PEOPLE IDENTIFICATION WITH LIMITED LABELS IN PRIVACY-PROTECTED VIDEO

Yi Chang, Rong Yan, Datong Chen, Jie Yang

School of Computer Science, Carnegie Mellon University
Pittsburgh, PA 15213
{changyi, yanrong , datong , yang+}@cs.cmu.edu

ABSTRACT

People identification is an essential task for video content analysis in a surveillance system. A good classifier, however, requires a large amount of training data, which may not be obtained in some scenario. In this paper, we propose an approach to augment insufficient training data with pairwise constraints that can be offered from video images that have removed people's identities by masking faces. We show user study results that human subjects can perform reasonably well in labeling pairwise constraints from face masked images. We also present a new discriminative learning algorithm that can handle uncertainties in pairwise constraints. The effectiveness of the proposed approach is demonstrated using video captured from a nursing home environment. The new method provides a way to obtain high accuracy of people identification from limited labeled data with noisy pairwise constraints, and meanwhile minimize the risk of exposing people's identities.

1. INTRODUCTION

Nowadays balancing the profusion of video surveillance and the expectation of individual privacy protection is becoming an urgent requirement in video analysis applications. As we know, the research monitored by IRB (Institutional Review Board) must protect privacy of human subjects as its obligation in processing data. Therefore, some attributes of the data, such as the identities and certain activities of human subjects, need to be protected from unauthorized personnel. A simple way of privacy protection is to mask faces in video images when the data is shown to unauthorized personnel. In some applications we may even need to remove some people from video. For example, video/audio analysis can be a very useful assistive tool for geriatric care. However, some of the patients living in the facility, who might not want to participate in the studies, are also captured by video cameras. Figure 1.1 shows an example of protecting the privacy of a patient unwilling to be included in the study. The left figure is the original image while the figure on the right is the expected result of privacy protection.

Automatic people identification is essential for many video analysis applications including privacy protection. Medical studies usually need to conduct a long-term recording (*e.g.*, a month or a few months) and thus produce a huge amount of video data. Manually identify human subjects in such prolonged video is a very difficult task, if not impossible. However, constructing automatic people identification also encounters the difficulty of privacy issue. On one hand, training a good people identification system requires a large amount of training data. On the other hand, in order to protect the privacy, only authorized personnel (a very limited number of doctors and nurses) are allowed to observe the unprotected data. It is difficult for authorized personnel to label such large mount of data.

Figure 1.1 An Illustration of Privacy Protection from Recorded Video

In this paper, we propose a method that can augment training data for training a people identification system from pairwise constraints labeled by unauthorized personnel from face masked data. Human faces in video images can be located by an automatic face detector and masked accordingly before showed to unauthorized personnel. The method exploits the human power of unauthorized personnel in labeling data without exposing identities of protected subjects. We perform user study to verify the hypothesis that human subjects can perform reasonably well in labeling pairwise constraints from face masked images. However, we cannot use the existing learning algorithm directly because of uncertainties in the pairwise constraints. We develop a new discriminative learning algorithm that can handle imperfect pairwise constraints and demonstrate the effectiveness of the proposed approach using video data captured from a nursing home environment.

In the previous research, quite a few researchers took account of privacy protection in video from different points of view. Senior et al. [8] presented a model to define video privacy, and implemented some elementary tools to re-render the video in a privacy-preserving manner. Tansuriyavong et al. [9] proposed a system that automatically identifies a person by face recognition, and displays the silhouette image of the person with a name list to balance the privacy-protecting and information-conveying. Brassil [2] implemented a system to permit individuals to protect privacy from video surveillance with the usage of mobile communications. Zhang et al. [11] 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 [4, 6, 12] discussed the privacy issue in the computer supported cooperative work domain. Furthermore, Newton et al. [7] proposed an algorithm to preserve privacy by de-identifying facial images. However, this algorithm is only apt to static facial images with high resolution, but not to dynamic video data with relatively low quality.

2. LABELING PAIRWISE CONSTRAINTS WITHOUT EXPOSING PEOPLE IDENTITIES

People identification is an essential task for video content analysis and privacy protection. In order to train a classifier for each person over a long period of time, we need a large amount of training data. Only authorized personnel (doctors or nurses) can label the training data because of the requirement of privacy protection. However, usually it is unlikely to ask these authorized personnel to label a large mount of training data. If we ask for unauthorized personnel to label more training data on the video data directly, nevertheless, it would severely break the privacy protection policy. Therefore the problem becomes how to obtain a good classifier with minimal efforts from authorized personnel in labeling training data and minimal risk of exposing identities of protected subjects to unauthorized personnel.

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. We propose a novel method to learn a classifier with two sets of labels to balance the insufficient training data and the privacy protection. Figure 2.1 shows the flow chart of the learning process, where privacy protection data could prevent unauthorized personnel to see the identities of protected people.

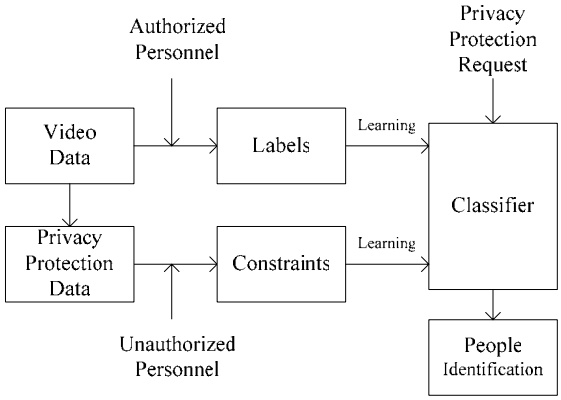


Figure 2.1 The Flow Chart of the Learning Process

A pairwise constraint between two examples demonstrates whether they belong to the same class or not. It provides some weak information in form of the relationship between the labels rather than the labels themselves. We use the imperfect pairwise constraint to model the imperfect pairwise labels from the user study. That is, we use two different sets of labeled data to build the classifier: a 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 masked face. During the process of labeling pairwise constraints, the system will automatically mask human faces using a face detection algorithm. We perform user study to test the hypothesis that human subjects can perform reasonably well in labeling pairwise constraints from face masked images.

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 face of the person, because they could use clothes, shape or gesture as the alternative information to make decision on pairwise constraints. To validate our hypothesis, we performed the following user study. To minimize the exposure of people identities in the labeling process, we only display the human silhouette images with blanked faces on the user interface shown to the unauthorized personnel. Firstly, we implemented a tool for user study of which the interface is displayed in Figure 2.2. The image on the top left side is the sample image, while the other images are all candidates to be compared with the sample images. In the experiments, volunteers are requested to label whether the candidate images are of the same person with the sampled image. All images are randomly selected from pre-extracted silhouette images and all candidate images do not belong to the same sequence as to the sample image. There are two modes in our user study tool. In the complex mode, there could be multiple candidate images mapping to the sample image, while in the simplified mode, only one candidate image that matching the sample image. Current user studies take the simplified mode as the basic test bed on the static images. In more details, the displayed images are randomly selected from a pool of 102 images, each of which is sampled from a different sequence of video.

According to the result of our user study, nine unauthorized personnel take a total of 180 runs to label the constraints with an overall accuracy around 88.89%, which supports the assumption that users could successfully label the pairwise constraints without exposing people identities, although these constraints are not fully correct. Finally, after filtering out the redundant labels, we obtained 140 correct constraints and 20 mistakenly labeled constraints.

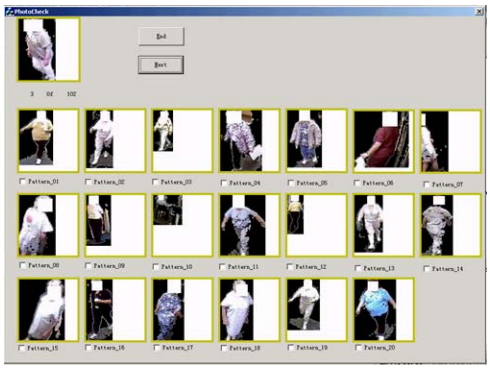


Figure 2.2 The Interface of User Study

3. DISCRIMINATIVE LEARNING WITH NOISY PAIRWISE CONSTRAINTS

In this section, we attempt to incorporate the additional pairwise constraints obtained from unauthorized personnel into a margin-based discriminative learning framework to boost the performance of people labeling. 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:

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

where x_i is the feature of the i^{th} training example, y_i denotes the corresponding label, and $f(x)$ is the classifier outputs. L denotes the empirical loss function, and $\lambda\Omega(\| f \|)$ 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. [10] extended above optimization objectives by introducing pairwise constraints as another set of empirical loss function,

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

where $L'(c_{ij}, f(x_i), f(x_j))$ is called pairwise loss function, and c_{ij} is a pairwise constraint between the i^{th} example and j^{th} 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. [10] choose the loss function to be a monotonic decreasing function of the difference between the predictions of a pair of pairwise constraints, *i.e.*,

$$L'(c_{i,j}, f(x_i), f(x_j)) = L(f(x_i) - c_{ij} f(x_j)) + L(c_{ij} f(x_j) - f(x_i)).$$

However, previous approaches usually assume the pairwise constraints are obtainable without any human errors. But our user study indicates that there are a small number of human errors on constraints labeling when faces are masked for privacy protection. Thereby we propose a new approach to improve discriminative learning with noisy pairwise constraints. In our approach, we introduce an additional term g_{ij} to model the uncertainty of each constraint achieved from the user study. The modified optimization objectives can be written as:

$$\frac{1}{m}\sum_{k=1}^{m} L(y_k, f(x_k)) + \frac{\mu}{|C|}\sum_{i,j} g_{ij} L'(c_{i,j}, f(x_i), f(x_j)) + \lambda \Omega(\| f \|_{\mathrm{H}})$$

where g_{ij} is the corresponding weight for the 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.

We normalize the sum of the pairwise constraint loss by the number of total constraints $|C|$ to balance the importance of labeling 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:

$$\frac{1}{m}\sum_{k=1}^{m} \log(1 + e^{-y_k f(x_k)}) + \frac{\mu}{|C|}\sum_{i,j} g_{ij} \log(1 + e^{f(x_i) - y_j f(x_j)})$$

$$+ \frac{\mu}{|C|}\sum_{i,j} g_{ij} \log(1 + e^{y_j f(x_j) - f(x_i)}) + \lambda \Omega(\| f \|_{\mathrm{H}})$$

The kernelized representation of the empirical loss function can be derived based on the Representer Theorem [5]. 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 of logistic regression loss function as the following formula,

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

$$+ \frac{\mu}{|C|} g_{ij} \cdot \vec{1}^{T} \log(1 + e^{-\alpha K'_P}) + \lambda \alpha K \alpha$$

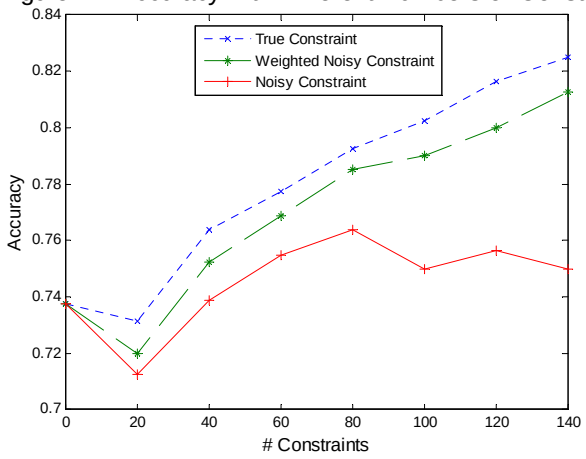
where Kp is the regressor matrix and K'p is the pairwise regressor matrix. Please see [10] for more details of their definitions. To solve the optimization problem, we apply the interior-reflective Newton methods to reach a global optimum. In the rest of this paper, we call this type of learning algorithms as weighted pairwise kernel logistic regression (WPKLR).

4. EXPERIMENTAL EVALUATION

In this paper, we applied the WPKLR algorithm to label the people identities from real surveillance video. We empirically chose the constraint parameter μ to be 20 and 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. Currently we particularly paid attention to video sequences that only contain 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 labeling 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.

Figure 4.1 Accuracy with Different Numbers of Constraint

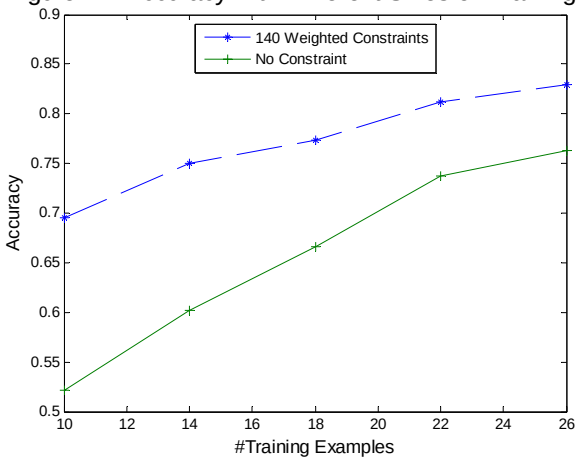


Our first experiment is to examine the effectiveness of pairwise constrains for labeling identities as shown in Figure 4.1. The learning curve of Noisy Constraint is completely based on the labeling result from the user study, but uniformly weighted all constraints as 1. Weighted Noisy Constraint uses 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 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 4.1 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 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 4.2 illustrates the performance with different number of training example. For all the settings, introducing 140 constraints could always substantially improve classification accuracy. Furthermore, pairwise constraints could even make more noticeable improvement given fewer training examples, which suggests constraints are helpful to reduce labeling efforts from authorized personnel.

Figure 4.2 Accuracy with Different Sizes of Training Set



5. CONCLUSION

In this paper, we present a framework to protect privacy of people in recorded video, and propose a learning method to label people from the video data with much less exposure of identities in the training process. Using only a small number of labeled data provided by the authorized personnel, we incorporate pairwise constraints that can be offered by a large group of unauthorized personnel even when they have no prior knowledge on the video data. Such kind of two-step labeling process could achieve minimal efforts from authorized personnel in labeling training data and minimal risk of exposing identities of protected people. According to our user study, we verify that human subjects could perform reasonably well in labeling pairwise constraints from face-masked images. Furthermore, we expand the learning methods to fit imperfect pairwise constraints, which could apply the pairwise constraint learning into more broad problems. Finally, we demonstrate the effectiveness of our automatic people labeling approach through the video captured from a nursing home environment.

With a good people identification algorithm, we could carry on a series of research on the recorded video, one of the future work includes developing a fully fledged privacy protection system that can be used for removing people from recorded video data. We will also explore different methods to minimize human efforts in labeling video data and minimize risk of exposing identities of protected people. Currently, we randomly choose the pairwise constraints for our user study, so the workload of constraint labeling processing is heavy. In the next step, we should also examine what are the most informative constraints, and select them to improve the efficiency of the unauthorized personnel labeling.

REFERENCES

[1] Boyle, M., Edwards, C., Greenberg, S. The Effects of Filtered Video on Awareness and Privacy. In Proc. of CSCW, 2000.

[2] Brassil, J. Using Mobile Communications to Assert Privacy from Video Surveillance. In Proc. of IPDPS, 2005.

[3] Cavallaro, A. Adding privacy constraints to video-based applications. European Workshop on the Integration of Knowledge, Semantics and Digital Media Technology, 2004.

[4] Hudson, S., Smith, I. Techniques for Addressing Fundamental Privacy and Disruption Tradeoffs in Awareness Support Systems. In Proc. of ACM Conference on Computer Supported Cooperative Work, 1996.

[5] Kimeldorf, G., Wahba, G. Some results on tchebycheffian spline functions. J. Math. Anal. Applic., vol.33, pp. 82-95, 1971.

[6] Lee, A., Girgensohn, A., Schlueter, K. NYNEX Portholes: Initial User Reactions and Redesign Implications. In Proc. of International Conference on Supporting Group Work, 1997.

[7] Newton, E., Sweeney, L., Malin, B. Preserving Privacy by De-identifying Facial Images. IEEE Transactions on Knowledge and Data Engineering, 17 (2) February 2005, pp. 232-243.

[8] Senior, A., Pankanti, S., Hampapur, A., Brown, L., Tian, Y., Ekin, A. Blinkering Surveillance: Enabling Video Privacy through Computer Vision. IBM Research Report, RC22886 (W0308-109), 2003.

[9] Tansuriyavong, S., Hanaki, S. Privacy protection by concealing persons in circumstantial video image. In Proc. of PUI, 2001.

[10] Yan, R., Zhang, J., Yang J., Hauptmann, A. A Discriminative Learning Framework with Pairwise Constraints for Video Object Classification. In Proc. of CVPR, 2004.

[11] Zhang, W., Cheung, S., Chen, M. Hiding Privacy Information In Video Surveillance System. In ICIP, 2005.

[12] Zhao, Q., Stasko, J. The Awareness-Privacy Tradeoff in Video Supported Informal Awareness: A Study of Image-Filtering Based Techniques. GVU Technical Report, GIT-GVU-98-16.