A Novel Approach for Privacy-Preserving Video Sharing

Jianping Fan, Hangzai Luo
Dept of Computer Science
UNC-Charlotte
Charlotte, NC 28223, USA
{jfan, hluo}@uncc.edu

Mohand-Said Hacid
UFR Informatique
Universite Claude Bernard
Lyon 1, Lyon, FRANCE
mshacid@bat710.univ-lyon1.fr

Elisa Bertino
Dept of Computer Science
Purdue University
W. Lafayette, IN 47907, USA
bertino@cs.purdue.edu

ABSTRACT

To support privacy-preserving video sharing, we have proposed a novel framework that is able to protect the video content privacy at the individual video clip level and prevent statistical inferences from video collections. To protect the video content privacy at the individual video clip level, we have developed an effective algorithm to automatically detect privacy-sensitive video objects and video events. To prevent the statistical inferences from video collections, we have developed a distributed framework for privacy-preserving classifier training, which is able to significantly reduce the costs of data transmission and reliably limit the privacy breaches by determining the optimal size of blurred test samples for classifier validation. Our experiments on a specific domain of *patient training and counseling videos* show convincing results.

Categories and Subject Descriptors
I.4.8 [**Image Processing and Computer Vision**]: Scene Analysis-*object recognition*, H.2.8 [**Database Management**]: Database Applications - *video databases*.

General Terms
Algorithms, Measurement, Experimentation

Keywords: Video content privacy, statistical inferences, privacy-preserving video sharing, unlabeled samples.

1. INTRODUCTION

Digital video plays an important role in supporting online patient training and counseling by illustrating real medical treatment procedures by means of videos [1]. In order to improve the quality of online patient training and counseling, it is very important to share patient training and counseling videos among multiple competitive groups and organizations (i.e., video owners). Increasing the amount of available videos results in a better offering for patients. However, privacy regulations, such as HIPAA, consumer backlash, and other privacy concerns often prevent multiple competitive video owners from sharing their patient training and counseling videos [2-4], because no good comprehensive framework is today available addressing all the following interrelated challenging problems:

(a) **Owner-Adaptive Video Privacy Modeling:** Many approaches to privacy protection have been recently developed [5-7]; these approaches however have a limited applicability because they do not cater for individual privacy preferences. We believe that a suitable approach for privacy-preserving video sharing must take into account a fundamental aspect of privacy well expressed by Alan Westin according to whom "privacy is the claim of individuals, groups, or institutes to determine for themselves when, how and to what extend information is communicated to others". We thus need techniques supporting *owner-adaptive video privacy modeling*.

(b) **Video Content Privacy Protection:** At the individual video clip level, the content privacy for the patient training and counseling videos encompasses two major aspects: (1) privacy for the human objects shown in video as the professional patient trainers or doctors; and (2) privacy for the human objects shown in video as the patients to illustrate the relevant clinic examples. In addition, video content privacy is also highly context-dependent and thus there is an urgent need to detect the privacy-sensitive video events. To protect the video content privacy, some techniques have been proposed that simply block the human faces [9-11]. However, in order to achieve more effective online patient training and counseling, it is also very important to let the patients see the real clinic examples at a high quality. Therefore, protecting the human object's privacy by simply blocking human faces may seriously reduce the video quality.

(c) **Statistical Inference Control:** Protecting the content privacy for individual video clips may not be enough to ensure privacy-preserving video sharing; we may also need to prevent statistical inferences from video collections [8, 17]. Statistical inferences represent an important challenge for video privacy protection because of non-sensitive data can be exploited to infer privacy-sensitive information. Such a challenge is beyond the reach of most existing privacy protection methods.

Based on these observations, we propose a novel framework able to assure the privacy of the video contents and control the statistical inferences in the specific domain of *patient training and counseling videos*. This paper is organized as follows. Section 2 presents the proposed framework for owner-adaptive video privacy modeling. To protect the

CIKM '05, October 31–November 5, 2005, Bremen, Germany

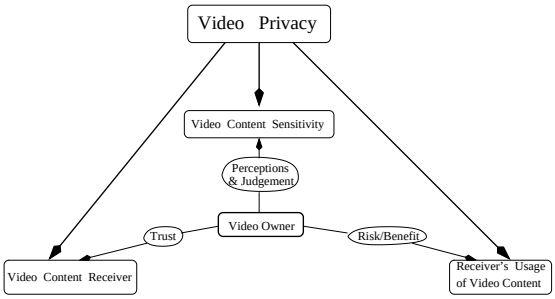

Figure 1: The proposed framework for owner-adaptive video privacy modeling.

video content privacy at the individual video clip level, Section 3 and Section 4 introduce a new algorithm for automatic detection of the privacy-sensitive video objects and video events. Section 5 introduces our approach to privacy-preserving video sharing for distributed classifier training and for preventing statistical inferences from video collections. Section 6 reports the results of an extensive evaluation we have performed on the proposed techniques. Finally, we conclude this paper in Section 7.

2. VIDEO PRIVACY MODELING

The definition of video privacy largely depends on three inter-related factors [10]: *video content sensitivity*, *video receiver*, and *receiver's usage of video contents*. Obviously, video privacy also depends on the video owner's *perceptions/judgement* of privacy of the videos being shared because privacy means different things to different people. In order to achieve owner-adaptive video privacy modeling, six inter-related factors need to be taken into account as shown in Fig. 1: *video content sensitivity*, *video receiver*, *receiver's usage of video contents*, *video owner's perceptions/judgement of video privacy*, *trust* between the video owner and the video receiver, and *risks/benefits* for video sharing. In addition, a good balance is crucial between the risks of privacy breaches and the benefits of video sharing. Based on this motivation, we propose a novel framework by taking all these inter-related factors into account in a comprehensive approach to achieve *owner-adaptive video privacy modeling*.

In order to implement the proposed framework for owner-adaptive video privacy modeling, we have defined a basic vocabulary of *privacy-sensitive video objects* in the specific domain of patient training and counseling videos; each video owner is thus able to select a subset of these privacy-sensitive video objects according to his/her individual privacy concerns.

3. VIDEO OBJECT DETECTION

To support our framework, the basic vocabulary of privacy-sensitive video objects is pre-defined by the video owners. To detect these privacy-sensitive video objects, video shots are first detected automatically [1]. To detect the privacy-sensitive video objects associated with each video shot, we have designed a set of automatic video object detection functions, where each video object detection function is able to detect only one certain type of these privacy-sensitive video objects in the basic vocabulary.

After the video shots are detected from a given video clip, our automatic video object detection functions are executed. To detect a given type of privacy-sensitive video object, our

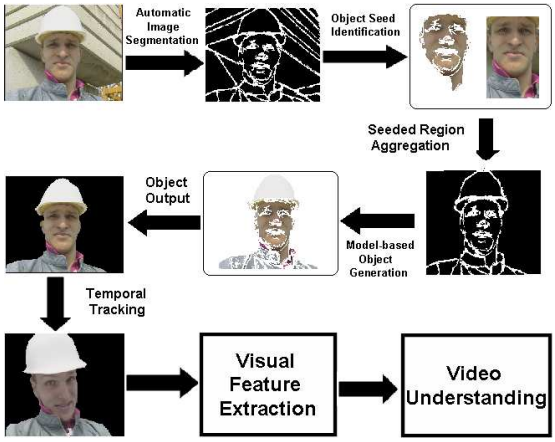

Figure 2: The flowchart of the proposed algorithm for human object detection and tracking.

detection function takes the following steps as shown in Fig. 2: (a) Automatic image segmentation is first performed on each video frame to obtain the homogeneous image regions [1, 16, 18]; (b) The given type of video object may have very different visual properties because of presence/absence of distinctive parts, variability in overall shape, changing appearance due to lighting conditions, viewpoints etc. Thus, automatic image segmentation itself is unable to directly detect the privacy-sensitive video objects and machine learning should be involved for region classification and object generation. Based on this understanding, the homogeneous image regions are classified into two classes by using SVM binary classifier, that is, into object regions *versus* non-object regions; (c) The connected object regions are then merged and aggregated according to a knowledge-based object model for generating the given type of privacy-sensitive video object; (d) Object tracking is finally performed to obtain the temporal relationships of object regions among video frames within the same video shot.

After the privacy-sensitive video objects are extracted, the original video streams are decomposed into a set of privacy-sensitive video objects such as human beings with race and gender, background, and areas of interest.

4. VIDEO EVENT DETECTION

Another difficulty in video privacy protection is that the video privacy is also highly context-dependent. To detect the privacy-sensitive video events, the video shots for a given video clip are classified into a set of pre-defined semantic video concepts that are sensitive to the context-dependent privacy breaches. The contextual relationships, between a given semantic video concept C_j and the relevant video objects, are interpreted by using a finite mixture model:

$$P(X, C_j, \Theta_{c_j}) = \sum_{i=1}^{\kappa_j} \omega_i P(X|C_j, \theta_i), \quad \sum_{i=1}^{\kappa_j} \omega_i = 1 \qquad (1)$$

where $\Theta_{c_j} = \{\kappa_j, \omega_{c_j}, \theta_{c_j}\}$ is the parameter set for model structure, weights, and model parameters; in particular, κ_j is the model structure, $\omega_{c_j} = \{\omega_1, \cdots, \omega_{\kappa_j}\}$ is the set of weights for κ_j mixture components; $\theta_{c_j} = \{\theta_1, \cdots, \theta_{\kappa_j}\}$ is the set of model parameters for κ_j mixture components, $P(X|C_j, \theta_i)$ is the mixture component (i.e., one specific video context class) that is used to approximate the class

distribution for one specific type of the relevant video objects, X is a set of m-dimensional object-based features.

4.1 Adaptive EM Algorithm

To learn the semantic video concept accurately, we have developed an ***adaptive EM algorithm*** to achieve more effective model selection and parameter estimation by using a maximum likelihood approach. Based on a limited number of labeled samples Ω_{c_j}, the optimal model parameters $\hat{\Theta}_{c_j}$ for the specific semantic video concept C_j are determined by:

$$\hat{\Theta}_{c_j} = arg \quad max \quad \{L(\Theta_{c_j})\} \tag{2}$$

where $L(\Theta_{c_j}) = -\sum_{X_i \in \Omega_{c_j}} \log P(X_i, C_j, \Theta_{c_j}) + \log p(\Theta_{c_j})$ is the objective function, $-\sum_{X_i \in \Omega_{c_j}} \log P(X_i, C_j, \Theta_{c_j})$ is the likelihood function, and $\log p(\Theta_{c_j}) = -\frac{m+\kappa_j+3}{2} \sum_{l=1}^{\kappa_j} \log \frac{N\omega_l}{12} - \frac{\kappa_j}{2} \log \frac{N}{12} - \frac{\kappa_j(N+1)}{2}$ is the *minimum description length* (MDL) term to penalize the complex models [14-15], N is the total number of samples that are used for classifier training, m is the feature dimensions.

To achieve more effective classifier training, our adaptive EM algorithm can re-organize the distribution of mixture components and select the optimal number of mixture components by performing automatic ***merging***, ***splitting*** and ***elimination*** of mixture components.

Our adaptive EM algorithm uses symmetric *Jensen-Shannon* (JS) *divergence* $JS(C_j, \theta_l, \theta_k)$ to measure the divergence between two mixture components $P(X|C_j, \theta_l)$ and $P(X|C_j, \theta_k)$ for the same concept model C_j.

$$\begin{aligned} JS(C_j, \theta_l, \theta_k) &= H(\pi_1 P(X|C_j, \theta_l) + \pi_2 P(X|C_j, \theta_k)) \\ &\quad -\pi_1 H(P(X|C_j, \theta_l)) - \pi_2 H(P(X|C_j, \theta_l)) \end{aligned} \tag{3}$$

where $H(P(\cdot)) = -\sum P(\cdot) \log P(\cdot)$ is the well-known Shannon entropy, π_1 and π_2 are the weights. In our experiments, we set $\pi_1 = \pi_2 = \frac{1}{2}$.

If the *intra-concept JS divergence* $JS(C_j, \theta_l, \theta_k)$ is too small, these two mixture components are strongly overlapped and may overpopulate the relevant sample areas; thus they are merged into a single mixture component $P(X|C_j, \theta_{lk})$. In addition, the *local JS divergence* $JS(C_j, \theta_{lk})$ is used to measure the divergence between the merged mixture component $P(X|C_j, \theta_{lk})$ and the local sample density $P(X, \theta_{lk})$. Our adaptive EM algorithm tests $\frac{\kappa_j(\kappa_j-1)}{2}$ pairs of mixture components that could be merged and the pair with the minimum value of the local JS divergence is selected as the best candidate for ***merging***.

Two types of mixture components may be ***split***: (a) The elongated mixture components which underpopulate the relevant samples (i.e., characterized by the local JS divergence); (b) The tailed mixture components which overlap with the mixture components from other concept models (i.e., characterized by the *inter-concept JS divergence*). To select the mixture component for splitting, two criteria are combined: (1) The *local JS divergence* $JS(C_j, \theta_i)$ to characterize the divergence between the ith mixture component $P(X|C_j, \theta_i)$ and the local sample density $P(X|\theta_i)$; (2) The inter-concept JS divergence $JS(C_j, C_h, \theta_i, \theta_m)$ to characterize the overlapping between the mixture components $P(X|C_j, \theta_i)$ and $P(X|C_h, \theta_m)$ from two relevant semantic video concepts C_j and C_h.

By *splitting* the elongated and tailed mixture components, some mixture components locating at the sample distribution boundary may be unrepresentative and be supported by few samples. If a specific mixture component is only supported by few samples, it may be removed from the underlying concept model. To determine the unrepresentative mixture component for ***elimination***, our adaptive EM algorithm uses the local JS divergence $JS(C_j, \theta_i)$ to characterize the representation of the mixture component $P(X|C_j, \theta_i)$ for the relevant samples. The mixture component with the maximum value of the local JS divergence is selected as the candidate for elimination.

To jointly optimize these three operations of merging, splitting and elimination, their probabilities are defined as:

$$\begin{cases} J_m(i,k,\theta_{ik}) &= JS(C_j, \theta_{ik}) + \varphi JS(C_j, \theta_i, \theta_k) \\ J_s(i,m,\theta_i) &= \frac{\varphi JS(C_j, C_h, \theta_i, \theta_m)}{JS(C_j, \theta_i)} \\ J_e(i,\theta_i) &= \frac{\varphi}{JS(C_j, \theta_i)} \end{cases} \tag{4}$$

where φ is a normalized factor and it is determined by:

$$\sum_{i=1}^{\kappa_j} J_e(i,\theta_i) + \sum_{i=1}^{\kappa_j} \sum_{k=i+1}^{\kappa_j} J_m(i,k,\theta_{ik}) + \sum_{i=1}^{\kappa_j} \sum_{m=1}^{\kappa_h} J_s(i,m,\theta_i) = 1 \tag{5}$$

The acceptance probability to prevent poor operation of merging, splitting or elimination is defined by:

$$P_{accept} = min\left(exp\left[-\frac{|L(C_j,\Theta_1) - L(C_j,\Theta_2)|}{\tau}\right], \quad 1\right) \tag{6}$$

where $L(C_j, \Theta_1)$ and $L(C_j, \Theta_2)$ are the objective functions for the models Θ_1 and Θ_2 (i.e., before and after performing the merging, splitting or elimination operation) as described in Eq. (2), τ is a constant that is determined experimentally. In our current experiments, τ is set as $\tau = 9.8$.

4.2 Learning with Unlabeled Samples

To learn the underlying concept model accurately, a large number of labeled samples is needed. When only a limited number of labeled samples is available for classifier training, it is difficult to select the optimal model structure and estimate the accurate model parameters. However, obtaining a large number of labeled samples is very expensive, and incorporating the outlying unlabeled samples for classifier training may lead to worse performance rather than improvement [12-13]. Thus, it is very important to develop new techniques able to eliminate the misleading effects of the outlying unlabeled samples.

After the weak classifier for the given semantic video concept C_j is learned from a limited number of available labeled samples, the Bayesian framework is used to achieve "soft" classification of unlabeled video clips. The confidence score for an unlabeled sample with the given semantic video concept C_j is defined as:

$$\psi(X_l, C_j, t) = \sqrt{\psi_\alpha(X_l, C_j, t)\psi_\beta(X_l, C_j, t)} \tag{7}$$

where $\psi_\alpha(X_l, C_j, t) = P(C_j|X_l, \Theta_{c_j})$ is the posterior probability for the unlabeled sample $\{X_l, S_l\}$ with the given semantic video concept C_j, $\psi_\beta(X_l, C_j, t) = -\log P(X_l, C_j, \Theta_{c_j})$ is the log-likelihood value of the unlabeled sample $\{X_l, S_l\}$ with the given semantic video concept C_h. For one specific unlabeled sample $\{X_l, S_l\}$, its confidence score $\psi(X_l, C_j, t)$ can be used as the criterion to indicate the possibility to be

taken as an **outlier** for the given semantic video concept C_j.

In order to eliminate the misleading effects of the outlying unlabeled samples for semi-supervised classifier training, the unlabeled samples are first categorized into two classes according to their confidence scores: (a) *certain unlabeled samples* with high confidence scores may originate from the known video context classes that have already been learned from the available labeled samples; (b) *uncertain unlabeled samples* with low confidence scores may originate from new concept, outliers or unknown video context classes that cannot be directly learned from a limited number of available labeled samples.

The certain unlabeled samples can be incorporated to improve the mixture density estimation incrementally (i.e., regularly updating the model parameters without changing the model structure) by reducing the density variance. With the updated concept model for the given semantic video concept C_j (i.e., incremental classifier), the confidence scores for some uncertain unlabeled samples may be changed over time when they originate from the unknown video context classes that cannot be interpreted intuitively by a limited number of labeled samples. For the uncertain unlabeled sample, the changing scale of its confidence scores with the given semantic video concept C_j is defined as:

$$y_l = |\psi(X_l, C_j, t+1) - \psi(X_l, C_j, t)| \tag{8}$$

where $y_l \geq 0$, $\psi(X_l, C_j, t)$ and $\psi(X_l, C_j, t+1)$ indicate its confidence scores with the same concept model C_j before and after the model update. The uncertain unlabeled samples with a large value of y_l may originate from the unknown video context classes induced by concept drift, and should therefore be used to achieve more accurate video concept interpretation and semi-supervised classifier training. Thus, we name the uncertain unlabeled samples with a large values of y_l as *informative unlabeled samples*. To address the concept drift problem, one or more new mixture components can be added to the residing areas for the informative unlabeled samples (i.e. ***birth***).

$$P(X, C_j, \Theta_{c_j}) = \omega_{\kappa_j+1} P(X|C_j, \theta_{\kappa_j+1})$$

$$+(1-\omega_{\kappa_j+1}) \sum_{l=1}^{\kappa_j} P(X|C_j, \theta_l)\omega_l \tag{9}$$

where ω_{κ_j+1} is the weight for the (κ_j+1)th mixture component $P(X|C_j,\ \theta_{\kappa_j+1})$ to characterize the appearance of unknown video context class for the given semantic video concept C_j.

On the other hand, the outlying unlabeled samples with the y_l value close to zero may originate from new concept or outliers. To eliminate the misleading effects of the outlying unlabeled samples, a penalty term γ_l is defined as:

$$\gamma_l = \begin{cases} 1, & certain \quad unlabeled \quad samples \\ \frac{e^{y_l} - e^{-y_l}}{e^{y_l} + e^{-y_l}}, & uncertain \quad unalbeled \quad samples \end{cases} \tag{10}$$

where $0 \leq \gamma_l \leq 1$, $\gamma_l = 0$ if $y_l = 0$. Thus, the penalty term γ_l can provide an effective solution to select the *informative unlabeled samples* for semi-supervised classifier training.

To avoid the problem of *overfitting the unlabeled samples*, the MDL term for model selection is updated by including the size of unlabeled samples which have nonzero-value of γ_l. In addition, the likelihood function as described in Eq. (2) is replaced by a joint likelihood function for both the labeled samples and the unlabeled samples. Thus, the joint objective function is defined as:

$$L(C_j, \Theta_{c_j}) = \log p(\Theta_{c_j}) - \sum_{X_i \in \Omega_{c_j}} \log P(X_i|C_j, \theta_l)\omega_l$$

$$-\lambda \sum_{X_n \in \overline{\Omega_{c_j}}} \gamma_n \log \left(\sum_{m=1}^{\kappa_j} P(X_n|C_j, \theta_m)\omega_m \right) \tag{11}$$

where the discount factor $\lambda = \frac{N_u}{N_L + N_u}$ is used to control the relative contribution of the unlabeled samples for semi-supervised classifier training, N_u is the total number of unlabeled training samples, N_L is the total number of labeled training samples. Using the joint objective function in Eq. (11) to replace the objective function in Eq. (2), our adaptive EM algorithm is applied to the mixture training sample set, both originally and probabilistically labeled, to learn the classifier accurately.

Once the classifiers for the semantic video concepts of particular interest are available, they are used to classify the video shots for the given video clip into a set of privacy-sensitive semantic video concepts. Semantic understanding of the given video clip is thus achieved. The context-dependent video shots that are mapped onto the same privacy-sensitive semantic video concept are then merged as the *privacy-sensitive video event*. Our experimental results for privacy-preserving video event detection are given in Fig. 3.

Figure 3: The experimental results for video event detection: (a) standing; (b) walking; (c) picking up; (d) carrying.

4.3 Video Content Privacy Protection

Once the detection functions for the privacy-sensitive video objects and video events are available, they are used to protect the content privacy at the individual video clip level. To filter out the privacy-sensitive human objects (i.e., doctors, professional patient trainers, patients in video), we use digital human models (i.e., virtual human objects) to replace the appearances of privacy-sensitive human objects in video. Thus, the blurred video streams are able to protect the privacy-sensitive information about who are in the

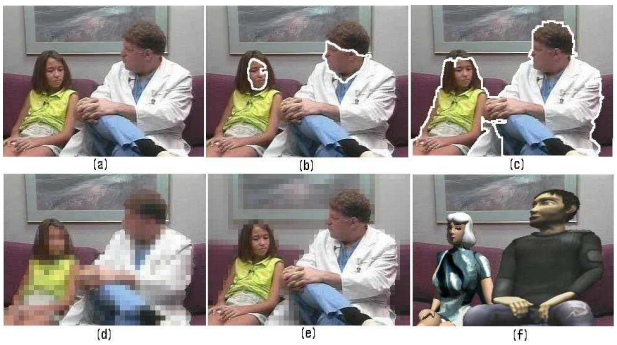

Figure 4: Experimental results for video content privacy protection: (a) original video; (b) face detection; (c) object detection; (d) simple object blocking; (e) simple background blocking; (f) virtual objects.

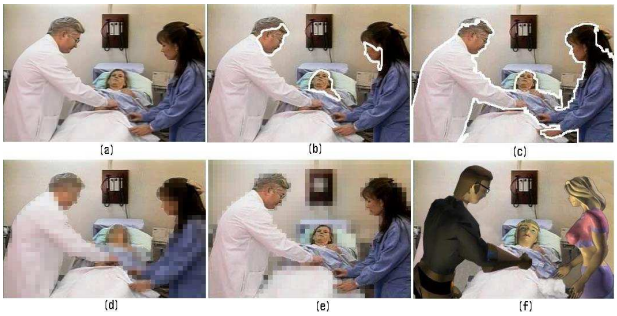

Figure 5: Experimental results for video content privacy protection: (a) original video; (b) face detection; (c) object detection; (d) simple object blocking; (e) simple background blocking; (f) virtual objects.

video scene. The blurred video streams are still able to provide enough non-sensitive information about the real medical treatment procedure for one certain infectious disease and enable high-quality online patient training and counseling. In addition, the blurred video streams are able to provide the non-sensitive information about the number of people in the scene and a rough idea about their postures, but it makes impossible for the receivers to guess who these persons are because no image details are conveyed in the blurred video streams. Experimental results on video content privacy protection are given in Fig. 4, Fig. 5, and Fig. 6.

To protect the context-dependent video content privacy, a set of video shots that are relevant to the detected privacy-sensitive video events are removed from the original video clip, and the residual non-sensitive video shots are re-packaged as a new MPEG video stream.

5. STATISTICAL INFERENCE CONTROL

To support more effective online patient training and counseling, it is very important to enable privacy-preserving video sharing among multiple competitive groups and organizations. However, sharing large-scale video clips may induce the privacy breaches because the dishonest users may use statistical inference techniques to infer the individual video owner's privacy. To prevent the statistical inferences from video collections, we propose a distributed framework that enables a ***privacy-preserving classifier training*** by treating κ individual video owners as κ horizontally partitioned data sources. For a given semantic video concept C_j, each

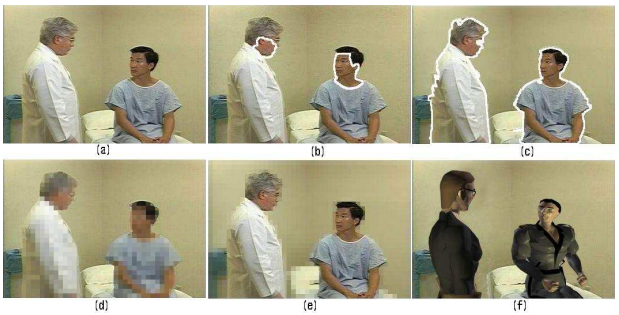

Figure 6: The experimental results for video content privacy protection: (a) original video; (b) face detection; (c) object detection; (d) simple object blocking; (e) simple background blocking; (f) virtual objects.

video owner can independently learn an individual *weak concept model* (i.e., local classifier) by using his/her own training samples (as shown in Fig. 7). Our model-based classifier training technique described in Section 4 can be used to select the optimal model structures and to estimate the accurate model parameters for these κ weak concept models.

In order to achieve more accurate classification of distributed video contents, it is very important to learn the classifier accurately by collecting the training samples from all these κ video owners. However, sending the training samples to the central site is undesirable from the privacy perspective because the dishonest users may be able to infer the individual video owners' privacy-sensitive information by using statistical inference techniques. To prevent statistical inference from video collections, we have proposed a distributed approach to enable privacy-preserving classifier training. Instead of sending the original training samples to the central site, each individual video owner has to send his/her weak concept model to the central site for *combined classifier training* (i.e., learning *global concept model* for accurately interpreting the given semantic video concept).

To enable privacy-preserving distributed classifier training, ***virtual samples*** are directly generated from the available weak concept models at the central site by using Markov Chain Monte Carlo sampling technique [7]. We call these training samples generated from the κ weak concept models at the central site as the *virtual samples* because they are not obtained directly from the original video streams. The virtual samples asymptotically have the same statistical properties as the original video data because both of them originate from the same mixture density function (i.e., same weak concept model). Such virtual samples are thus able to effectively train the combined classifier [7]. In addition, the virtual samples are also sufficiently different from the original video data and thus they are able to protect the privacy of the original video streams. Without having available the blurred video streams, it is impossible for the dishonest users at the central site to reliably relate the virtual samples to the original video streams and to violate the individual video owner's privacy. Thus, generating the virtual samples from these κ weak concept models at the central site can significantly reduce the privacy breaches and can also drastically reduce the costs for data transmission.

Based on these observations, our framework for combined classifier training takes the following steps: (a) The mixture components from the κ weak concept models are combined to obtain a "pseudo-complete" global concept model

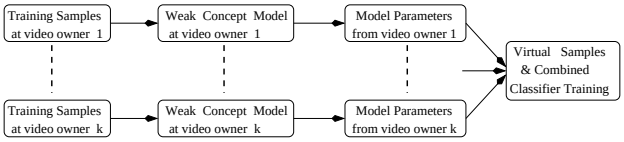

Figure 7: The distributed framework for privacy-preserving classifier training.

for interpreting the given semantic video concept C_j more accurately. The virtual samples from these κ weak concept models are integrated as the *combined virtual samples* to learn the underlying global concept model for accurately interpreting the given semantic video concept C_j. (b) Based on the available mixture components shared from these κ weak concept models, our adaptive EM algorithm is used to select the optimal model structure and estimate the accurate model parameters for the global concept model by performing automatic *merging*, *splitting*, and *elimination* of mixture components. (c) The mixture components with less prediction power on the combined virtual samples are eliminated. The overlapped mixture components from different weak concept models are merged into a single mixture component. The elongated mixture components that underpopulate the combined virtual samples are split into multiple representative mixture components.

By integrating all these κ weak concept models shared from κ data sites, the global concept model for interpreting the given pattern or concept C_j is defined as:

$$P(X, C_j, \Theta_{c_j}) = \sum_{l=1}^{\kappa_{c_j}} P(X|C_j, \theta_l)\omega_l$$

where $\kappa_{c_j} = \sum_{h=1}^{\kappa} M_h$ is the total number of the mixture components shared from the κ individual data sites, $M_h \leq \kappa_h$, and M_h is the number of mixture components shared from the hth data site (i.e., the hth weak concept model has κ_h mixture components totally).

If one mixture component, $P(X|C_j, \theta_m)$, is ***eliminated***, the global concept model for accurately interpreting the given pattern or concept C_j is then refined as:

$$P(X, C_j, \Theta_{c_j}) = \frac{1}{1-\omega_m} \sum_{l=1}^{\kappa_{c_j}-1} P(X|C_j, \theta_l)\omega_l, \ m \neq l \quad (12)$$

If two mixture components $P(X|C_j, \theta_m)$ and $P(X|C_j, \theta_l)$ are ***merged*** as a single mixture component $P(X|C_j, \theta_{ml})$, the global concept model for accurately interpreting the given semantic video concept C_j is refined as:

$$P(X, C_j, \Theta_{c_j}) = \sum_{h=1}^{\kappa_{c_j}-2} P(X|C_j, \theta_h)\omega_h + P(X|C_j, \theta_{ml})\omega_{ml} \quad (13)$$

If one mixture component, $P(X|C_j, \theta_h)$, is ***split*** into two new mixture components, $P(X|C_j, \theta_r)$ and $P(X|C_j, \theta_t)$, the global concept model for accurately interpreting the given semantic video concept C_j is refined as:

$$P(X, C_j, \Theta_{c_j}) = \sum_{h=1}^{\kappa_{c_j}-1} P(X|C_j, \theta_h)\omega_h + P(X|C_j, \theta_r)\omega_r + P(X|C_j, \theta_t)\omega_t \quad (14)$$

By using our adaptive EM algorithm to directly combine these κ weak concept models that are independently learned from κ individual video sources, our framework for combined classifier training is expected to derive the global concept model able to interpret the given semantic video concept C_j more accurately.

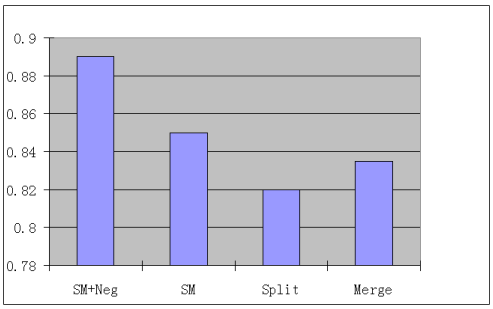

Figure 8: The classifier performance (i.e., precision ρ) under different combinations of operations: parkinson disease self-treatment.

In order to validate the combined classifier (i.e., global concept model) at the central site, each individual video owner has to share a limited number of *blurred test samples*. However, it is impossible to prevent misuse of these blurred test samples once they are released. In order to prevent statistical inferences from the blurred test samples at the central site, we have proposed a novel approach whose goal is to estimate the optimal size of such samples. Such optimally sized samples are able to prevent statistical inferences while reliably validating the combined classifier.

Because each individual video owner O_l sends not only the blurred test sample set S but also his/her weak concept model to the central site, it is possible for the dishonest users to incorporate the O_l's weak concept model with his/her blurred test sample set S to infer the O_l's private information. In order to present our approach for preventing statistical inferences, we first need to define a metrics to estimate the individual video owner O_l's *privacy disclosure* induced when sharing the weak concept model and the blurred test sample set S with size n. The metrics we adopt is defined as follows:

$$\rho(C, n, O_l) = |H(P(C, O_l|X, \Theta_c, S)) - H(P(C))| \quad (15)$$

where $H(\cdot)$ is the well-known Shannon entropy, $P(C, O_l|X, \Theta_c, S)$ is the posterior probability of the users' prediction of the O_l's privacy C after exploiting the O_l's blurred test sample set S and his/her weak concept model, $P(C)$ is the prior probability of the users' prediction of the O_l's privacy C.

To incorporate the O_l's blurred test sample set S for classifier validation, it is very important to determine what size of the blurred test sample set gives statistically significant validation results while preventing the dishonest users from inferring the individual video owner's private information. We use the well-known distribution-independent bound (i.e., Chebychev inequality [8]) to determine the minimum size n_{min} of the blurred test sample set S:

$$Prob(p - \hat{p} \geq \frac{\sigma}{\sqrt{2\alpha n_{min}}}) \leq \alpha \quad (16)$$

where $Prob(\cdot)$ is the underlying probability distribution such as Gaussian distribution, α is the pre-defined bound for the expected error rate of the combined classifier, $0 \leq \alpha \leq 1$, p

is the error rate.

$$\begin{cases} \hat{p} & = & \frac{1}{n_{min}} \sum_{i=1}^{n_{min}} x_i \\ \sigma^2 & = & \frac{1}{n_{min}-1} \sum_{i=1}^{n_{min}} (\hat{p} - x_i)^2 \end{cases} \quad (17)$$

where $\hat{p}$ is the average error rate of the blurred test sample set S, σ^2 is the variance of the blurred test sample set S.

On the other hand, it is also critical to determine the maximum size n_{max} of the blurred sample set S that may result in privacy breaches [8, 17]; this size is etimated as follows:

$$\varrho(C, n_{max}, O_l) = \inf_{X \in S} \left\{ \sum_{\Theta_c \in \Re} P(C, O_l | X, \Theta_c, S) \right\} \leq \delta \quad (18)$$

where δ is the pre-defined confidence bound, $P(C, O_l|X, \Theta_c, S)$ is the highest posterior probability for the user's prediction of the O_l's privacy C, and $\Re$ is credible set for the potential predictors which have the highest posterior probability close to δ. In our experiments, we set $\delta = 50\%$ so that data mining tools cannot obtain reliable results [5-12].

Thus, the optimal size $n_{optimal}$ of the blurred test sample set S is determined by the following low and up bounds:

$$n_{optimal} \in [n_{min}, \quad n_{max}) \quad (19)$$

where the n_{min} and n_{max} are the low and up bounds determined by Eqs. (16) and (18).

To achieve a good balance between limiting the privacy breaches and enabling reliable classifier validation, the optimal size $n_{optimal}$ of the O_l's blurred test sample set S to be shared is determined by an optimization procedure:

$$\begin{aligned} & Min\{\rho(C, n_{optimal}, O_l)\} \\ subject\ to: & \\ & n_{optimal} \in [n_{min}, n_{max}) \end{aligned} \quad (20)$$

By optimizing the criterion given by Eq. (20), the optimal size $n_{optimal}$ of the blurred test samples from each individual video owner, that are necessary to reliably validate the combined classifier while limiting the privacy breaches, can be obtained accurately.

By determining the optimal size of the blurred test samples to be shared, our framework is able to enable *privacy preserving distributed classifier training* and to effectively prevent statistical inferences from video collections.

6. ALGORITHM EVALUATION

Our experimental *algorithm evaluation* focuses on: (a) evaluating the performance of our adaptive EM algorithm with different combinations of merging, splitting and elimination; (b) evaluating the performance of our classifier training technique when using different sizes of unlabeled samples; (c) evaluating our distributed framework for privacy preserving classifier training to prevent statistical inferences.

The *benchmark metric* for the classifier evaluation includes *precision* ρ and *recall* ϱ. They are defined as:

$$\rho = \frac{\vartheta}{\vartheta + \gamma}, \qquad \varrho = \frac{\vartheta}{\vartheta + \nu} \quad (21)$$

where ϑ is the set of true positive samples that are related to the corresponding concept and are classified correctly, γ is the set of true negative samples that are irrelevant to the corresponding concept and are classified incorrectly, and ν is the set of false positive samples that are related to the corresponding concept but are misclassified.

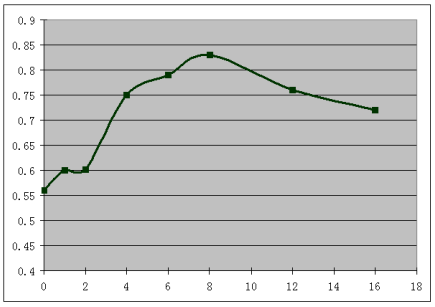

Figure 9: The empirical relationship between the classifier performance (i.e., precision ρ) and the ratio between the unlabeled samples and the labeled samples, ρ versus $\lambda' = \frac{N_u}{N_L}$: parkinson disease self-treatment.

In our adaptive EM algorithm, multiple operations, such as merging, splitting, and elimination, have been integrated to re-organize the distributions of mixture components, select the optimal model structure and construct more flexible decision boundaries among different concepts according to the real class distributions of the training samples. Thus, our adaptive EM algorithm is expected to have better performance than the traditional EM algorithm and its recent variants [14-15].

In order to evaluate the real benefits of the integration of these three operations (i.e. merging, splitting, and elimination), we have tested the performance differences of our adaptive EM algorithm with different combinations of these three operations. As shown in Fig. 8, we have tested the performance of the classifiers under different combinations of three operations: only splitting, only merging, combining splitting and merging (i.e. SM), combining splitting, merging and elimination (i.e., SM + Neg). From these experimental results, one can find that our adaptive EM algorithm can improve the classifiers' performance significantly.

Given a limited number of labeled samples, we have tested the performance of our classifiers by using different sizes of unlabeled samples for classifier training (i.e. with different size ratios $\lambda' = \frac{N_u}{N_L}$ between the unlabeled samples N_u and the labeled samples N_L). The average performance differences are given in Fig. 9 and Fig. 10.

When a limited number of labeled samples is available and more unlabeled samples are involved for semi-supervised classifier training (i.e., $\lambda' = \frac{N_u}{N_L}$ becomes bigger), we have also obtained a decrease in the classifier's performance because large-scale outlying unlabeled samples have dominated the statistical properties of the joint class distribution and misleaded the classifier. Ideally, it is possible for the dishonest users to integrate large-scale non-sensitive data (i.e., unlabeled samples) with a limited number of privacy-sensitive blurred test samples (i.e., labeled samples) to infer the individual video owner's privacy. However, this empirical observation (i.e., decrease of prediction accuracy) has provided very convincing evidence for the efficiency of our proposed solution on preventing statistical inferences: *It is impossible for the dishonest users to obtain reliable results when only a limited number of blurred test samples are shared.*

To evaluate our distributed framework for privacy preserving classifier training, we partitioned each data set into three individual groups and performed classifier training on

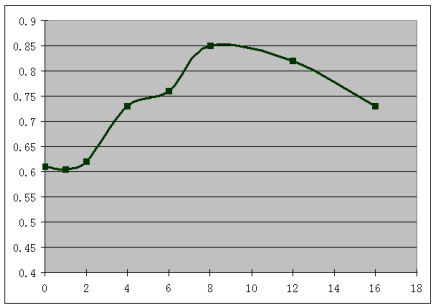

Figure 10: The empirical relationship between the classifier performance (i.e., precision ρ) and the ratio between the unlabeled samples and the labeled samples, ρ versus $\lambda' = \frac{N_u}{N_L}$: diabetic disease self-treatment.

these three individual data groups independently. We have obtained the empirical relationships between the quality of the global concept model (i.e., precision of the combined classifier) and the privacy disclosures as shown in Fig. 11.

For validating the combined classifier at the central site, each individual video owner has to share not only his/her weak concept model but also a limited number of blurred test samples. To prevent statistical inferences, we have also obtained the empirical relationships between the privacy disclosures and the number of blurred test samples to be shared as shown in Fig. 12. One can find that sharing more blurred test samples decreases the individual video owner's ability on controlling the statistical inferences and results in the privacy breaches.

7. CONCLUSIONS

To enable privacy-preserving video sharing among multiple competitive groups and organizations, we have proposed a novel framework able to both protect the video content privacy and control the statistical inferences. By detecting the privacy-sensitive video objects and video events automatically, our proposed algorithm is able to effectively protect the video content privacy at the individual video clip level. By determining the optimal size of blurred test samples for classifier validation, our proposed framework for privacy-preserving distributed classifier training is able to not only limit the privacy breaches but also improve the classifier's accuracy significantly. Our experiments in the specific domain of online patient training and counseling videos show that our techniques are effective.

8. REFERENCES

[1] J. Fan, H. Luo, A.K. Elmagarmid, "Concept-oriented indexing of video databases: towards semantic sensitive retrieval and browsing", *IEEE Trans. on Image Processing*, vol.13, no.7, pp.974-992, 2004.

[2] E. Newton, L. Sweeney, B. Malin, "Preserving privacy by de-identifying facial images", Technical Report CMU-CS-03-119, 2003, also presented at *IEEE Trans. on Knowledge and Data Engineering, 2005.*

[3] J. Wickramasuriya, M. Datt, S. Mehrotra, N. Venkatasubramanian, "Privacy protecting data collection in media spaces", ACM Multimedia, 2004.

[4] A. Senior, S. Pankanti, A. Hampapur, L. Brown, Y. Tian, A. Ekin, "Blinkering surveillance: enabling video privacy through computer vision", IBM TR W0308-109, 2003.

[5] R. Agrawal, R. Srikant, "Privacy-preserving data mining", ACM SIGMOD, pp.439-450, 2000.

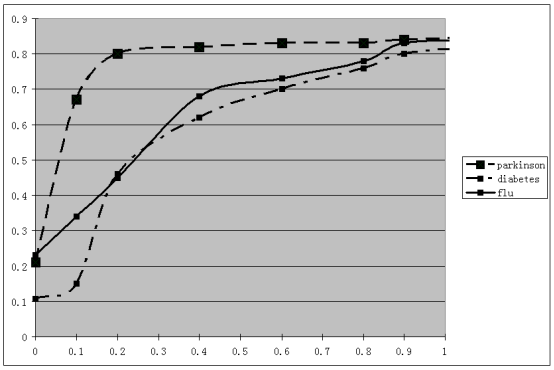

Figure 11: The empirical relationship between the classifier performance (i.e., precision ρ) and the privacy disclosure.

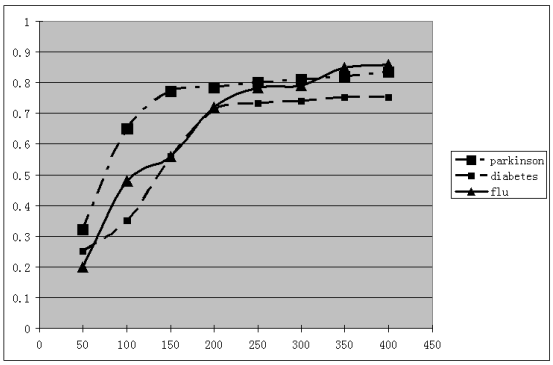

Figure 12: The empirical relationship between the privacy disclosure and the number of blurred test samples to be shared.

[6] D. Agrawal, C. Aggarwal, "On the design and quantification of privacy preserving data mining algorithms", ACM PODS, 2001.

[7] S. Merugu, J. Ghosh, "Privacy-preserving distributed clustering using generative models", IEEE ICDM, 2003.

[8] C.J. Adcock, "Sample size determination: A review", *The Statistician*, vol.46, no.2, pp.261-283, 1997.

[9] M. Boyle, C. Edwards, S. Greenberg, "The effects of filtered video on awareness and privacy", ACM CSCW, 2000.

[10] A. Adams, "Users' perception of privacy in multimedia communication", ACM CHI, 1999.

[11] S. Patil, A. Kobsa, "The challenges in preserving privacy in awareness systems", TR, UC Davis, 2003.

[12] K. Nigam, A. McCallum, S. Thrun, T. Mitchell, "Text classification from labeled and unlabeled documents using EM", *Machine Learning*, vol.39, no.2, 2000.

[13] M. Szummer and T. Jaakkola, "Information Regularization with Partially Labeled Data", Proc. NIPS, 2002.

[14] G. McLachlan and T. Krishnan, *The EM algorithm and extensions*, New York, John Wiley & Sons, 2000.

[15] M. Figueiredo and A.K. Jain, "Unsupervised learning of finite mixture models", *IEEE Trans. PAMI*, vol.24, pp.381-396, 2002.

[16] H. Luo, J. Fan, Y. Gao, G. Xu, "Multimodal Salient Objects: General Building Blocks of Semantic Video Concepts", CIVR, 2004.

[17] E. Bertino, B. Ooi, Y. Yang, R. Deng, "Privacy and ownership preserving of outsourced medical data", ICDE, 521-532, 2005.

[18] Y. Gao, J. Fan, H. Luo, G. Xu, "Salient Objects: Semantic Building Blocks for Image Concept Interpretation", CIVR, 2004.