

行政院國家科學委員會專題研究計畫 成果報告

隱私相機：自動地把路人從無預警的拍攝裡消除的相機 研究成果報告(精簡版)

計畫類別：個別型

計畫編號：NSC 95-2221-E-002-150-

執行期間：95年08月01日至96年07月31日

執行單位：國立臺灣大學資訊工程學系暨研究所

計畫主持人：朱浩華

計畫參與人員：博士班研究生-兼任助理：游創文

碩士班研究生-兼任助理：吳昊極、林東昀

處理方式：本計畫可公開查詢

中華民國 96 年 09 月 10 日

行政院國家科學委員會補助專題研究計畫 ☒ 成果報告
☐ 期中進度報告

(計畫名稱)

計畫類別：☒ 個別型計畫 ☐ 整合型計畫

計畫編號：NSC 95 - 2221 - E - 002 - 150 -

執行期間：95 年 8 月 1 日至 96 年 7 月 31 日

計畫主持人：朱浩華

共同主持人：

計畫參與人員：吳昊極 林東昀 游創文

成果報告類型(依經費核定清單規定繳交)：☒ 精簡報告 ☐ 完整報告

本成果報告包括以下應繳交之附件：

☐ 赴國外出差或研習心得報告一份

☐ 赴大陸地區出差或研習心得報告一份

☐ 出席國際學術會議心得報告及發表之論文各一份

☐ 國際合作研究計畫國外研究報告書一份

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、
列管計畫及下列情形者外，得立即公開查詢

☐ 涉及專利或其他智慧財產權，☐ 一年 ☐ 二年後可公開查詢

執行單位：臺灣大學資訊工程

中 華 民 國 96 年 9 月 10 日

行政院國家科學委員會專題研究計畫成果報告

隱私相機：自動地把路人從無預警的拍攝裡消除的相機

Privacy-aware cameras: automating the removal of accidental passers from unintentional cameras

計畫編號：NSC 95 — 2221 — E — 002 — 150 —

執行期限：95 年 8 月 1 日至 96 年 7 月 31 日

主持人：朱浩華 國立臺灣大學資訊工程學系

計畫參與人員：吳昊極 林東昀 游創文 臺灣大學資訊工程學系 網路與多媒體研究所

Abstract

Digital cameras have become one of the most widely used gadgets in our everyday lives to capture and share moments of day-to-day experiences, anytime, anywhere. However, this ubiquity of cameras also raises serious social privacy concerns for many people who wish to retain their privacy and secrecy. In this project, we propose the *privacy-aware cameras* to address one aspect of this privacy concern – in the use of digital cameras in public spaces where camera snapshots may unintentionally catch co-located and walk-by passers. By tracking the whereabouts of passers in the proximity of a camera, our cameras can help to filter them out.

We have prototyped this privacy-aware camera and performed live experiments. Experimental results have shown that it works well given a small number of passers.

I. Background

Digital cameras have become one of the most widely used gadgets in our everyday lives to capture and share moments of our day-to-day experiences, anytime, anywhere. Given the small form factor and low hardware cost, digital cameras have become popular additions in most of the mobile phones sold nowadays. Some new mobile phones even have dual cameras – one camera facing the front and the other one facing the back. Based on reports from IDC [6], camera phones have made up more than half of all mobile phones sold in Europe and Japan in 2004. In the U.S., camera phones are expected to reach past the 50% mark at the end of 2005. A parallel but complementary development is in the high storage (1+ gigabyte) of flash memory cards that can be plugged into these camera phones to store almost unlimited number of high quality pictures. Together, they enable a new *freedom* of digital photography to snap pictures at every mindful or sometimes mindless moment. However, this new freedom of digital photography also raises serious social privacy concerns for many people who wish to retain their privacy and secrecy. People are now living in public spaces full of such privacy-intrusive digital cameras and camera users.

In this project, we propose a technical solution to address one aspect of this privacy concern – in the use of

digital cameras in public spaces where camera snapshots may *unintentionally* catch co-located and walk-by passers who strongly object to any parts of them being captured in other people's pictures. Note that our solution is targeting for *unintentional capture*, not intentional capture [1]. Unintentional capture means that the picture takers have no intention of capturing these passers in pictures. As a matter of fact, the picture takers may consider these accidental passers as undesirable background noises or unwanted intrusion adversely affecting the picture quality.

Our proposed solution is called the *privacy-aware cameras*. Cameras are enhanced to become (*location*) aware of the presences of privacy-concerning passers in the surrounding environments. Given this awareness, whenever a picture is taken, the privacy-aware camera can *protect the passers' privacy by automatically removing the parts of the pictures that show the passers*. Although this idea seems sensible for the passers, it raises a practical question for the camera owners – what is the incentive for the picture takers to buy cameras that help protect only the privacy of passers? Worse, the removal of passers from the pictures can leave *odd-looking blank hole(s)*, which can further degrade the quality of the pictures. To address this incentive question, we must find an attractive benefit for the picture takers. Our solution must ensure that the removals of the passers are *clean* and *invisible*, as if the picture is taken without the passers physically being there. With this benefit, the picture takers can enjoy the *freedom* in taking pictures at any time without having to worry about passers, considered as undesirable background noises, showing up in their pictures. This benefit can be illustrated in Figure 1. To provide this clean and invisible passer removal, we incorporate a pictorial composition technique called *photomontage* [7] that combines additional picture(s) taken shortly after (refer to these additional pictures as “background frames”), to seamlessly and realistically repair the missing background left by the removed passers.

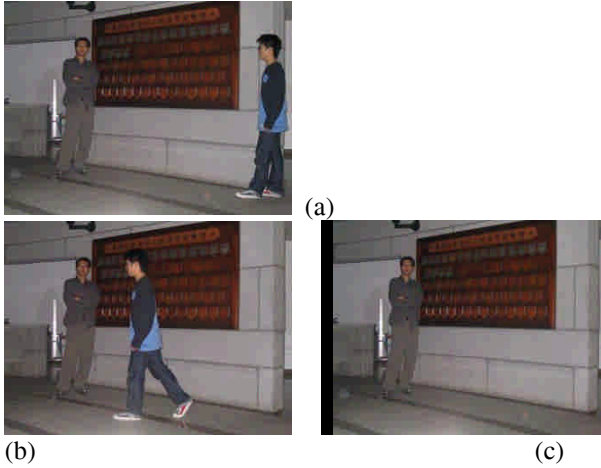


Fig 1. The passer removal by privacy-aware camera. The left image (a) is the original picture captured by the privacy-aware camera. After the composition with a background frame shown as the middle image (b) of Fig 1, the privacy-aware camera can produce a new picture without the passer, shown as the right image (c).

To provide privacy-awareness and seamless passers' removal, we have integrated (1) a vision tracking system, (2) a localization system, and (3) an image processing (photomontage) system into our privacy-aware camera. First, the vision tracking system is used to *track* the presence of people and mark their pixel areas within the camera view. Second, the localization system is invoked to label people within camera view as either passer(s) or target(s). Third, the image processing system applies the photomontage technique to seamlessly repair the passers' pixel areas in the original picture, using realistic backgrounds taken at a later time (called the background frames) where the passers have moved away from these background pixel areas. The result is that (1) the passers are protected from unintentional capture by having the parts of their images removed from the camera pictures, and (2) the picture-takers can capture high quality pictures that are free from these passers' intrusions into their pictures. This is a *win-win* scenario for both the passers and picture-takers who sometimes must peacefully share the same public spaces at the same time.

Our privacy-aware cameras differ significantly from previous work in privacy protection for cameras. In general, the previous work can be categorized into the following three general categories: (1) adding noises to pictures such that identifiable parts of the passers, such their faces, are blurred [4][5], (2) damaging pictures such that passers cannot be recognized, e.g., by shining a light beam on the camera lens at the time of picture shooting [1] [9], and (3) disabling cameras when passers are in proximity [2][3][8]. Although the first and third approaches – damaging pictures and disabling cameras, can help to protect the passers' rights, this protection for the passers is provided at the cost of the picture-takers' capture right in the public spaces. The second approach – blurring the part of the pictures containing the passers, is similar to the first and second approaches in that it degrades the quality of the pictures in order to protect the passers' privacy. In comparison,

our approach takes a different view by considering these passers as unwanted noises in the pictures and then filtering these noises out for the picture-takers. Our noise-removal concept is the opposite of the noise-insertion concept used in the previous approaches. The advantage in the noise-removal approach is that the privacy of passers can be protected without sacrificing the quality of pictures captured by the picture-takers. The reminder of this project report is organized as follows. Section 2 describes the related work. Section 3 specifies our design goals and assumptions. Section 4 presents our design and implementation of the privacy-aware camera. Section 5 shows our experimental setup, experimental results, and limitations of our current prototype. Section 6 draws our conclusion and future work.

II. Related Work

Many technical solutions have been proposed to prevent or react to undesired camera capture. These works can be categorized into three general approaches: (1) adding noise to pictures, (2) damaging pictures such that passers cannot be recognized, and (3) disabling cameras when unwilling passers are in proximity.

Some approaches focus on how to destroy camera by emitting light beam on the lens of camera. For example, solutions, such as Georgia tech's capture-resistant environment [1], leverage the light beam to destroy any possible picture when a CCD retro-reflection is detected by the camera in the environment. A solution called "Eagle Eye" [9] integrates the flashlight unit and a light sensor into a small wearable unit. When a flashlight is detected by the light sensor, the wearable unit automatically emits light beam to destroy the pictures.

Other approaches are to add noises to pictures so that identifiable parts of passers, such their faces, are blurred or masked. For example, Home Media Space [4] uses the blurred based mechanism and rotation of camera to protect the privacy of users in home space. IBM also presents a project called privacy camera [5] that every person will be automatically masked in the view of surveillance camera.

The third approach is to disable camera when unwilling passers are in proximity. A proposal called Safe Haven [3] would transmit signals to all camera phones in a particular area to disable their recording features. The Cloak system [8] addresses privacy concerns by having users carry a "privacy enabling device" (PED), which informs the environment that the cameras should be sanitized after the capture.

III. Design Goal and Assumptions

Although the ultimate designs objective is to create an automated passer removal system that can achieve both *high accuracy* and *precision* while operating in *restriction-free public spaces* with minimum *hardware* cost to passers and picture-takers, this is believed to be a grand challenge requiring extensive future research efforts.

We acknowledge this fact, and consider our privacy-aware cameras as an early effort to address this problem. Since our work is yet a perfect solution, we would like to state our assumptions.

- We assume that an existing localization infrastructure is installed in the public space to locate passers and cameras. There are a variety of choices for indoor and outdoor localizations such as GPS, WiFi, Zigbee, cellular network, ultrasound, RFID, etc. Most of these localization systems require each passer, target, and picture-taker to wear or carry a mobile unit. We argue that this assumption is realistic, considering that this mobile unit can be incorporated into a mobile phone, which is carried by most of people nowadays.
- In addition to carrying this mobile unit, each passer and target is willing to inform nearby cameras about his/her current location relative to the camera under *masked identify*. We argue that this assumption is not impractical, considering that cameras only know that “there is a passer (person) standing at some degree angel to the camera”. The cameras do not need to know the identity of any passer(s).
- Each passer and target sets his/her capture preference, as to which cameras are allowed to capture his/her image. This capture preference is specified by each user on his/her mobile unit. By default, the capture preference will answer to no-capture to all unknown cameras.
- The privacy-aware camera (camera phone) is embedded with sufficient processing and memory capabilities to run low-resolution vision tracking algorithm in real time and image processing algorithm (photomontage) in non-real time.

IV. Design and Implementation

The design of our privacy-aware camera is shown in Fig. 2. Its design consists of the following three phases: *Vision processing phase*, *Localization phase*, and *Image processing phase*. A general overview of these phases is described using the following example: Joe, a picture-taker, is taking a picture while Jane, an accidental passer, walks into the picture. In the first phase, when Joe clicks on the snapshot button on his camera, vision processing system is invoked to track the position of Jane and mark her pixel areas within the camera view at the time of picture-capture. At the same time, Joe’s camera broadcasts a message to Jane (and any all nearby persons), asking for her locations and capture permission. In the second phase, Jane’s mobile unit responds to Joe’s camera about her location with a masked identify, e.g., “a passer at location $\langle x, y \rangle$ does not wish to be captured by your camera”. Joe’s camera can then correctly label Jane’s pixel area within camera view as a passer, by combining Jane’s *location (at the horizontal plane)* relative to the camera and Jane’s *position (at the vertical plane)* within the camera view. In the third phase, the camera automatically takes a background frame when Jane walks outside of her pixel region. Then, the camera runs the image processing

system, photomontage, to seamlessly repair the Jane’s pixel area in the original picture, using realistic back-grounds from the background frames. The result is that (1) Jane, the passer, is happy because she is protected from unintentional capture, and (2) Joe, the picture-takers, is also happy because his picture is free from Jane’s intrusions. More details about the three phases are elaborated as follows.

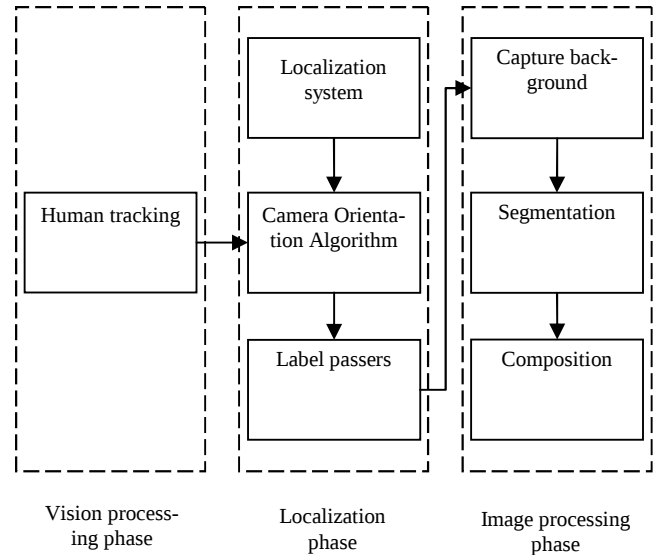


Fig 2. The design of privacy-aware camera

Vision processing phase

The first phase is the vision processing phase. This phase is started when the picture taker is clicking on the camera’s snapshot button. Two tasks are performed during this phase: (1) tracking the positions of passers & target in the picture’s vertical plane, and (2) inferring the relative angles between all passers and the camera in the horizontal plane. The relative angles are needed in the next localization phase to correctly label passers in the picture for removal.

To track the positions of passers (and/or target) in the camera view, we combine two vision algorithms: motion factor and human detection. The motion factor facilitates the tracking of moving passers/targets continuously; whereas the human detection enables the tracking of stationary passers/target in the camera. We use the Intel’s OpenCV software package [12], which contains library for both motion factor tracking and human detection.

Given the positions of passers and target in the camera vertical view, we can calculate the relative angles between passers/targets and the camera on the horizontal plane. The angels can be computed as follows. For each camera, we calibrate a linear mapping function from a passer’s pixel count (to the picture’s middle line on the vertical plane) to his/her camera angle (on the horizontal plane). This mapping function can be illustrated in Fig 3. θ is the angle between the camera and the passer

on the horizontal plane. AB is the distance from the middle line to the position of a passer in the picture. BE is the pixel count distance from the lens to the middle of camera view. Since AB is orthogonal to BE , we can compute α as $\tan^{-1}(AB/BE)$. Because α and θ are the pair of opposite vertical angle, we can infer the relative angle θ between the passer and the camera, which is equaled to α .

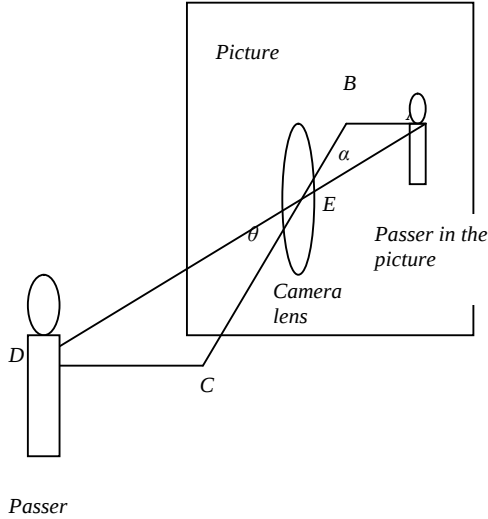


Fig 3. The formation of picture

Localization phase

The localization phase is invoked when the camera broadcasts a request for passers' / target's locations (on the horizontal plane) as well as their capture permission. Upon receiving their anonymous location information, the camera tries to correctly label the positions of passers and targets on the source picture. Since our test bed is an indoor environment (e.g., a museum), we deploy the Cricket ultrasonic location system. The Cricket ultrasonic system is setup as follows. A grid of ultrasonic listeners is installed on the ceiling, and these listeners periodically check the ultrasonic signals transmitted from the beacons carried by passers, targets, and picture-takers. By applying the triangulation method, $\langle x, y \rangle$ coordinates of passers, targets, and picture-takers can be calculated.

In order to correctly label the positions of passers and targets in the picture, it requires determining the global orientation of the camera view. Although the global orientation can be obtained directly from a magnetic-based orientation sensor, we have found that market-available orientation sensors are either too expensive, too bulky, or not accurate enough. As a result, we have come up with a software algorithm to compute the camera orientation.

The idea behind the camera orientation algorithm is shown in Fig. 4. The correct camera orientation can be

found by searching for an orientation angle such that the passers'/targets' relative angles to the camera match their locations on the horizontal plane.

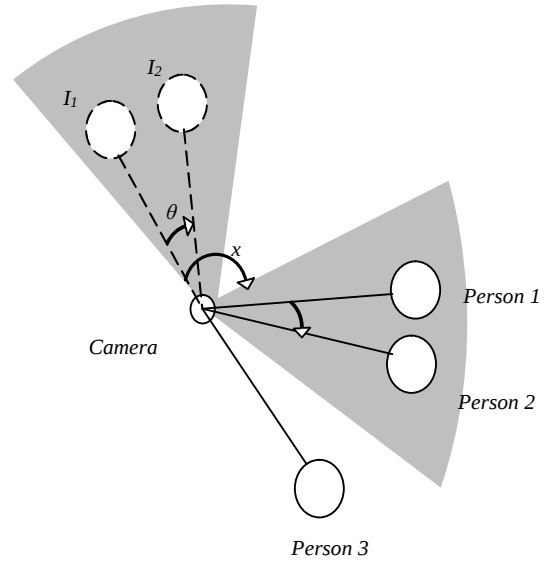


Fig 4. The Camera Orientation Algorithm. This is the map of location system. Person 1, Person 2, and Person 3 are located in the map. The upper fan consists of I_1 and I_2 represents the information gotten from the vision phase. The camera orientation can be found by searched for an orientation angle x such that the passers'/targets' relative angles to the camera match their location on the horizontal plane.

Although the camera orientation algorithm can calculate the angle of the camera orientation, we find several problems that will result in error calculation. Because the inference of orientation is based on the computer vision, the overlapping people will be identified as one person in image. In addition, the error of location system will also result in the calculation error. The other problem is that if there is more than one matching in the map, we can not identify which orientation is correct. In other words, the However, if we add some more information, such as moving or static of people, we can reduce the possibility of matching error.

Passer labeling is to label passer in the image. After getting "not allow to be captured" information from the passer, we can label the passer in the map of location system. Based on the labeled map information and the camera orientation, we can identify the target and passer in the view of camera. Then we can infer the possible passers' pixel area in the picture by the relative angle information which is illustrated in Section 4.2.

Image processing phase

The image processing phase consists of three tasks: background capture, segmentation, and composition. The goal of the background capture task is to get the least possible background frames to fill the backgrounds of passers in the picture. After the location phase, the composite task in the privacy-aware camera automatically captures background frames. As the frames are continuously retrieved, (the pixel region of passer is identified by the vision processing phase), the

privacy-aware camera checks whether the retrieved frames cover all backgrounds left by the passers. Then, the algorithm ends and returns all possible frames for the composition task.

The goal of the segmentation task is to retrieve the part of the passers in the picture. The segmentation task will automatically identify the contour of the passers in the picture and select the similar color pixels from the pixel region of passers. Then the pixel region of the passes is selected from the picture. Finally, the pixel regions of the passers will be removed from the picture. Then, the composite task is performed to fill the pixel regions of the passers by overlapping the background frames captured previously. Finally, to make the overlapping seamless, the edge around the passers' holes is smoothed by using pixel values around the holes.

Implementation

The prototype implementation of the camera is shown in Fig. 5. The Cricket mobile unit [13] for tracking the locations of the passers, targets, and the picture-takers are shown in Fig. 6. For software, we have used Intel OpenCV [12] vision package for human motion tracking, Adaboost [14] for human detection, and Photomontage [7] for clean and seamless passers' removal.



Fig 5 The Camera consist of a Palmtop PC and a Logitech Webcam



Fig 6 The cricket mote

V. Experimental setup, results, and prototype limitations

We have conducted several live experiments to evaluate the privacy-aware camera prototype. We define the following performance metrics for evaluation in the experiments:

- *Capture delay* measures the time duration between the time the source picture is captured and the time all overlay frames are captured;
- *Human detection success rate* measures the percentage of all pictures from which the vision tracking phase correctly recognizes all the humans who come within the camera view;
- *Label success rate* tracks the percentage of all pictures from which the localization phase correctly labels all passer(s) and target(s) who come within the camera view;
- *Removal success rate* tracks the percentage of all pictures from which the image processing phase correctly applies photomontage to remove all passers who come within the camera view.

These performance metrics can be affected by a number of factors, such as the number of passers/target in proximity of the camera, the passers' walking speed, the accuracy of the vision tracking algorithm, and the accuracy of our indoor localization system. In the experiments, we choose to analyze only the first factor and its effect on the evaluation metrics.. This is done by varying the number of nearby passers within & outside the camera view. As for other factors, the passers' walking speed is set to be random human walking speed. The vision tracking algorithm and the localization system are fixed as described in the previous section.

Scenarios	Performance metrics			
	Capture delay (seconds)	Human detection success rate	Label success rate	Passer removal success rate
<i>1T 1P 0O</i>	1.7	96%	92%	89%
<i>1T 1P 1O</i>	2	94%	83%	75%
<i>1T 2P 0O</i>	2.5	81%	79%	75%
<i>1T 1P 2O</i>	2	97%	75%	71%
<i>1T 2P 1O</i>	2.8	79%	71%	63%
<i>1T 3P 0O</i>	3.2	58%	66%	41%

Table 1: Show the results of the live experiments. The experiments evaluate 4 performance metrics: delay, human detection success rate, label success rate, and passer removal success rate. The experiments vary the number of nearby passers within & outside the camera view: T is the number of target, P is the number of passers within the camera view, and O is the number of passers who are outside the camera view.

The experiments consist of six scenarios shown in Table 1. For each scenario, we vary the number of nearby passers within & outside the camera view: T (target) is the number of target person (s) within the camera view and should not be removed; P (inside passers) is the number of passers within the camera

view and should be removed; and O (outside passers) tracks the number of passers outside the camera view. In each scenario, after 50 pictures are taken, we observe and measure the four performance metrics shown in Table 1.

The experimental results show that under small number of passers (e.g., one passer), the capture delay is short (1.7 seconds) and passers' removal success rate is high (89%). However, as the number of passers increases, both the capture delay increases and passers' removal success rate decreases. The reason can be explained as follows. As more passers enter the camera view, the capture delay becomes the time when the slowest passer moves outside of its background area. The cause for the drops in the three success rates can be attributed to several factors. First of all, the human detection fails more frequently given a higher probability of multiple passers overlapping within the same picture area. Since the human detection (the 1st phase) provides inputs to the passer label (the 2nd phase), which further provides inputs to the passer removal (the 3rd phase), there is a chain effect of error.

We have also identified two limitations in our current prototype implementation. The first limitation is that we cannot remove passers who are stationary. This is due to the photomontage that needs to acquire the true background pixel areas occupied by the passers in the picture, and these background pixel areas can be obtained only if the passers move outside of their current background pixel regions. There are several possible solutions to address this limitation. The first possible solution is to alarm the passers, e.g., through a flashing light on their badge, asking them to move a bit in order to avoid unintentional capture. The second possible solution is to apply *inpainting* techniques [10][11] that can infer what the passers' backgrounds would look like from their nearby pixels.

The second limitation is shaking cameras. Excessive amount of camera shaking will also causes seamful or visible overlay in photomontage areas of pictures, because the passers' background areas may have changed between the original pictures and the later background pictures.

VI. Conclusion and future work

This project report describes the design, implementation and evaluation of our privacy-aware camera. The privacy-aware camera provides a technical solution to support (1) privacy protection for the passers from unintentional capture and (2) better picture quality for picture-takers free from intrusions of co-located passers. The unique aspect of this approach is that can accommodate both the rights of passers and picture takers who are sharing the same public spaces. The associated cost is additional hardware & software carried by the passers and picture-takers, and installed on the cameras. Through live experiments, we have shown that under small number of passers (i.e., one passer), our privacy-aware cameras can correctly remove passers at a high success rate (89%).

References

1. Khai N. Truong, Shwetak N. Patel, Jay W. Summet, and Gregory D. Abowd. Preventing Camera Recording by Designing a Capture-Resistant Environment. In the Seventh International Conference on Ubiquitous Computing (ACM Ubicomp 2005). Tokyo, Japan, September 11, 2005.
2. Halderman, J.A, Waters, B., and Felten E.W. Privacy Management for Portable Recording Devices. In the 3rd Workshop on Privacy in Electric Society (WPES 2004). Washinton, DC. October 2004.
3. Iceberg's Safe Haven.
<http://www.iceberg-ip.com/index.htm>
4. Carman Neustaedter and Saul Greenberg. The Design of a Context-Aware Home Media Space for Balancing Privacy and Awareness. In the fifth International Conference on Ubiquitous Computing (ACM Ubicomp 2003). Seattle, Washington, October 12, 2003
5. IBM's Video Privacy
<http://www.research.ibm.com/peoplevision/videoprivacy.html>
6. IDC's Worldwide Camera Phone and Videophone 2005-2009 Forecast
<http://www.idc.com/getdoc.jsp?containerId=34338>
7. Aseem Agarwala, Mira Dontcheva, Maneesh Agrawala, Steven Drucker, Alex Colburn, Brian Curless, David Salesin, Michael Cohen. Interactive Digital Photomontage. ACM Transactions on Graphics (Proceedings of SIGGRAPH 2004), 2004.
8. Brassil, J. Using Mobile Communications to Assert Privacy from Video Surveillance. To appear at 1st International Workshop on Security in Systems and Networks 2005. April 2005.
9. Eagle Eye. Bulletin of the Connecticut Academy of Science and Engineering. Vol. 12, No. 2, 1997.
10. M. Bertalmio, G. Sapiro, V. Caselles and C. Ball-ester. "Image Inpainting". Proceedings of SIGGRAPH 2000, New Orleans, USA, July 2000.
11. Antonio Criminisi, Patrick Perez, Kentaro Toyama. Object Removal by Exemplar-based Inpainting. In the Proc. IEEE Computer Vision and Pattern Recognition, Madison, WI, Jun. 2003.
12. Intel OpenCV Library.
<http://www.intel.com/technology/computing/opencv/index.htm>
13. Nissanka B. Priyantha, Anit Chakraborty, and Hari Balakrishnan
Proc. of the Sixth Annual ACM International Conference on Mobile Computing and Networking (MOBICOM), August 2000.
14. Paul Viola, Michael Jones, Robust Real-time Object Detection. Paul Viola, Michael Jones, Robust Real-time Object Detection. In International Journal of Computer Vision 2001.