Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2007, Article ID 75427, 9 pages
doi:10.1155/2007/75427

Research Article

Tools for Protecting the Privacy of Specific Individuals in Video

Datong Chen, Yi Chang, Rong Yan, and Jie Yang

School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA

Received 25 July 2006; Revised 28 September 2006; Accepted 31 October 2006

Recommended by Ying Wu

This paper presents a system for protecting the privacy of specific individuals in video recordings. We address the following two problems: automatic people identification with limited labeled data, and human body obscuring with preserved structure and motion information. In order to address the first problem, we propose a new discriminative learning algorithm to improve people identification accuracy using limited training data labeled from the original video and imperfect pairwise constraints labeled from face obscured video data. We employ a robust face detection and tracking algorithm to obscure human faces in the video. Our experiments in a nursing home environment show that the system can obtain a high accuracy of people identification using limited labeled data and noisy pairwise constraints. The study result indicates that human subjects can perform reasonably well in labeling pairwise constraints with the face masked data. For the second problem, we propose a novel method of body obscuring, which removes the appearance information of the people while preserving rich structure and motion information. The proposed approach provides a way to minimize the risk of exposing the identities of the protected people while maximizing the use of the captured data for activity/behavior analysis.

1. INTRODUCTION

In the last few years, significantly more video cameras continue to be deployed in a variety of locations for different purposes, such as video surveillance and human activity/behavior analysis for medical applications. These systems have posed significant questions about privacy concerns. There are many challenges for privacy protection in video. First, we have to deal with a huge amount of the video data. A video stream captured by a surveillance camera within 24 hours consists of 2 592 000 frames of image (in 30 fps) per day and more than 79 million image frames per month. Medical studies usually need to conduct a long-term recording (e.g., a month or a few months) with dozens of cameras, and thus produce a huge amount of video data. Second, labeling data is a very labor-intensive task but many automatic video analysis algorithms and systems rely on a large amount of training data to achieve a reasonable performance. This problem becomes even worse when the privacy protection issue is taken into account, because we have only limited personnel who can access the original data. Third, we have to deal with the real-time issue because many video analysis tasks require video data to be processed in real time.

In the previous research, quite a few researchers took account of privacy protection in video from different points of view. Senior et al. [1] presented a model to define video privacy, and implemented some elementary tools to rerender the video in a privacy-preserving manner. Tansuriyavong and Hanaki [2] proposed a system that automatically identifies a person by face recognition, and displays the silhouette image of the person with a name list in order to balance the privacy protecting and information conveying. Brassil [3] implemented a system to allow individuals to protect privacy from video surveillance with the usage of mobile communications. Zhang et al. [4] proposed a detailed framework to store privacy information in surveillance video as a watermark and monitor the invalid person in a restricted area but protect the privacy of the valid persons. In addition, several research groups [5–7] discussed the privacy issue in the computer-supported cooperative work domain. Furthermore, Newton et al. [8] proposed an effective algorithm to preserve privacy by deidentifying facial images. Boyle et al. [9] discussed the effects of blurring and pixelizing on awareness and privacy.

In this paper, we present our efforts in developing tools for protecting the privacy of specific individuals in video. Our problem is slightly different from the previous problems, where a common practice of privacy protection in video is to

obscure human faces as those appearing in TV news. But in this work, since we are interested in privacy protection for medical applications, obscuring faces might not be sufficient for some cases. For example, video/audio analysis can be a very useful assistive tool for geriatric cares. However, some of the patients living in the facility, who do not want to participate in the studies, are also captured by video cameras. In order to protect privacy of those individuals, simply obscuring the face is not satisfactory. Those individuals are required to be removed from the video right after the recording by the regulation. A solution is to completely remove those individuals from video by masking their whole bodies. But this solution makes some studies, such as the social interaction between those individuals and other patients, impossible. Therefore, our goal is to maximize the benefits of the captured video data while effectively protecting the privacy of different individuals. In this paper, we propose to protect privacy by removing appearance information while keeping the structural information of human bodies. We use a pseudogeometric model, that is, edge motion history image (EMHI), to preserve body structure and motion information for activity analysis. In order to obscure those people from video recordings, we have to identify those people from the video first. But as one of the constraints, the university's IRB (Institutional Review Board) has required to protect the identities of patients before unauthorized personnel can access the data. This means that only authorized personnel (e.g., doctors and nurses) can help to identify those people. Manually identifying those individuals in such prolonged video is a very difficult task, if not impossible, because of not only the large data volume but also the high frequency of people appearing and disappearing in the camera scene. Therefore, automatic people identification is crucial for protecting the privacy in video. However, constructing an automatic person identification system also encounters the difficulty of privacy protection issue. On one hand, training a good person identification system requires a large amount of training data. On the other hand, it is difficult for authorized personnel to provide such a large amount of labels. Therefore we augment the learning process with insufficient labeled data and additional pairwise constraints that can be labeled by unauthorized personnel without exposing the patient identity information.

The rest of the paper is organized as follows. Section 2 describes the problem and overviews the developed tools. Sections 3–6 present the development of the people identification tools using noisy pairwise constraints. Section 7 introduces the method for obscuring people and Section 8 concludes the paper.

2. PROBLEM DESCRIPTION

In this research, we would like to develop tools for protecting the privacy of specific individuals in video. Specifically we need to completely remove those individuals' appearance information from video, under the constraint that only authorized personnel can access original data. Therefore, our problem is made up of two subproblems: (1) identify people with the limited labeled data, and (2) remove appearance information but keep the structure information of their bodies.

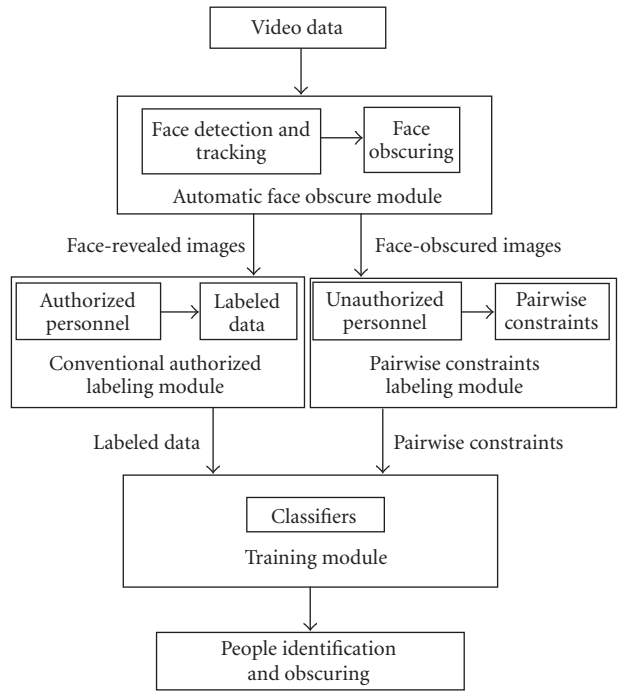


Figure 1: The proposed approach consists of four modules: automatic face obscuring module, conventional authorized labeling module, pairwise constraint labeling module, and training module, and appearance obscures module.

To address the first subproblem, we use a system that identifies people based on color appearances, because current recognition algorithms are not robust enough to produce useful results given data of this quality. We propose a method that can augment labeled data by training a person identification system from both identity labeled data and pairwise constraints. The basic idea is to let authorized personnel label identities of people on a small set of data and ask unauthorized personnel to label pairwise constraints from the video data with human faces automatically masked. We then use the true labels as well as pairwise constraints to train the people classifier.

The proposed approach consists of five modules as shown in Figure 1. The first module automatically locates human faces and computes their obscure masks. An algorithm is proposed to robustly detect human faces by integrating face detection and bidirectional tracking, which is discussed in Section 3. The training data for constructing a person identification system can be labeled from two different modules. One is the conventional labeling module, in which the authorized personnel can label identities of human subjects from the original video data. The labeling results are the subjects' images associated with the identities. The other is the pairwise constraint labeling module, which is used to label pairwise constraints from face-obscured video data. When labeling a pairwise constraint, a user is asked to

judge if two images belong to the same class without identifying who they are. The judgment on a selected image pair is called a pairwise constraint. In Section 4, we describe a user study, which verifies that humans can perform reasonably well in labeling pairwise constraints from face-masked images. Compared to the conventional labeling process, it is much cheaper to obtain a large number of pairwise constraints by exploiting unauthorized human power without exposing identities of human subjects in the video.

The fourth module trains the classifier for identifying people using both labeled data and pairwise constraints. Note that the previous work on using pairwise constraints assumes the existence of noiseless pairwise constraints. However, we have to deal with noisy pairwise constraints in the proposed method because it is difficult for the unauthorized annotators to label perfect pairwise constraints from face-obscured data. Therefore, we propose a novel discriminative learning algorithm based on conventional margin-based learning algorithms to handle imperfect pairwise constraints in the training process. The final module obscures appearances of selected individuals to protect patients' privacy from public access. The appearance of a protected subject is removed from both face and body texture while the structures of the body and motion are preserved.

3. THE AUTOMATIC FACE OBSCURING MODULE

This module first detects and tracks faces in video frames, and then creates obscure masks using the face locations and scales. In this section, we only focus on describing the face detection and tracking process, which must achieve a high recall in order to protect patients' privacy. Large variances on face poses, sizes, and lighting conditions are major challenges in analyzing surveillance video data, which cannot be covered by either profile faces or even intermediate estimations. In order to achieve a high recall, we utilize a new forward-backward face localization algorithm by combining face detection and face tracking technologies.

Many visual features have been used to detect faces, for example, color [10] and shape [11], which are effective and efficient in some well-controlled environments. Most recent face detection algorithms employ texture feature or appearances and train face detectors statistically using learning techniques, such as Gaussian mixture models [12], PCA (principal components analysis), neural networks [13], and SVM (support vector machine) [14]. Viola and Jones [15] applied the boosting technique to combine multiple weak classifiers to achieve fast and robust frontal face detection. To detect faces in varying poses, profile faces [16] and intermediate pose appearance estimations [17] have been studied but the problem is still a great challenge.

Face tracking follows a human head or facial features through a video image sequence using temporal correspondences between frames. In this paper, we are only interested in tracking human heads, which can be achieved by tracking segmented regions [18], color models or color histograms [19–21], or shapes [11]. A tracking process includes predicting and verifying the face location and size in an image frame given the information in the consecutive frames. Kalman filters [22] and particle filters can be used to perform the prediction adaptively.

To effectively obscure human faces in video, we propose a bidirectional tracking algorithm to combine face detection, tracking, and background subtraction into a unified framework. In this algorithm, we first perform background subtraction to extract foreground and then run face detection on the foreground. Once a face is detected, we track the face simultaneously in both backward and forward directions in video.

3.1. Background subtraction

A background is dynamically learned by using the kernel density estimation [23]. Given a set of appearances $A = (A_{t_1}, A_{t_2}, \ldots, A_{t_n})$ of a layer extracted with rectangular windows from n frames, we can normalize the size of each appearance and represent it as $\overline{A} = (\overline{A}_{t_1}, \overline{A}_{t_2}, \ldots, \overline{A}_{t_n})$. Let $A_t(x)$ be a pixel value at a location x in the rectangle appearance patch of A_t. Given the observed pixel value $A_t(x)$ in a tracking candidate window A_t (can also be normalized to $\overline{A}_t$), we can estimate the probability of this observation as

$$\Pr\left(\overline{A_t}(x)\right) = \frac{1}{n}\sum_{i=1}^{n} \alpha K\left(\overline{A_t}(x), \overline{A_{t_i}}(x)\right), \tag{1}$$

where K is a kernel function defined as a Gaussian function:

$$K(x_1, x_2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\|x_1 - x_2\|^2/2\sigma^2}. \tag{2}$$

The constant σ is the bandwidth. Using the color values of a pixel, the probability can be estimated as

$$\Pr\left(\overline{A_t}(x)\right) = \frac{1}{n}\sum_{i=1}^{n} \alpha \prod_{j\in(R,G,B)} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(\overline{A_t}(x)^j - \overline{A_{t_i}}(x)^j)^2/2\sigma^2}, \tag{3}$$

where α is the weight associated with the number of appearance samples in the model A:

$$\alpha_i = \frac{1}{|A|}. \tag{4}$$

Given a background model and a new image, foreground regions can be extracted by computing the probability of each pixel in the image using (3) with a cutoff threshold of 0.5.

3.2. Face detection

Two face detectors are used in parallel on the extracted foregrounds in this paper. The first face detector is the Schneiderman-Kanade [16] face detector. The detector extracts wavelet features in multiple subbands from a large amount of labeled images and trains neural networks using a boosting technique. The detector is used to detect only frontal faces in this paper, though it can be extended for several other poses, which are pretrained as face profiles.

The second face detector is a head-and-shoulder analyzer based on the boundary of a foreground region. The shape of a combination of head and shoulder is a good evidence to detect the face (head) of a standing or sitting person with a large variation of head poses.

SVMs are trained to detect head-and-shoulder patterns on the basis of bag-of-segments. To extract this feature, long up-boundaries are first tracked in the background-subtracted image. We then scan a boundary contour with a 5-overlapped-circle template. The related positions of the 5 circles are fixed. We vary the sizes of the template from 25 pixels to 125 pixels $(25, 45, \ldots, 125)$ in height. The template extracts 5 segments at each location as shown in Figure 2. We represent each segment using the second, third, and fourth orders of moments after normalizing with the first order of moment.

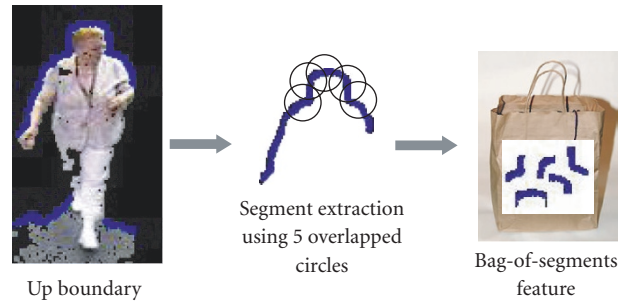


Figure 2: Feature extraction of head-and-shoulder detection.

3.3. *Face tracking*

A detected face is tracked in both backward and forward directions. We track a face using an approach based on online region confidence learning. This approach associates different local regions of a face with different confidences on the basis of their discriminative powers from their background and probabilities of being occluded. To this end, face appearances are dynamically accumulated using a layered representation. Then a detected (or tracked) face area is partitioned into regular and overlapping regions. We learn the confidences of these regions online by exploiting the most discriminative features to local background, and the occlusion probability in the video. The learned regions confidences are modeled as bias terms in a mean-shift tracking algorithm. This approach has advantages of using region confidences against occlusions and a complex background [11].

The performance of the face detection and tracking algorithm is evaluated by a public CHIL database (chil.server.de). In 8 000 testing frames, the algorithm detected 98% (recall) faces in the ground truth with at least 50% area covered by the detection results with a 95% precision.

4. LABELING PAIRWISE CONSTRAINTS WITHOUT EXPOSING PEOPLE IDENTITIES

To address the leakage of the authorized human power in labeling, we use two labeling modules, including the conventional labeling module for authorized personnel and the pairwise constraint labeling model for unauthorized personnel. In the second labeling module, we can employ a large number of unauthorized personnel to provide data labels for training. The challenge is how to obtain useful data labels from unauthorized personnel while still maintaining the privacy of protected subjects from these unauthorized personnel.

Instead of labeling the identities of the subjects in video data directly, we propose an alternative solution by labeling the pairwise constraints so that the subject identities are not exposed. By definition, a pairwise constraint between two examples indicates whether they belong to the same class or not. For example, we show a number of snapshots of face-obscured images to an annotator and ask him/her to pick out two snapshots that are most likely to be the same person. Such a constraint provides additional weak information in a form of the relationship between the labels rather than the labels themselves. There are two problems to be considered when using pairwise constraints to improve the training of classifiers.

(1) The labeled pairs may or may not correspond to the same subject. The accuracy of this labeling process is crucial for a further training task.

(2) How to improve a classifier with imperfect pairwise constraints?

5. A USER STUDY OF THE PAIRWISE CONSTRAINT LABELING QUALITY

Can we obtain satisfactory pairwise constraints without exposing people's identities? Our intuition is that it is possible for unauthorized personnel to obtain highly accurate constraints without seeing the faces, because they could use clothes, shape, or other cues as the alternative information to make decisions on pairwise constraints. To validate our hypothesis, we performed the following user study.

We only display the human silhouette images with obscured faces in the user interface shown to human subjects. A screen shot of the interface is shown in Figure 3. The image on the top-left side is the sample image, while the other images are all candidates to be compared with the sample image. In the experiments, the volunteers were requested to label whether the candidate images contained the same person as the sample image. All images were randomly selected from preextracted silhouette images and all candidate images do not belong to the same sequence as the sample image. There are two modes in our user study tool. In the complex mode, there are multiple candidate images matching to the sample image, while in the simplified mode, only one candidate image matches the sample image. Current user studies take the simplified mode as the basic test bed on the static images. In more detail, the displayed images were randomly selected from a pool of 102 images, each of which was sampled from a different sequence of videos. These video sequences were captured by a surveillance camera in a nursing home environment.

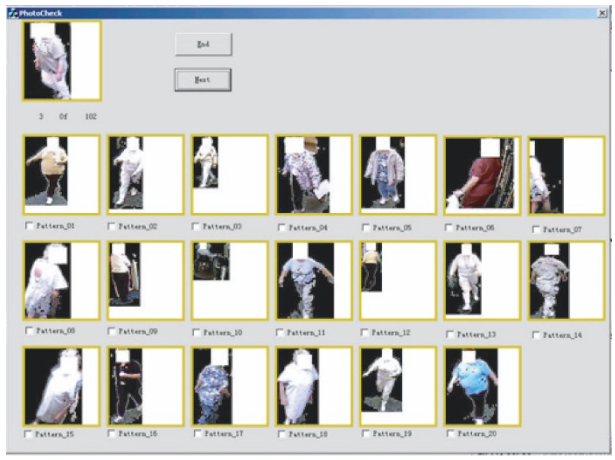


Figure 3: The interface of the labeling tool for user study.

In the user study, nine human subjects took a total of 180 runs to label the pairwise constraints. In all 160 labeled pairwise constraints, 140 constraints correctly correspond to the identities of the subjects and 20 of them are errors, which achieved an overall accuracy around 88.89%. The result shows that human annotators could label the pairwise constraints with a reasonable accuracy from face-obscured video data. But this study also indicates that these pairwise constraints are not perfect. There is a certain amount of errors in the labels, which can pose a challenge for the following training phase.

6. DISCRIMINATIVE LEARNING WITH NOISY PAIRWISE CONSTRAINTS

To improve upon the classifiers solely using these training examples, we attempt to incorporate the imperfect pairwise constraints labeled from unauthorized personnel as complementary information. That is, we use two different sets of labeled data to build the classifier: one set of labeled data provided by authorized personnel from original video; the other set of imperfect pairwise constraints labeled by unauthorized personnel from privacy-protection data with obscured faces.

We propose a novel algorithm to incorporate the additional pairwise constraints obtained from unauthorized personnel into a margin-based discriminative learning. Typically, the margin-based discriminative learning algorithms focus on the analysis of a margin-related loss function coupled with a regularization factor. Formally, the goal of these algorithms is to minimize the following regularized empirical risk:

$$R_f = \sum_{i=1}^{m} L(y_i, f(x_i)) + \lambda\Omega(\|f\|), \tag{5}$$

where x_i is the feature of the ith training example, y_i denotes the corresponding label, and $f(x)$ is the classifier output. L denotes the empirical loss function, and $\Omega(\|f\|)$ can be regarded as a regularization function to control the computational complexity. In order to incorporate the pairwise constraints into this framework, Yan et al. [24] extended above optimization objectives by introducing pairwise constraints as another set of empirical loss functions,

$$\sum_{k=1}^{m} L(y_k, f(x_k)) + \mu \sum_{i,j} L'(c_{ij}, f(x_i), f(x_j)) + \lambda\Omega(\|f\|_H), \tag{6}$$

where $L'(c_{ij}, f(x_i), f(x_j))$ is called pairwise loss function, and c_{ij} is a pairwise constraint between the ith example and the jth example, which is 1 if two examples are in the same class, -1 otherwise. In addition, c_{ij} could be 0 if this constraint is not available.

Intuitively, when $f(x_i)$ and $c_{i,j}f(x_j)$ have different signs, the pairwise loss function should give a high penalty, and vice versa. Meanwhile, the loss functions should be robust to noisy data. Taking all these factors into account, Yan et al. [24] choose the loss function to be a monotonic decreasing function of the difference between the predictions of a pair of pairwise constraints, that is,

$$\begin{aligned} &L'(c_{i,j}, f(x_i), f(x_j)) \\ &\quad = L(f(x_i) - c_{ij}f(x_j)) + L(c_{ij}f(x_j) - f(x_i)). \end{aligned} \tag{7}$$

Equation (7) assumes perfect pairwise constraints. In the paper, we extend it to improve discriminative learning with noisy pairwise constraints. In our extension, we introduce an additional term g_{ij} to model the uncertainty of each constraint achieved from the user study. The modified optimization objective can be written as

$$\begin{aligned} &\frac{1}{m}\sum_{k=1}^{m} L(y_k, f(x_k)) \\ &\qquad + \frac{\mu}{|C|}\sum_{i,j} g_{ij}L'(c_{i,j}, f(x_i), f(x_j)) + \lambda\Omega(\|f\|_H), \end{aligned} \tag{8}$$

where g_{ij} is the corresponding weight for each constraint pair c_{ij} that represents how likely the constraint is correctly labeled from the user study. For example, if n out of m unauthorized personnel consider these two examples belonging to the same class, we could compute g_{ij} to be n/m. In practice, we can only obtain the positive c_{ij} sign values using a manual labeling procedure or a tracking algorithm. Therefore, we can omit the sign matrix c_{ij} in the future discussion.

We normalize the sum of the pairwise constraint loss by the number of total constraints $|C|$ to balance the importance of labeled data and pairwise constraints. In our implementation, we adopt the logistic regression loss function as the empirical loss function due to its simple form and strict convexity, that is, $L(x) = \log(1 + e^{-x})$. Therefore, the empirical loss function could be rewritten as follows:

$$\begin{aligned} &\frac{1}{m}\sum_{k=1}^{m} \log\left(1 + e^{-y_k f(x_k)}\right) + \frac{\mu}{|C|}\sum_{i,j} g_{ij}\log\left(1 + e^{f(x_i)-y_j f(x_j)}\right) \\ &\qquad + \frac{\mu}{|C|}\sum_{i,j} g_{ij}\log\left(1 + e^{y_j f(x_j)-f(x_i)}\right) + \lambda\Omega(\|f\|_H). \end{aligned} \tag{9}$$

6.1. Kernelization

The kernelized representation of the empirical loss function can be derived based on the representer theorem [25]. By projecting the original input space to a high-dimensional feature space, this representation could allow a simple learning algorithm to construct a complex decision boundary. This computationally intensive task is achieved through a positive definite reproducing kernel K and the well-known "kernel trick." We derive the kernelized representation as the following formula:

$$\frac{1}{m}\cdot\vec{1}^{T}\log\left(1+e^{-\alpha K_P}\right)+\frac{\mu}{|C|}g_{ij}\cdot\vec{1}^{T}\log\left(1+e^{\alpha K_P'}\right)+\frac{\mu}{|C|}g_{ij}\cdot\vec{1}^{T}\log\left(1+e^{-\alpha K_P'}\right)+\lambda\alpha K\alpha, \tag{10}$$

where K_p is the regressor matrix and K_p' is the pairwise regressor matrix. Please see [24] for more details of their definitions. To solve the optimization problem, we apply the interior-reflective Newton methods to reach a global optimum. In the rest of this paper, we call this type of learning algorithms a weighted pairwise kernel logistic regression (WPKLR).

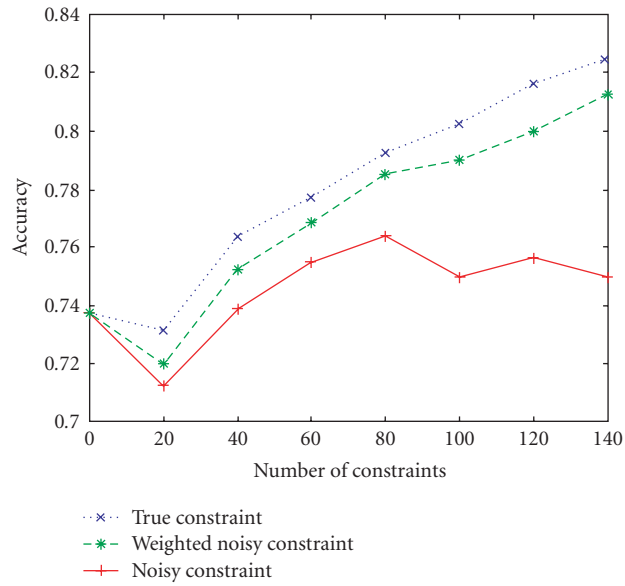


Figure 4: Accuracy with different numbers of constraints.

6.2. Experimental evaluations

In this paper, we applied the WPKLR algorithm to identify people from real surveillance video. We empirically chose the constraint parameter μ to be 20 and the regularization parameter λ to be 0.001. In addition, we used the radial basis function (RBF) as the kernel with ρ to be 0.08. A total of 48 hours video in total was captured in a nursing home environment in 6 consecutive days. We used a background subtraction tracker to automatically extract the moving sequences of human subjects, and we particularly paid attention to video sequences that only contained one person. By sampling the silhouette image in every half second from the tracking sequence, we constructed a dataset including 102 tracking sequences and 778 sampling images from 10 human subjects. We adopt the accuracy of tracking sequences as the performance measure. By default, 22 out of 102 sequences are used as the training data and others as testing, unless stated otherwise.

We extracted the HSV color histogram as image features, which is robust in detecting people identities and could also minimize the effect of blurring face appearance. In the HSV color spaces, each color channel is divided into 32 bins, and each image is represented as a feature vector of 96 dimensions. Note that in this video data, one person could wear different clothes on different days in various lighting environments. This setting makes the learning process more difficult, especially with limited training data provided.

Our first experiment is to examine the effectiveness of pairwise constraints for labeling identities as shown in Figures 4 and 5. The learning curve of noisy constraint is completely based on the labeling result from the user study, but uniformly weighted all constraints as 1. Weighted noisy constraint uses different weights for each constraint. In current experiments, we simulated and smoothed the weights based on the results of our user study. The underlying intuition is that the accuracy of a particular constraint can be approximated by the overall accuracy of all constraints with enough unauthorized personnel for labeling. True constraint assumes that the ground truth is available, and thus the correct constraints are always weighted as 1 while wrong constraints are ignored. Although the ground truth of constraints is unknown in practice, we intentionally depict its performance to serve as an upper bound of using noisy constraints. Figure 4 demonstrated the performance with aforementioned three types of constraints. In contrast to the accuracy of 0.7375 without any constraints, the accuracy of weighted noisy constraint grows to 0.8125 with 140 weighted constraints, achieving a performance improvement of 10.17%. Also, the setting of weighted noisy constraint substantially outperforms the noisy constraint, and it can achieve the performance near to true constraint. Note that when given only 20 constraints, the accuracy is slightly degraded in each setting. A possible reason is that the decision boundary does not change stably with a small number of constraints. But the performance always goes up after a sufficient number of constraints are incorporated.

Our next experiment explores the effect of varying the number of training examples provided by the authorized personnel. In general, we hope to minimize the labeling effort of authorized personnel without severely affecting the overall accuracy. Figure 5 illustrates the performance with a different number of training examples. For all the settings, introducing 140 constraints could always substantially improve classification accuracy. Furthermore, pairwise constraints could make even more noticeable improvement given fewer training examples, which suggests that constraints are helpful to reduce labeling efforts from authorized personnel.

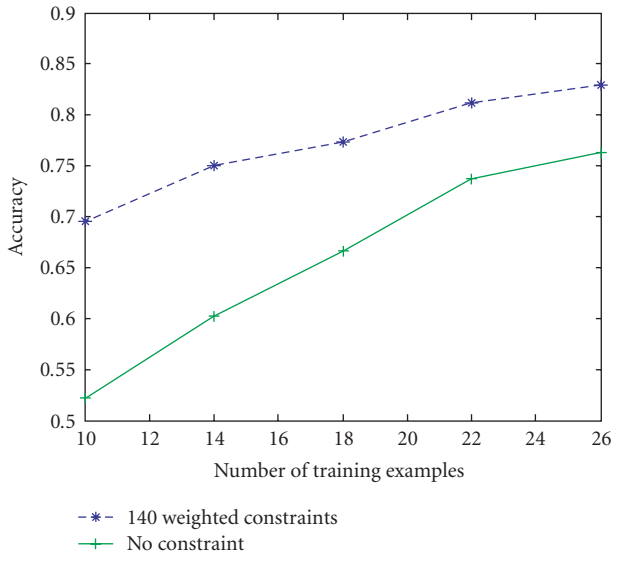


Figure 5: Accuracy with different sizes of training sets.

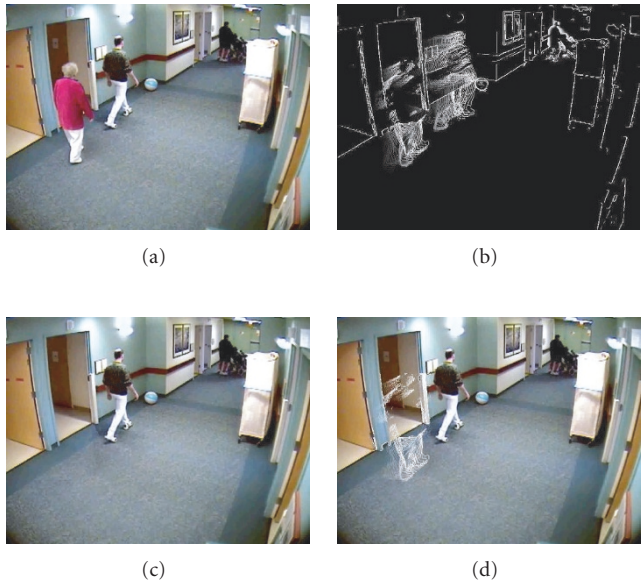


Figure 6: An example of people obscured by using the EMHI. (a) The original image, (b) its EMHI result, (c) the background restoration of the woman in pink identified from the original video frame. The background is learned in the background subtraction introduced in Section 3. (d) The final obscured image.

7. HUMAN BODY OBSCURING

The user study in Section 5 shows that identities of people are not completely obscured by only masking the faces, because other people can recognize their familiars by only looking at body appearances. To obscure protected subjects for the public access purpose while keeping the activity information, Hodgins et al. [26] proposed the geometric models, which include stick figures, polygonal models, and NURBS-based models with muscles, flexible skin, or clothing. The advantage of geometric models is its ability to discriminate motion variations. The drawback is that geometric models, for example, the stick models, are defined on the joints of human bodies, which is difficult to automatically extract from video.

In this paper, we propose a pseudogeometric model, namely edge motion history image (EMHI) to address the problem of body obscuring. EMHI captures the structure of human bodies using edges detected in the body appearances and their motion. Edges can be detected in a video frame, especially around contours of a human body. This detection can be performed automatically, but it is not able to extract edges perfectly and consistently through a video sequence. To integrate noisy edge information in multiple frames and improve the discrimination of the edge-based model, we use the motion history image (MHI [27]) techniques.

Let $E_t(x)$ be a binary value to indicate if pixel x is located on an edge at time t. An EMHI $H_t^\tau(x)$ is computed from the EMHI of the previous frame $H_{t-1}^\tau(x)$ and the edge image $E_t(x)$ as

$$H_t^\tau(x) = \begin{cases} \tau & \text{if } E_t(x) = 1, \\ \max\left(0, H_{t-1}^\tau(x) - 1\right) & \text{otherwise.} \end{cases} \quad (11)$$

In an EMHI, edges are accumulated through the time line to smooth the noisy edge detection results and preserve motion information of the human activities. Figure 6 shows an original video frame, its EMHI result, background restoration, and the final obscured image. The proposed EMHI algorithm completely removes the identity information of the woman in pink from the video while keeping the action information of the woman. Figure 6(a) is the original image. Figure 6 also illustrates possible ways to protect privacy of specific individuals in video. Figure 6(c) shows the result of completely removing the woman in pink from the original image. Figure 6(b) is the result of applying the EMHI to the entire image. Figure 6(d) is the result of applying the EHMI to only the woman in pink.

The EMHI obscuring process is automatic and does not require silhouettes. The obscured image totally preserves the location of the woman in pink. The body texture is obscured and only body contours are partially preserved, which protects the identity of the woman. The activity of the woman is preserved very well. People can easily tell that someone is walking from this ghost-like image.

8. CONCLUSION

In this paper, we have described several useful tools for protecting the privacy of specific individuals in surveillance video. These tools provide a robust algorithm of face localization to obscure all faces in the video. The face masked video can be then used to provide labels of pairwise constraints by collecting identical people snapshots in face-obscured images. The pairwise constraints can be provided by a large group of unauthorized personnel even when they have no prior knowledge of the subjects in the video data. According to our user study, we verified that human subjects could perform reasonably well in labeling pairwise constraints from

face-obscured images. At the same time, the authorized personnel provide a small number of labeled data for learning. We proposed a learning algorithm called WPKLR to train a people identifier with both identity-labeled data and pairwise constraints. Furthermore, we expand the learning methods to deal with imperfect labeling of pairwise constraints. This approach could make use of minimal efforts from authorized personnel in labeling the training data while still minimizing the risk of exposing identities of protected people. Based on people identification results, the tools can further remove the appearances of specific individuals from video while preserving the structure of the body and motion information for activity/behavior analysis. We demonstrate the effectiveness of our automatic people labeling approach through the video captured from a nursing home environment.

Our pairwise constraint labeling experiments show that people's identities can be potentially revealed from the face-obscured images. To avoid revealing the identities of protected subjects, unauthorized people must never see the subjects before. Therefore, the unauthorized people do not have a chance to interpret the subjects' identities even if they have figured out the pairwise constraints between subjects.

Although both the face detection and people classification cannot provide 100% accuracy, the proposed system is still able to reduce most of the labeling effort of the authorized personnel. In the future, more efficient face detection and people classification algorithms will focus on improving the automated modules of the system. We also plan to implement user studies to evaluate performance of the tools in both privacy protection and activity analysis.

ACKNOWLEDGMENTS

This research is partially supported by the Army Research Office under Grant no. DAAD19-02-1-0389, and the NSF under Grants no. IIS-0205219 and no. IIS-0534625.

REFERENCES

[1] A. Senior, S. Pankanti, A. Hampapur, L. Brown, Y.-L. Tian, and A. Ekin, "Blinkering surveillance: enabling video privacy through computer vision," Tech. Rep. RC22886 (W0308-109), IBM, White Plains, NY, USA, 2003.

[2] S. Tansuriyavong and S.-I. Hanaki, "Privacy protection by concealing persons in circumstantial video image," in *Proceedings of the Workshop on Perceptive User Interfaces (PUI '01)*, pp. 1–4, Orlando, Fla, USA, November 2001.

[3] J. Brassil, "Using mobile communications to assert privacy from video surveillance," in *Proceedings of the 19th IEEE International Parallel and Distributed Processing Symposium (IPDPS '05)*, p. 290, Denver, Colo, USA, April 2005.

[4] W. Zhang, S.-C. S. Cheung, and M. Chen, "Hiding privacy information in video surveillance system," in *Proceedings of International Conference on Image Processing (ICIP '05)*, vol. 3, pp. 868–871, Genova, Italy, September 2005.

[5] S. E. Hudson and I. Smith, "Techniques for addressing fundamental privacy and disruption tradeoffs in awareness support systems," in *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW '96)*, pp. 248–257, Boston, Mass, USA, November 1996.

[6] A. Lee, A. Girgensohn, and K. Schlueter, "NYNEX portholes: initial user reactions and redesign implications," in *Proceedings of the International ACM SIGGROUP Conference on Supporting Group Work (GROUP '97)*, pp. 385–394, Phoenix, Ariz, USA, November 1997.

[7] Q. Zhao and J. Stasko, "The awareness-privacy tradeoff in video supported informal awareness: a study of image-filtering based techniques," Tech. Rep. GIT-GVU-98-16, Graphics, Visualization, and Usability Center, Atlanta, Ga, USA, 1998.

[8] E. M. Newton, L. Sweeney, and B. Malin, "Preserving privacy by de-identifying face images," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 2, pp. 232–243, 2005.

[9] M. Boyle, C. Edwards, and S. Greenberg, "The effects of filtered video on awareness and privacy," in *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW '00)*, pp. 1–10, Philadelphia, Pa, USA, December 2000.

[10] J.-C. Terrillon, M. N. Shirazi, H. Fukamachi, and S. Akamatsu, "Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images," in *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 54–61, Grenoble, France, March 2000.

[11] D. Chen and J. Yang, "Online learning of region confidences for object tracking," in *Proceedings of the 2nd Joint IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS '05)*, pp. 1–8, Beijing, China, October 2005.

[12] K.-K. Sung and T. Poggio, "Example-based learning for view-based human face detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 39–51, 1998.

[13] H. A. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 23–38, 1998.

[14] E. Osuna, R. Freund, and F. Girosi, "Training support vector machines: an application to face detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '97)*, pp. 130–136, San Juan, Puerto Rico, USA, June 1997.

[15] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '01)*, vol. 1, pp. 511–518, Kauai, Hawaii, USA, December 2001.

[16] H. Schneiderman and T. Kanade, "A statistical method for 3D object detection applied to faces and cars," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '00)*, vol. 1, pp. 746–751, Hilton Head Island, SC, USA, June 2000.

[17] S. Gong, S. McKenna, and J. J. Collins, "An investigation into face pose distributions," in *Proceedings of the 2nd International Conference on Automatic Face and Gesture Recognition*, pp. 265–270, Killington, Vt, USA, October 1996.

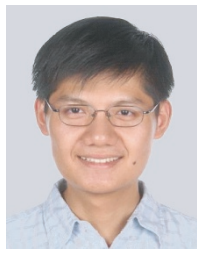
[18] G. D. Hager and K. Toyama, "X vision: a portable substrate for real-time vision applications," *Computer Vision and Image Understanding*, vol. 69, no. 1, pp. 23–37, 1998.

[19] Y. Raja, S. J. McKenna, and S. Gong, "Tracking and segmenting people in varying lighting conditions using colour," in *Proceedings of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 228–233, Nara, Japan, April 1998.

[20] K. Schwerdt and J. L. Crowley, "Robust face tracking using color," in *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 90–95, Grenoble, France, March 2000.

[21] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, "Pfinder: real-time tracking of the human body," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 780–785, 1997.
[22] A. Gelb, Ed., *Applied Optimal Estimation*, MIT Press, Cambridge, Mass, USA, 1992.
[23] A. Elgammal, R. Duraiswami, D. Harwood, and L. S. Davis, "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance," *Proceedings of the IEEE*, vol. 90, no. 7, pp. 1151–1163, 2002.
[24] R. Yan, J. Zhang, J. Yang, and A. Hauptmann, "A discriminative learning framework with pairwise constraints for video object classification," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '04)*, vol. 2, pp. 284–293, Washington, DC, USA, June-July 2004.
[25] G. Kimeldorf and G. Wahba, "Some results on Tchebycheffian spline functions," *Journal of Mathematical Analysis and Applications*, vol. 33, no. 1, pp. 82–95, 1971.
[26] J. K. Hodgins, J. F. O'Brien, and J. Tumblin, "Perception of human motion with different geometric models," *IEEE Transactions on Visualization and Computer Graphics*, vol. 4, no. 4, pp. 307–316, 1998.
[27] J. W. Davis and A. F. Bobick, "The representation and recognition of human movement using temporal templates," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '97)*, pp. 928–934, San Juan, Puerto Rico, USA, June 1997.

Datong Chen is a Systems Scientist in the Computer Science Department of the Carnegie Mellon University. He got his Ph.D. degree from Swiss Federal Institute of Technology in 2003, and M.S. and B.E. degrees from Harbin Institute of Technology in 1997 and 1995, respectively. Before doing his Ph.D. degree, he worked in the Telecooperation Office of the University of Karlsruhe. His research interests focus on assistive technology, pattern analysis, multimedia data mining, and statistical machine learning.



Yi Chang was born in Hunan Province, China. He received his B.S. degree in computer science from Jilin University, Changchun, China, in 2001, and M.S. degree from Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2004, and M.S. degree in Carnegie Mellon University, Pittsburgh, Pa, in 2006. His research interests include information retrieval, multimedia analysis, natural language processing, and machine learning.

Rong Yan is a Research Staff Member in IBM TJ Waston Research Center, Hawthorne, NY. He obtained his Ph.D. degree in language and information technologies from Carnegie Mellon University in 2006 and a B.E. degree in computer science from Tsinghua University, Beijing, in 2001. His research interests include multimedia retrieval, video content analysis, and machine learning. He is the author/coauthor of a book chapter and more than 35 refereed journal and conference publications. He received the ACM Multimedia Best Paper Runner-Up Award in 2004.

Jie Yang is a Senior Systems Scientist in the Human-Computer Interaction Institute, Carnegie Mellon University. He obtained his Ph.D. degree in electrical engineering from University of Akron, Akron, Ohio, in 1991. He joined the Interactive Systems Lab in 1994, where he has been leading research efforts to develop visual tracking and recognition system for multimodal human-computer interaction. His research interests are multimodal interfaces, computer vision, and pattern recognition.