology from Keio University, Yokohama, in 2004 and 2006, respectively. Since joining NTT Cyber Space Laboratories in 2006, he has been engaged in research on computer vision and pattern recognition. He was awarded the Young Researchers’ Award from IIECE in 2006 and the IEVC Best Paper Award from IIEEJ in 2007. He is a member of IEICE.
Paper: Object Detection under Illumination Variation with Privacy Protection

Hiroyuki Arai  He received the M.S. degree in physics from Hokkaido University, Hokkaido, in 1991. He joined NTT in 1991 and engaged in research on a map recognition system. He was transferred to NTT DATA in 2001 and developed image processing techniques. He was transferred to NTT Cyber Space Labs. in 2005. From 2000 to 2005, he was a fellowship-researcher in the “Natural Vision Project” of the National Institute of Information and Communications Technology. He is a member of IEICE and ITE.

Hitoshi Nakazawa  He received the B.S. degree in electronic engineering from Ibaraki University, Ibaraki, in 1984. Since joining Nippon Telegraph and Telephone Public Corporation (now NTT) in 1984, he has mainly been engaged in R&D of facsimile intelligent communication systems (F-net), digital rights management systems, desktop conference systems, and intelligent monitoring systems (NiMSA). He is a member of IEICE.

Takayuki Yasuno  (Member)  He received the B.E. and M.S. degrees from Keio University, Yokohama, Japan, in 1986, 1988, respectively. In 1988, he joined NTT. He was a visiting researcher in Waseda University, Tokyo, from 1994 to 1997. He is now a senior manager at NTT Intellectual Property Center. He holds a Dr. Eng. degree. He is interested in computer vision and visual communication systems. He is a member of IPSJ and IIEEJ.

Hideki Koike  He received the M.S. degree in mathematics from Tohoku University, Miyagi, in 1985. He joined NTT Labs. in 1985 and engaged in research on image processing. He was transferred to NTT COMWARE in 2001, and engaged in research on RFID. He moved to NTT Cyber Space Labs. in 2007 and is engaged in research on computer vision. He is a member of IEICE.