Person De-Identification in Videos

Prachi Agrawal and P. J. Narayanan

***Abstract*—Advances in cameras and web technology have made it easy to capture and share large amounts of video data over to a large number of people. A large number of cameras oversee public and semi-public spaces today. These raise concerns on the unintentional and unwarranted invasion of the privacy of individuals caught in the videos. To address these concerns, automated methods to *de-identify* individuals in these videos are necessary. De-identification does not aim at destroying all information involving the individuals. Its ideal goals are to obscure the identity of the actor without obscuring the action. This paper outlines the scenarios in which de-identification is required and the issues brought out by those. We also present an approach to de-identify individuals from videos. Our approach involves tracking and segmenting individuals in a conservative voxel space involving x, y, and time. A de-identification transformation is applied per frame using these voxels to obscure the identity. Face, silhouette, gait, and other characteristics need to be obscured, ideally. We show results of our scheme on a number of videos and for several variations of the transformations. We present the results of applying algorithmic identification on the transformed videos. We also present the results of a user-study to evaluate how well humans can identify individuals from the transformed videos.**

***Index Terms*—Biometrics, blurring, de-identification, identification of persons, video surveillance.**

I. Introduction

ADVANCES in cameras and web technology have made it easy to capture and share large amounts of video data over the internet. This has raised new concerns regarding the privacy of individuals. For example, when photographs of a monument are taken to create a panoramic view of the scene, people present are not aware of it and their consent is not taken before making them public. Technologies like Google Street View, EveryScape, Mapjack, and so on have a high chance of invading into one's private life without meaning to do so. Parents have also expressed concern on the possible compromise of the security of their children. The recent furore over Street View in Japan and the U.K. underscores the need to address the privacy issue directly. An increasing number of video cameras observe public spaces like airports, train stations, shops, and streets. While there may be a possible security need to see the individuals in them, identifying the action suffices in most cases. The actor needs to be identified only rarely and only to authorized personnel.

Manuscript received February 23, 2010; revised May 13, 2010; accepted July 6, 2010. Date of publication January 13, 2011; date of current version March 23, 2011. This work was supported in part by the Naval Research Board of India. This paper was recommended by Associate Editor Y. Rui.

The authors are with the Center for Visual Information Technology, International Institute of Information Technology, Hyderabad 500032, India (e-mail: prachi@research.iiit.ac.in; pjn@iiit.ac.in).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2011.2105551

There is, thus, a need to *de-identify* individuals from such videos. De-identification is a process which aims to remove all identification information of the person from an image or video, while maintaining as much information on the action and its context. Recognition and de-identification are opposites with the former making use of all possible features to identify and the latter trying to obfuscate the features to thwart recognition. De-identification should be resistant to recognition by humans and algorithms. Identifying information captured on video can include face, silhouette, posture, gait, and so on.

The privacy issues are genuine and will grow with wider adaptation of video technology. Automated methods to de-identify individuals without affecting the context of the action in the video are needed to address them. Face de-identification in images has been attempted before. However, in a recent work, Kumar *et al.* [1] concluded with the help of a study that context and background help humans recognize faces. Human subjects had 99.20% recognition accuracy on original portrait images. The accuracy dropped to 97.53% when only the tightly cropped face was shown. However, 94.27% accuracy was obtained on original images with blacked out face region. This paper showed that context, background, hair, and so on, help humans in recognition. Videos present more challenges as they capture more information (silhouette, gait, and other aspects) which can also be used for identification compared to images.

In this paper, we present the problem of de-identification of individuals in videos. We first present a general framework under which the requirements of de-identification can be studied. We also present an algorithm for de-identification and study its impact for various combinations of the steps and parameters involved. The performance of our method is studied for identification by algorithms as well as by humans in user studies involving over 100 individuals. The framework to analyze de-identification in videos is a contribution of the paper along with the specific algorithms presented. Preliminary results on the problem appeared in an earlier paper [2].

Section II outlines the different scenarios where de-identification is a requirement, and various aspects related to it. Previous work on de-identification is given in Section III. Section IV explains our method and Section V presents experimental results and a detailed user-study. The concluding remarks are in Section VI.

II. De-Identification: General Framework

De-identification involves the detection and a transformation of images or videos of individuals to make them unrecognizable, without compromising on the action and other contextual

content. It is easy to hide the identity of individuals by replacing a conservative area around them by, say, black pixels. However, this hides most information on what sort of human activity is going on in that space, which may be important for various studies. The goal is to protect the privacy of the individuals while providing sufficient feel for the human activities in the space being imaged. There is a natural trade-off between protecting privacy and providing sufficient detail. Privacy protection provided should be immune to recognition using computer vision as well as using human vision.

A. Different Scenarios and De-Identification

Three types of videos need de-identification to not compromise the privacy of individuals. *Casual videos* are captured for other purposes and get shared. Examples include images used by projects like Google StreetView, the net-cameras fitted in public spaces that can be viewed over the internet, videos or photos on sharing sites, and so on. Individuals appear in these videos purely unintentionally and there is no need to know their identities. All individuals should therefore be de-identified irrevocably and early, perhaps at the camera itself. *Public surveillance videos* come from cameras watching spaces such as airports, streets, stores, and so on. There is no intention to capture any specific set of persons, but there is an explicit intention to capture people occupying the space. These videos may be viewed at a monitoring station to look for anomalies and to judge how users react to situations or products. These may be displayed on public monitors and a recorded version may be accessible to many people. The types of actions performed by individuals in these videos may be important, but not their identities. Hence de-identification is necessary. *Private surveillance videos* come from cameras placed at the entrances of semi-private spaces like offices. Individuals entering them have a purpose and access is often limited to authorized persons only. The videos may be of higher quality and are likely to have a more detailed view of the individuals. De-identification may not be essential, but could be recommended to take care of potential viewing by non-authorized people.

B. Criteria for De-Identification

The characteristics or features used to recognize humans in videos is the focus of a de-identification transformation. These include the following.

1) Face plays a dominant role in automatic and manual identification. Thus, the de-identification transformation should pay more attention to detect and obfuscate faces in the video more than other aspects.
2) The body silhouette and the gait are important clues available in videos which need to be obfuscated. Humans exploit them effectively and algorithmic identification schemes using them have been developed with some success [3], [4]. While obfuscating the silhouette with a small loss of detail is easy, gait is hard to hide. The silhouette should be dilated or expanded to remove its information content. A tight segmentation of the individuals may preserve the silhouette. Gait relates to the temporal variation of a person's arms and silhouette. Masking it needs the temporal silhouettes to be changed in a non-predictable way.
3) Other information about individuals may be critical to specific aspects of privacy, such as the race and gender. Both are hard to mask completely. Though race may relate closely to skin color and can be masked by RGB or hue-space transformations, they destroy the naturalness of the videos in our experience. Gender is more subtle and no clearly defined manifestation has been agreed on, which makes obfuscation of gender hard. We do not address gender or race hiding in this paper, though they may pose critical privacy issues.

C. Subverting De-Identification

We now discuss ways by which the de-identification can be subverted or "attacked" to reveal the identity of individuals involved. The de-identification process has to be satisfactorily robust to these methods.

1) Reversing the de-identification transformation is the most obvious line of attack. The transformation should, thus, be irreversible. We use a blurring involving several neighboring voxels in space and time to prevent direct reversal. However, an indirect reversal approach is to estimate the blurring function from the de-identified frames and then get the original frames back using reconstruction and comparison. Frames of the de-identified video may also be treated as multiple low-resolution observations when a form of blurring is used. Techniques similar to those used in super-resolution may facilitate the reversal of the blurring partially or completely. However, these techniques assume that the point spread function (PSF) of the camera (or the blurring function) which results in the low resolution image is the same at every pixel of the image. An intuitive method to prevent this kind of attack is to make the blurring function random which will make the estimation impossible. The key is to randomize the function in such a way that does not adversely affect the image quality and smoothness. This is achieved as discussed in detail in Section IV.
2) Recognizing persons from face, silhouette, gait, and so on, is being pursued actively in computer vision. The problem may be set as a series of verification problems, given a list of people. The de-identification transformation has to be robust to the common computer vision algorithms. We conducted experiments to validate our system's robustness against some common computer vision algorithms, namely, face detection and person detection. Since identification requires more intricate feature information than detection, we show failure of state-of-the-art detection algorithms as proof of robustness of our algorithm against computer vision algorithms. The results are shown in Section V.
3) Manual identification is another way to subvert de-identification, though it is considerably more expensive. It is not clearly known what properties or features humans use to identify and recognize individuals. However, general blurring and color manipulation makes

recognition highly unlikely even by humans. User study is an effective way to judge the effectiveness of the de-identification approach and to compare between multiple approaches. A detailed user study was conducted to evaluate the effectiveness of our algorithms in different typical scenarios. An effective de-identification algorithm should preserve the action in a video, while removing the identity of the actor completely. Another purpose of the user study was to weigh the importance of gait as a feature for recognition. The results are in accordance with our expectations and are presented in Section V.

4) Brute-force verification is a way to attack a de-identified video. Such attacks are possible if some knowledge of the de-identification algorithm and its parameters are available. Different combinations of algorithms and their parameters can be applied on target individuals, with an automated or manual comparison performed in the de-identified space. A match in the transformed space can strongly indicate a match in the original space. This way of attack cannot be prevented easily; they can only be made arbitrarily hard by the underlying combinatorics. The evaluation of the effectiveness of these methods was left out as it is out of the scope of this paper.

It should be noted that only transformations that ignore the input video can theoretically be totally safe. Brute-force attack is possible on others. Such a transformation will replace individuals in the video with a constant (say, black or white) or random color. We rule out such methods as they destroy all information on the action performed.

D. Storage of Videos

The de-identification process should support untransformed video to be viewed if the situation demands. This is essential to support the primary purpose of watching the space, whether for security or information. That is, the de-identification should be selectively reversed when needed. It is important that individuals do not appear in the clear at any time in the video otherwise. The safest approach is to de-identify the video at the capture-camera. Only the transformed video is transmitted or recorded. Clear video can be viewed only by reversing the transformation. This requires the de-identification to be reversible, which poses some risk of being attacked. The parameters needed for reversing the transformation should be saved along with the video using sufficiently strong encryption. Another approach is to store the original video, with sufficiently hard encryption, along with the de-identified video. The required keys for decryption are available only with authorized persons. This needs additional storage space, which can be reduced by saving only the portions that contain humans. Another relevant issue is the computational cost of de-identification. Videos are bulky and their transformation requires serious computing power. We do not address the computational issues in this paper, though they are important.

III. Related Work

In the past, outlines of privacy preserving systems have been presented to highlight the underlying issues [5], [6]. These were only sketches and not reports of an implemented de-identification system. Most implementations of privacy protection schemes focus on faces [7]–[10]. However, face is only one out of a long list of identifiable features of an individual: body structure, silhouette, gait, gender, race, and so on also aid recognition and hence should be masked adequately. Although face de-identification is not enough when it comes to providing privacy (especially in a video), the motivation behind all these schemes was similar to ours: protecting privacy of an individual. Hence, we provide a brief description of the privacy protection schemes implemented in the past.

Commonly used face de-identification schemes rely on methods that work well against human vision such as pixelation and blurring. More recent methods such as the k-Same [7] and k-Same-Select [8] implement the k-anonymity protection model which provides provable privacy and preserve data utility. Gross *et al.* [8], [9] combined a model-based face image parametrization with a formal privacy protection model. They also proposed a semi-supervised learning based approach for multi-factor models for face de-identification. Phillips [10] proposed an algorithm for privacy protection through the reduction of the number of eigenvectors used in reconstructing images from basis vectors. In other work [11], the need for automatic techniques for protecting the privacy of people captured in images by Google Street View was recognized and addressed by a method to obscure the faces and number plates of cars in these images. However, the primary focus of this paper is on handling large scale data and reducing the number of false positives in order to maintain the visual quality of images, while keeping recall as high as possible.

Face modification has also been attempted as a way of image manipulation [12]–[14]. Bitouk *et al.* [14] replaced faces from one image into another, by aligning the faces in the two images automatically to a common coordinate system. Blanz *et al.* [13] estimated the shape, pose, and direction of illumination in the target and source faces, and fit a morphable 3-D model to each face optimizing all the parameters. They rendered the new face by transferring the scene parameters of the target image to the source 3-D model. However, face modification is different from de-identification.

There has been little work in the past dealing with entire human body for de-identification. Chen *et al.* [15] presented a system to protect the privacy of pre-specified individuals in a video taken in a hospital. They used an automatic people identification system that learned from limited labeled data. They also proposed a method for human body obscuring using motion history information of the edges. This method hides the identity of the actor, but it also removes all the information on the action. Park *et al.* [16] introduced the concept of personal boundary and incorporated it in a context adaptive human movement analysis system. Foreground pixels are divided into coherent blobs based on color similarity. Multiple blobs constitute a human body and are tracked across the frames. These blobs are used to block human identity. The problem with this approach is that it preserves the overall silhouette of the person which can aid recognition.

Another technique used for protecting privacy is based on segmenting the privacy information from a video and

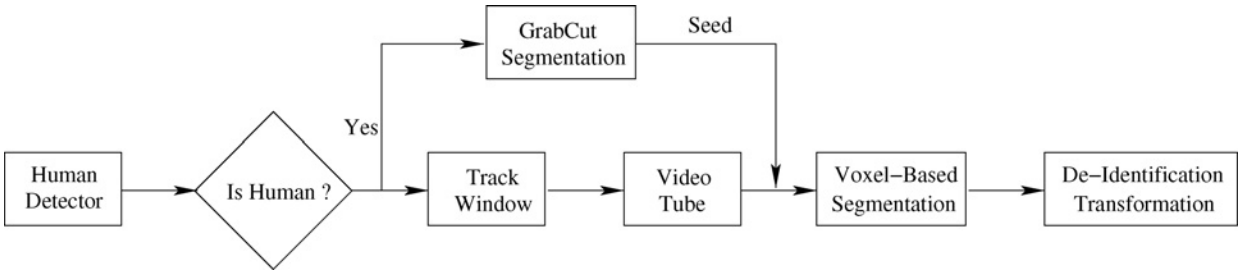


Fig. 1. Overview of the method.

encrypting the information to hide it from the end user. Different frameworks have been proposed to hide the data in the video itself, e.g., as a watermark [17] or as encrypted information in DCT blocks [18]. This information can be retrieved later on request. These schemes work on the entire human body instead of just faces, but they also remove all the information content related to action in the video. They could benefit from a framework that provides a variable amount of control to the users over the information viewed in a video [17]. Neustaedter *et al.* [19] also suggested a prototype design of a smart system which learns from the visual feedback it receives and provides a varying level of privacy based on the feedback.

There have been studies in the past to evaluate the performance of simple privacy protection techniques for day-to-day home-office situations and people's perception of de-identification in general. Boyle *et al.* [20] showed that blur filtration balances privacy and awareness for day-to-day office situations. Neustaedter *et al.* [19] showed that blur filtration is insufficient to provide an adequate level of privacy for risky home situations. They concluded from a survey that people will be suspicious of any system initially but could learn to trust it after a period of usage, like Active Badge System [21].

Detecting and segmenting humans in images and videos is a very active area of research today which may help a complete de-identification system [22], [23]. Recognizing humans from faces, silhouettes, gait, and so on, is also an active area; success in those provides more methods a de-identification system should guard against.

Algorithm 1 Pseudo code: Overview of the method

```
Apply HOG human detector [24].
if Bounding Box (BB) of human overlaps a TrackedWindow then {Same Person}
    Replace old TrackedWindow with BB.
else {New Person}
    Add BB as new TrackedWindow.
    Perform GrabCut [25] with BB as input.
    Build GMMs.
end if
For each new frame, update the existing TrackedWindows after patch-based tracking [26].
Form each person's video tube by stacking their TrackedWindows across time.
Divide the video into fixed 4 × 4 × 2 voxels.
if Voxel planes in a person's video tube = 4 then
    Build a 3-D Graph on voxels of the video tube.
    Perform Graph Cut [27], [28].
    Retain the last voxel plane.
    Apply de-identification transformation on the segmented frames using one of the techniques mentioned in Section IV-C.
    Apply the randomization kernel.
end if
```

IV. De-Identification: Proposed Approach

An overview of our method is outlined in Fig. 1. The system is comprised of three modules: Detect and Track, Segmentation, and De-identification.

A. Detect and Track

The first step is to detect the presence of a person in the scene. HOG based human detector gives good results with a low miss rate [24]. Other human detectors may also be employed [29], [30]. A robust tracking algorithm is required, as any error in tracking will increase the chances of recognition. We use a patch-based recognition approach for object tracking [26]. The object is divided into multiple spatial patches or fragments, each of which is tracked in the next frame by a voting mechanism based on the histogram of the corresponding image patch. The voting score for different positions and scales from multiple patches is minimized in a robust manner to combine the vote maps and select the most appropriate candidate. This approach is robust to partial occlusions and pose variations. It also takes into account the relative spatial distributions of the pixels, unlike traditional histogram-based tracking methods [31], [32].

Although the algorithm allows for voting on different scales of the object, to avoid errors resulting from partial occlusions and fast changing scale, we apply the human detector every F frames. The output of the human detector becomes the input to the tracking module. The value of F depends on the amount of movement in the video. If the scale of the human doesn't change much over the course of the video, then a high value of F can be chosen. If the scale changes every few frames, then F is small. We set the value of F to 40 for our experiments.

B. Segmentation

The bounding boxes of the human in every frame, provided by the tracking module, are stacked across time to generate

a *video tube* of the person. Multiple video tubes are formed if there are multiple people in the video. Segmentation of the person is performed on the video tube as follows. The video space is first divided into fixed voxels of size $(x \times y \times t)$ in the spatial (x, y) and temporal (t) domains. This reduces the computation required in the large video space. Also, a block-based segmentation removes fine silhouette information while preserving gross outlines. Fine boundaries of a person reveal a lot about the body shape and gait, and can aid recognition [3], [4]. The values of x and y are typically set to 4 each and t can be anything between 2 and 10, depending on the degree of movement in the frames.

Segmentation assigns each voxel v a label, 1 for foreground and 0 for background. For this, the video tube is divided into blocks of B voxel-planes in time. A voxel-plane is a collection of voxels obtained by combining $[F_n, F_{n+1}, \cdots F_{n+t-1}]$ frames in the video space, where t is the size of each voxel in the temporal domain. The voxels are treated as superpixels and a 3-D graph is constructed per block, where each node corresponds to a voxel [33]. One voxel-plane overlap is used between consecutive blocks to enforce continuity across the blocks. B must be small (between 3 and 10) for good results, but not too small, as it would make the overall computation time high.

The energy term E associated with the graph is of the form

$$E(\underline{\alpha}, \underline{\theta}, \underline{v}) = U(\underline{\alpha}, \underline{\theta}, \underline{v}) + \lambda_1 V_1(\underline{v}) + \lambda_2 V_2(\underline{v}) \tag{1}$$

where U is the data term and V_1, V_2 are the smoothness terms corresponding to the intra-frame and inter-frame connections between two voxels, respectively. The Gaussian mixture models (GMMs) are used for adequately modeling data points in the color space [34]. $\underline{\theta} = \{\theta^0, \theta^1\}$ are two full-covariance Gaussian color mixtures, one each for foreground and background, with K clusters each. Hence, $k \in [1, K]$, $\alpha = \{0, 1\}$ and $\theta^\alpha = \{w_k^\alpha, \mu_k^\alpha, \Sigma_k^\alpha\}$. We used $K = 6$ for the results presented here. These GMMs provide seeds to the graph, as well as help in defining the energy terms. The energy E is defined such that a minimization of it provides a segmentation that is coherent across time and space.

The data term U, similar to the one used by GrabCut [25] is defined as $U(\underline{\alpha}, \underline{\theta}, \underline{v}) = \sum_n D(\alpha_n, \theta_k, v_n)$ where n is the number of voxels and

$$D(\alpha_n, \theta_k, v_n) = \min_{k=1\cdots K} [-\log w_k^{\alpha_n} + \frac{1}{2} \log \det \Sigma_k^{\alpha_n} + \frac{1}{2} \bar{v}_n^T \Sigma_k^{\alpha_n -1} \bar{v}_n] \tag{2}$$

where $\bar{v}_n = v_n - \mu_k^{\alpha_n}$. The representative color v_n for a voxel should be chosen carefully. The average color of a voxel is not a good representative as we initialize the GMMs based on pixel colors. The average color, which is a mixture of several colors, might not lie close to any GMM, despite being a foreground or a background pixel. The problem is intensified in the case of boundary voxels, where the average color would be a mixture of the foreground and background colors. Our solution is biased toward segmenting more voxels as foreground than background, which would be difficult in case of average color. To this end, we first compute the distance D_0 and D_1 to the background and foreground respectively for each pixel in a voxel, using pixel color instead of v_n in (2). The pixels are sorted on the ratio $\frac{D_0}{D_1}$ in the decreasing order. We choose the color of mth pixel after sorting as the representative color v_n. The value of m is kept low so that voxels with even a few foreground pixels are biased toward the foreground. This is important for de-identification as the foreground needs to be segmented conservatively. We also identify seed voxels for the graphcut segmentation based on D_0 and D_1. If the distance to foreground, D_1, is very low for the mth pixel, the voxel is a seed foreground. However, if the distance to background, D_0, is very low for the $(N - m)$th pixel (where N is the number of pixels in the voxel), the voxel is a seed background.

The smoothness terms V_1 and V_2 are also similar to the ones used in GrabCut, defined as $V(\underline{v}) = \sum_{v_p, v_q \in \underline{v}} \delta_{pq} \cdot V_{pq}$, where δ_{pq} is 1 when v_p and v_q are neighbors and 0 otherwise, and

$$V_{pq} = \exp^{-\beta \|v_p - v_q\|^2} \tag{3}$$

where v_p is the mean color of a voxel. β is the expected value calculated as $\beta = (2\mathcal{E}(\|v_p - v_q\|^2))^{-1}$, where $\mathcal{E}$ is the expectation operator [25].

A mincut on the above graph minimizes the energy E efficiently [27], [28]. A rigid but blocky (because of voxelation) outline of the human is obtained after segmentation. Initialization of foreground and background seeds is done by performing GrabCut [25] on the first frame that contains the human. The foreground and background GMMs are also initialized in this process.

C. De-Identification

After the segmentation of the person, the de-identification transformation is applied on the human being present. We explore two de-identification transformations: 1) exponential blur of pixels of the voxel, and 2) line integral convolution (LIC). We explore these transformations in isolation as well as in different combinations, and evaluate the performance of each of these.

In exponential blur, all neighboring voxels of a foreground voxel within the distance a participate in de-identification. The parameter a controls the amount of de-identification; more the value of a, more is the de-identification. Typically a lies between 1 and 5. The output color for each pixel in a foreground voxel is a weighted combination of its neighboring voxels' average colors. Each voxel is weighted based on the pixel's distance from the center of that voxel. If v_i is a foreground voxel and v_p is its neighboring voxel, the weights corresponding to the (l, m, n)th pixel of v_i can be calculated $\forall v_p \in \Gamma_i$ as

$$\gamma(l, m, n) = e^{-\frac{d^2_{(l,m,n),v_p}}{8a^2}} \tag{4}$$

where Γ_i is the set of voxels which lie within distance a from v_i, and $d_{(l,m,n),v_p}$ is the distance of the (l, m, n)th pixel of v_i from the voxel center v_p.

The weights γ have certain inherent properties. The distance $d_{(l,m,n),v_p}$ depends only on l, m, n and the relative position of v_p with respect to the current voxel. Hence, once the value of a is fixed, the weight vector (of size Na^3) is fixed. Because this weight vector is the same for every voxel, it can

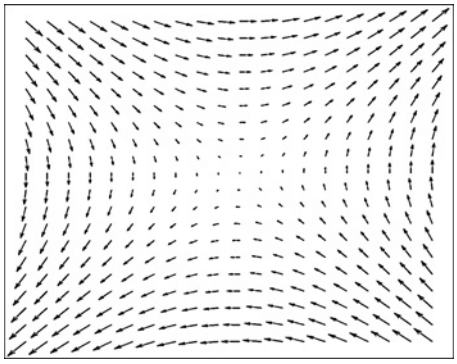


Fig. 2. Saddle shaped vector field used for LIC.

be pre-computed once and used for every voxel. Moreover, the distance $d_{(l,m,n),v_p}$, and hence the weight vector $\gamma(l, m, n)$, vary smoothly within a voxel and across two voxels. This prevents abrupt changes in color at voxel boundaries. Also, because the weight corresponding to a distant voxel is low compared to a nearby voxel, the voxels at distance a will have less contribution to a pixel's color. Hence, the color of pixels on the either side of voxel boundaries changes smoothly, as only voxels at distance a are added or removed from their active neighborhood. This kind of smooth temporal blurring of the space-time boundaries aims to remove any gait information of the individual.

The second de-identification transformation is based on LIC. LIC is used for imaging vector fields [35] on a texture. A long and narrow filter kernel is generated for each vector in the field whose direction is tangential to that of the vector and length is $2L$. L lies typically between 2 and 20. The bounding box around the human is mapped one-to-one onto the vector field. The pixels within the bounding box and under the filter kernel are summed, normalized and placed in an output pixel image for the corresponding position. This process is repeated for all foreground pixels obtained after segmentation. LIC distorts the boundaries of the person which tends to obfuscate silhouettes. Different vector fields can be used for achieving different effects. We used a saddle shaped vector field (Fig. 2) for our experiments. The amount of de-identification is controlled by the line length parameter, L, of the convolution filter.

When used in isolation, blur is more effective to hide gait and facial features, while LIC distorts the silhouettes more. Hence, we tried a combination of these two transformations where we perform LIC on the voxels followed by a voxel based exponential blur. To make identification based on the color of face and clothes difficult, intensity space compression (ISC) was additionally tried as a subsequent step. The intensity values of the foreground pixels are compressed after an exponential blur or LIC. The result is boosted up by a fixed value after the compression. It provides greater de-identification, but the video loses more context information. The results are presented in Figs. 3 and 4.

D. Randomization

Since the super-resolution techniques assume the PSF to be same at every pixel, the easiest way to thwart a reversal attack using them is to randomize the blurring function at every pixel. This trivial adjustment makes estimation of the blurring function impossible, and hence direct comparison based reconstruction techniques will not work. Instead of making the whole blurring function random at every pixel which would result in non-smooth, low quality and blocky images, we make use of a separate randomization layer as the final step. This is achieved by using a blurring kernel (one out of a fixed pool of N kernels), chosen randomly for every pixel. The pool contains low pass filters of frequencies and construction slightly different from each other. This blurring is thus sufficiently random, but not so much to introduce sharp lines in the output image. Similar effect could be achieved by adding a small random value to the blurring weight corresponding to each pixel in the previous step. However, the resulting kernel will not be consistent with the notion of an ideal blurring kernel where the weights fall off consistently with respect to distance, and might introduce discontinuities around the boundaries of two voxels.

V. Experimental Results

We implemented the above system and conducted the experiments on standard data sets like CAVIAR, BEHAVE, and so on, and on our own that provide more clearly visible individuals in videos. We divide the video into $N = 4 \times 4 \times 2$ sized voxels. The parameter m which decides the representative color v_n of a voxel used in defining the data term in (2) was kept as 3 (10% of N) for our experiments. Increasing the voxel size across time domain increases the blockiness across the frames. If a person is moving fast enough, it can introduce jumps in the segmented output around the boundary. Different parameters were tried for each of the de-identification transformations; $a = 2$ and 4 for exponential blur, $L = 10$ and 20 for LIC on pixels, and $vL = 2$ and 5 for LIC on voxels. $L = 20$ in pixel space is equivalent to $vL = 5$ in voxel space as 5 voxels cover 20 pixels in one dimension. Similar comparisons can be made between $L = 10$ and $vL = 2$.

Our implementation is not real time currently. It takes about 10 to 12 s on an average to completely process and obtain results on a block of size 4 voxel planes on an Intel 2.4 GHz processor with 1 GB RAM. The tracking module takes about 8–10% of the running time. Graph cut in itself takes only about 2–3% of the total time to run. The de-identification and randomization modules together take over 12% of the time. The rest of the time is spent in voxelizing the video, calculating the energy functions for t-edges and n-edges of the graph, and so on. The inherent parallelism of many of these modules may be explored for a real-time implementation on the GPU. However, we do not address the real-time issue in this paper.

Visual results in selected frames are shown in Figs. 3–5. Fig. 3 shows the output of different de-identification transformations on a single frame from different videos. Increasing the value of a and L increases the de-identification achieved, but it results in more loss of information in a scene. In general, Blur-4 and LIC-20 perform better than Blur-2 and LIC-10 in masking the identity of people. However the output of LIC-20 sometimes looks unnatural and ghost-like. The combination of LIC and Blur works better than either by itself; the

Fig. 3. First column shows the clear frame. The next five columns show the output of Blur-2, Blur-4, LIC-10, LIC-20, and Blur-2 followed by an intensity space compression, in that order.

user-study conducted on the videos conforms with the statement and is discussed in the next section. The effect of changing the parameters of the transformations can be seen in the figures. The ISC, as shown in Figs. 3 and 4, can remove color dominated information such as race, but can accentuate the body structure of the person. Fig. 5 shows frames of de-identified videos in which people are performing different activities. As can be seen, the activity is recognizable but the person is not, which is the underlying goal of de-identification. More results can be seen at http://cvit.iiit.ac.in/projects/de-id/index.html.

A. Algorithmic Evaluation

To gauge the robustness of our system against algorithmic recognition techniques, we tested the de-identified videos on a standard face detector and a human detector, which are used as the first step by most recognition algorithms. We used OpenCV's implementation of the Viola-Jones face detection algorithm [36] for face detection and the HOG based human detector for person detection [24]. On a total of 24 de-identified videos, and 6110 frames in which a person was present, the face detector resulted in 0.2% hits and the human detector resulted in 56.2% hits, on an average. Table I summarizes the output for different transformation combinations. An increase in the de-identification transformation parameter reduces the number of hits, as expected.

However, when the detectors were tested on clear videos, we get 97.2% and 7.8% hit rates in the case of person detector and face detector, respectively.[1] A fall in the hit rate in de-identified videos, especially of the face detector, can be taken as a confirmation that our system is robust against recognition algorithms, as the fine details which are the requirement of any recognition algorithm are removed from the videos. The person detector worked in more than half the cases on an average, which is acceptable, as it only indicates that a human is present in the video.

[1]The videos contained people at a large distance from the camera, as in surveillance, and frontal and profile faces. All these explain the low hit rate in the case of face detector.

TABLE I
PERCENTAGE OF CORRECT ANSWERS FOR THE FACE AND HUMAN DETECTORS

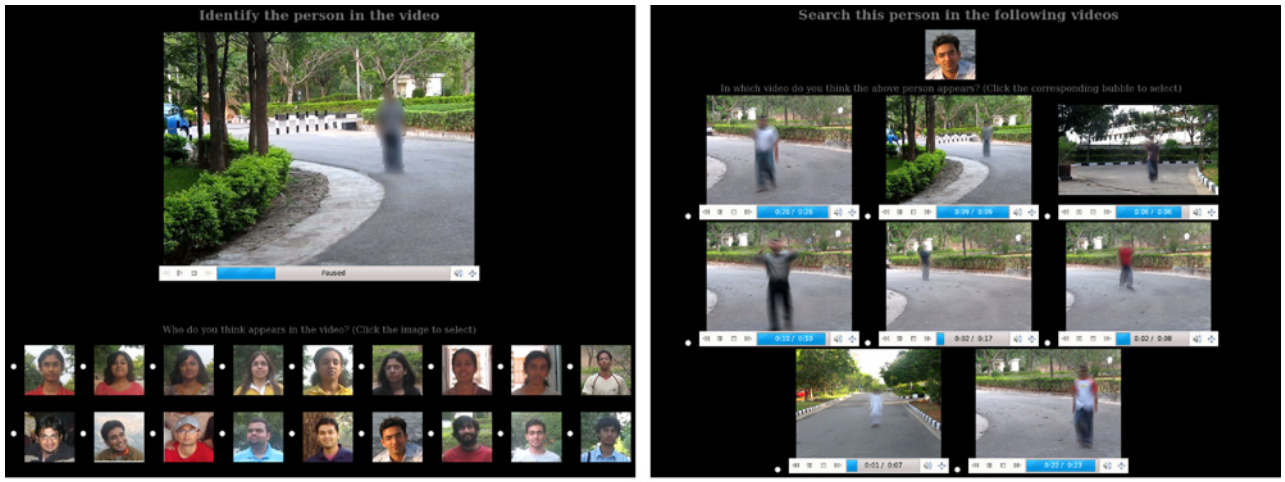
Algorithm, Parameter	Percentage of Success	
	Human Detection	Face Detection
Blur, $a = 2$	89.7	1.0
Blur, $a = 4$	56.1	0
LIC, $L = 10$	73.6	0.3
LIC, $L = 20$	22.4	0
$vL = 2, a = 2$	64.4	0
$vL = 2, a = 4$	59.3	0
$vL = 5, a = 2$	44.9	0
$vL = 5, a = 4$	38.9	0

B. User Study

Recognition by humans is one of the ways to subvert de-identification. It is difficult to quantitatively state the effectiveness of the system as it is not known which features humans use to identify and recognize individuals. Hence, two user studies were conducted to test the usefulness of the system. The videos used in these studies were our own, taken in different plausible settings, featuring students of our institute. Preliminary results of an early user study on standard data sets like CAVIAR, BEHAVE, and so onrevealed that these data sets are not challenging enough for studies on de-identification. In BEHAVE, the scene is captured by a distant camera placed high up looking down the road. The actors' faces are very small and not recognizable. In CAVIAR, the faces are recognizable when the actors came close to the camera. However, for unfamiliar people, the only cue available

Fig. 4. Results on two different videos. The clear frames are shown in the odd rows while corresponding de-identified frames in the even rows.

Fig. 5. De-identified frames showing people performing different activities; the activity is recognizable but the person is not.



(a) (b)

Fig. 6. Screenshot of the portal used for the user study for (a) identification and (b) search.

TABLE II
NUMBER OF CORRECT IDENTIFICATIONS FOR *Search* AND *Identification* EXPERIMENTS IN THE USER STUDY

	From Images		From Videos		Activity
Algorithm, Parameter	Identification	Search	Identification	Search	Recognition
Blur, $a = 2$	4(B)	4(A)	4(B)	4(A)	7(B)
Blur, $a = 4$	1(D)	3(C)	0(D)	1(C)	6(D)
LIC, $L = 10$	3(F)	2(E)	4(F)	2(E)	7(F)
LIC, $L = 20$	1(H)	3(G)	1(H)	3(G)	7(H)
$vL = 2, a = 2$	3(C)	3(D)	2(C)	3(D)	11(C)
$vL = 2, a = 4$	0(G)	1(H)	1(G)	4(H)	8(G)
$vL = 5, a = 2$	2(A)	0(B)	2(A)	2(B)	8(A)
$vL = 5, a = 4$	2(E)	0(F)	3(E)	4(F)	7(E)

Sets A to H had 9, 8, 11, 9, 9, 8, 10, and 10 users, respectively.

to identify these actors (even in most clear videos) is the color of their clothes. Since we did not have access to an image of these actors other than that in the video itself, the user study on these de-identified videos necessarily meant matching the color of clothes in the candidate images and videos. Our experiments also showed that users were employing the clothing information only for identification. Hence, there was a need to create our own data set for the evaluation of our method, in which faces are clearly visible for even unfamiliar people to recognize. We could also take different images and videos of our actors in different clothing and scenarios to conduct a detailed user study as explained below.

The individuals in our data set were asked to perform actions like waving hand (as a gesture to greet), talking on the phone, turning head left or right, carrying a bag, eating or drinking, and so on. The user study gauged identification of the person and the action performed. Clear videos were also used as examples to enable the learning of gait, silhouette, and other aspects. A demo of our system was put up for evaluation at a technical event at our institute, which was attended by several hundred visitors from outside. The study was conducted on 74 such visitors. The subjects were completely unfamiliar with the individuals appearing in the videos. As shown in Table V-A, eight different parameter combinations of the de-identification transformations were included in the study. The study consisted of eight sets of six videos each. Half the videos in each set was de-identified using one combination and was used for the *identification* experiment. The other half was de-identified using another combination and was used for the *search* experiment. In this manner, all the eight parameter combinations were covered for both the experiments (identification and search) in these eight sets. The people taking the study were also divided into eight sets (named A to H), and each user took the study on identification and search in one set. The users were shown a randomly chosen video for identification and another for search. This was to ensure that the outcome of the experiment is not affected by the type of videos used for the purpose.

For the identification experiment, the users were asked to match the individual in the de-identified video against a pool of 20 candidate photographs (face and upper body only) shown in clear (Fig. 6(a)). They were also asked to select the action in the video. Next, the users were shown clear videos of those 20 candidates from which they could learn their walking style, posture, and so on. These videos were taken in a different setting and in different clothes than the de-identified videos to ensure there is no unnecessary learning from the background, context, and so on. The users could go back and change their

previous answer. Similarly, for the search experiment, the users were asked to search for an individual in a pool of eight de-identified videos (Fig. 6(b)). They were first shown a clear image of the person and were asked to find him/her. Then they were shown a clear video of the same person, and were given an option to go back and change their answer. All their answers were recorded and are summarized in Table II.

The numbers in Table II represent the correct identifications and searches by the users. The alphabet in the parentheses represents the set of people who took that particular experiment. The first column of the table represents the different algorithms and their parameters used. The next two columns are for different tests (identification and search from images, and then from videos). The last column shows the number of times the activity was correctly recognized by the users for a particular parameter.

The study can be divided into three categories for the sake of analysis. One category deals with the effect of a certain de-identification algorithm and parameter on the recognition ability of the users. Another category compares the improvement in performance of the users due to learning the gait and silhouette, in identification and search. The third category analyzes the ability of the users to recognize the activity for different parameters. The user study results are mostly as expected. Individuals with very special walking styles or body structures had much better recognition. The users could recognize the activity in the de-identified video in most cases for all parameters, at an average of about 80%. The impact of parameters is also as expected. The trade-off between privacy and context in the de-identified videos is apparent from the results. As the parameter controlling the amount of de-identification increases, the percentage of correct answers decreases. This necessarily means that as the actors became less identifiable, the video started losing the context and detail, as expected. This is almost always true, except in few cases, as explained later.

There are a few observations to be made from the study.

1) In general, search is easier than identification, as it is easier to learn about one person in search than about all possible candidates in identification. Hence, very few people changed their answers in the case of identification when they were shown clear videos after images to learn the gait. It also makes sense intuitively as verification is easier than identification.
2) A combination of LIC and Blur is better than these transformations in isolation. While this is true in most cases, more users changed their answers (usually to correct ones) when they were shown clear videos for the combinatorial cases. While this might look like an anomaly, it could be because the faces were obscured totally by the combined transformations. Hence, there was an increased reliance on the clear videos of these individuals, which is more pronounced in the case of search than identification for reasons explained earlier.
3) The users fared better in identification from videos than images for a particular de-identification transformation combination. The users spent only about 4–5 min on an average to complete the entire study and may have had only limited have enough time to learn the gait, and so on from the videos. The users did not change their answers when they moved from images to videos in most cases. If they did, they changed their answer to the correct one. In some cases, like identification and search in Blur-4 and identification in $vL = 2, a = 2$, some users who gave correct answers from the images changed their answers when they were shown videos. While part of this anomaly could be attributed to the anxiety of people when they are a part of such user studies, most of it stemmed largely from the fact that the videos which were used for the user study contained two men and two women whose height, build and walking styles were similar. Moreover, Blur-4 hides facial features more than any other transformation, and there was more reliance on videos for recognition. Two cases out of three in which the anomaly occurred are from the same set, which means that that particular set of users were more anxious and confused than others and changed their answers to the wrong ones after seeing the videos.
4) As the parameter controlling the amount of de-identification increases, the percentage of correct answers decreases. However, across different algorithms, LIC–10 is more effective than Blur-2. $vL = 2, a = 4$ is similar to $vL = 5, a = 2$. While users perform better in identification on one case, they perform better on another in search. $vL = 5, a = 2$ and $vL = 5, a = 4$ are also similar in performance, with only major difference being in search from videos. A possible explanation for this anomaly is that the set of users who took this particular experiment (F) were good at recognition, as is also apparent from the high numbers corresponding to the other experiment conducted with the same set, LIC-10. Another anomaly occurs between $vL = 2, a = 4$ and $vL = 5, a = 4$. In the case of $vL = 5, a = 4$ under identification, the percentage of correct answers is more than the corresponding figures in $vL = 2, a = 4$. The anomalies can be attributed to different set sizes, difference in the difficulty of de-identified videos across sets, and randomness which is unavoidable in any user study.

TABLE III
NUMBER OF CORRECT IDENTIFICATIONS IN THE USER STUDY ON FAMILIAR PEOPLE

	Familiar		Casually Familiar	
Algorithm, Parameter	Correct	Incorrect	Correct	Incorrect
Blur, $a = 2$	24	6	11	19
Blur, $a = 4$	21	9	10	20
LIC, $L = 10$	24	6	15	15
LIC, $L = 20$	23	7	13	17

To test the effect of familiarity on recognition ability, another user study was conducted. We showed four different sets of six videos each, processed with a different parameter value in each set, to 40 individuals. Half of them were from the same lab as the individuals appearing in the videos and

TABLE IV
HUMAN EXPERIENCE SCORES ON A SCALE OF 1 (LOW) TO 7 (HIGH)

Algorithm, Parameter	Naturalness
Blur, $a = 2$	5.2
Blur, $a = 4$	3.5
LIC, $L = 10$	3.9
LIC, $L = 20$	2.6
$vL = 2, a = 2$	5.2
$vL = 2, a = 4$	3.9
$vL = 5, a = 2$	3.8
$vL = 5, a = 4$	3.0
ISC	2.2

were quite familiar with them. Others were from different labs and were only casually familiar with these individuals. Users were asked to match the individuals appearing in the video against a palette of 30 photographs shown. They were also asked to state the factor that helped them in the recognition. The results are summarized in Table III. The numbers in the table represent the correct identifications by the users. Overall correct recognition was fairly high due to the familiarity of the users with the subjects. The users rated the gait or the walking style to be a big give-away. For example, individual 4, for whom the highest recognition was reported (about 80%), has a very unique walking style. For individual 2, only about 20% of the answers were correct because this person has no unique body shape or walking style. The correct answers for this person were only from those sets in which low values of parameters for Blur and LIC were used.

Another user study was conducted on 30 people to capture the users' experience of the processed videos. We showed each user nine videos, each processed with a different parameter combination (including ISC). The users were asked to rate each video on a scale of 1–7 to specify how natural (or acceptable) they found a particular parameter, where a score of 1 meant very unnatural and unacceptable while 7 meant completely acceptable. The results are shown in Table IV. All the parameter combinations scored above 3 on an average, while LIC-20 scored 2.6 and ISC scored only 2.2. Blur scored about 4.5 on an average (with answers ranging from 3 to 7), which is slightly better than LIC which scored about 3.5 on an average (with answers ranging from 2 to 6). The average scores of LIC and Blur combinations were between 3 and 6, with scores decreasing as the parameter values were increased. The difference in the naturalness scores of LIC and Blur was not significant enough to affect the choice between these two algorithms.

C. Limitations

Our algorithm consists of many modules and the results are sensitive to proper functioning of all the modules involved. It is necessary for each module to function perfectly for our de-identification to work. Failure to do so in even one frame can jeopardize privacy. Each module has its own limitations. The tracking module misses the extended arms, feet, hair, or even the face sometimes which might compromise privacy. The segmentation module is largely dependent on color, and gives errors when the background is cluttered or the background and foreground have the same color around the segmentation boundary (Fig. 7). The seeds for segmentation and the GMMs depend on the success of GrabCut, which is a very crucial step. Also, a miss by the HOG detector will certainly prove fatal for the de-identification process.

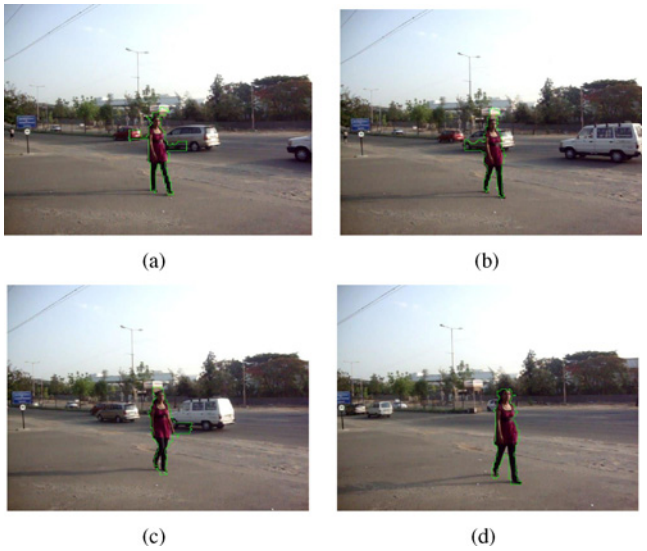


Fig. 7. Segmentation result on a video with dynamic background. (a)–(d) Frames of a sequence with clutter in the background. Results show that a cluttered background could lead to bad segmentation, hence extra blurring around the edges.

D. Discussion

The results suggest that a high level of blurring should be used for effective de-identification. While the facial and other features can be masked adequately, the gait and other temporal characteristics are hard to mask. Our user study confirms that de-identifying an individual to others familiar with him/her is a very challenging task. Without familiarity, gait and other characteristics are of low value and face plays the most important role. The studies also suggest that an action can be recognized with more accuracy than the person performing that action in a de-identified video.

VI. CONCLUSION

In this paper, we analyzed the issues relating to de-identification of individuals in videos to protect their privacy by going beyond face recognition. We also presented a basic system to protect privacy against algorithmic and human recognition. We presented results on a few standard videos as well as videos we collected that are more challenging to hide identity in. We also conducted a user study to evaluate the effectiveness of our system. Our studies indicate that gait and other temporal characteristics are difficult to hide if there is sufficient familiarity with the subjects and the user. Blurring is a good way to hide the identity if gait is not involved. We proposed to conduct further studies to evaluate the de-identification system against recognition by computer vision algorithms. That is likely to be easier than guarding against manual identification of individuals.

REFERENCES

[1] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Attribute and simile classifiers for face verification," in *Proc. IEEE ICCV*, Sep. 2009, pp. 365–372.

[2] P. Agrawal and P. J. Narayanan, "Person de-identification in videos," in *Proc. ACCV*, 2009, pp. 266–276.

[3] R. T. Collins, R. Gross, and J. Shi, "Silhouette-based human identification from body shape and gait," in *Proc. IEEE Conf. Face Gesture Recognit.*, May 2002, pp. 351–356.

[4] J.-H. Yoo, D. Hwang, and M. S. Nixon, "Gender classification in human gait using support vector machine," in *Proc. ACIVS*, 2005, pp. 138–145.

[5] A. W. Senior, "Privacy enablement in a surveillance system," in *Proc. ICIP*, 2008, pp. 1680–1683.

[6] X. Yu, K. Chinomi, T. Koshimizu, N. Nitta, Y. Ito, and N. Babaguchi, "Privacy protecting visual processing for secure video surveillance," in *Proc. ICIP*, 2008, pp. 1672–1675.

[7] E. Newton, L. Sweeney, and B. Malin, "Preserving privacy by de-identifying facial images," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 2, pp. 232–243, Feb. 2005.

[8] R. Gross, E. Airoldi, B. Malin, and L. Sweeney, "Integrating utility into face de-identification," in *Proc. Workshop Privacy Enhancing Technol.*, 2005, pp. 227–242.

[9] R. Gross, L. Sweeney, F. de la Torre, and S. Baker, "Semi-supervised learning of multi-factor models for face de-identification," in *Proc. IEEE Conf. CVPR*, Jun. 2008, pp. 1–8.

[10] P. Phillips, "Privacy operating characteristic for privacy protection in surveillance applications," in *Proc. AVBPA*, 2005, p. 869.

[11] A. Frome, G. Cheung, A. Abdulkader, M. Zennaro, B. Wu, A. Bissacco, H. Adam, H. Neven, and L. Vincent, "Large-scale privacy protection in Google Street View," in *Proc. IEEE ICCV*, Sep.–Oct. 2009, pp. 2373–2380.

[12] A. Agarwala, M. Dontcheva, M. Agrawala, S. M. Drucker, A. Colburn, B. Curless, D. Salesin, and M. F. Cohen, "Interactive digital photomontage," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 294–302, 2004.

[13] V. Blanz, K. Scherbaum, T. Vetter, and H.-P. Seidel, "Exchanging faces in images," *Comput. Graph. Forum*, vol. 23, no. 3, pp. 669–676, 2004.

[14] D. Bitouk, N. Kumar, S. Dhillon, P. Belhumeur, and S. K. Nayar, "Face swapping: Automatically replacing faces in photographs," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 1–8, 2008.

[15] D. Chen, Y. Chang, R. Yan, and J. Yang, "Tools for protecting the privacy of specific individuals in video," *EURASIP J. Adv. Signal Process.*, vol. 2007, no. 1, p. 107, 2007.

[16] S. Park and M. Trivedi, "A track-based human movement analysis and privacy protection system adaptive to environmental contexts," in *Proc. AVSBS*, 2005, pp. 171–176.

[17] W. Zhang, S. C. S. Cheung, and M. Chen, "Hiding privacy information in video surveillance system," in *Proc. ICIP*, vol. 3. 2005, pp. 868–871.

[18] S. C. S. Cheung, J. K. Paruchuri, and T. P. Nguyen, "Managing privacy data in pervasive camera networks," in *Proc. ICIP*, 2008, pp. 1676–1679.

[19] C. Neustaedter, S. Greenberg, and M. Boyle, "Blur filtration fails to preserve privacy for home-based video conferencing," *ACM Trans. Comput.-Hum. Interact.*, vol. 13, no. 1, pp. 1–36, 2006.

[20] M. Boyle, C. Edwards, and S. Greenberg, "The effects of filtered video on awareness and privacy," in *Proc. CSCW*, 2000, pp. 1–10.

[21] R. Want, A. Hopper, V. Falcao, and J. Gibbons, "The active badge location system," *ACM Trans. Inf. Syst.*, vol. 10, no. 1, pp. 91–102, 1992.

[22] X. Ren, A. C. Berg, and J. Malik, "Recovering human body configurations using pairwise constraints between parts," in *Proc. ICCV*, 2005, pp. 824–831.

[23] G. Mori and J. Malik, "Recovering 3-D human body configurations using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 7, pp. 1052–1062, Jul. 2006.

[24] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. CVPR*, vol. 1. 2005, pp. 886–893.

[25] C. Rother, V. Kolmogorov, and A. Blake, "Grabcut: Interactive foreground extraction using iterated graph cuts," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 309–314, 2004.

[26] A. Adam, E. Rivlin, and I. Shimshoni, "Robust fragments-based tracking using the integral histogram," in *Proc. CVPR*, vol. 1. 2006, pp. 798–805.

[27] Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images," in *Proc. ICCV*, 2001, pp. 105–112.

[28] V. Kolmogorov and R. Zabih, "What energy functions can be minimized via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 2, pp. 147–159, Feb. 2004.

[29] P. Tu, T. Sebastian, G. Doretto, N. Krahnstoever, J. Rittscher, and T. Yu, "Unified crowd segmentation," in *Proc. ECCV*, vol. 4. 2008, pp. 691–704.

[30] L. Bourdev and J. Malik, "Poselets: Body part detectors trained using 3-D human pose annotations," in *Proc. ICCV*, 2009, pp. 1365–1372.

[31] D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," in *Proc. CVPR*, 2000, pp. 2142–2149.

[32] Z. Zivkovic and B. J. A. Kröse, "An em-like algorithm for color-histogram-based object tracking," in *Proc. CVPR*, vol. 1. 2004, pp. 798–803.

[33] Y. Li, J. Sun, and H.-Y. Shum, "Video object cut and paste," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 595–600, 2005.

[34] Y.-Y. Chuang, B. Curless, D. Salesin, and R. Szeliski, "A Bayesian approach to digital matting," in *Proc. CVPR*, vol. 2. 2001, pp. 264–271.

[35] B. Cabral and L. C. Leedom, "Imaging vector fields using line integral convolution," in *Proc. 20th Conf. SIGGRAPH*, 1993, pp. 263–270.

[36] P. A. Viola and M. J. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. CVPR*, vol. 1. 2001, pp. 511–518.

Prachi Agrawal received the B.Tech. degree in electronics and communication engineering from the International Institute of Information Technology Hyderabad (IIIT-H), Hyderabad, India, in 2008. She is currently pursuing the Masters by Research degree from the Center for Visual Information Technology, IIIT-H.

Her current research interests include computer vision, video processing, and biometrics.

P. J. Narayanan received the Ph.D. degree in computer science from the University of Maryland, College Park.

He is currently a Professor and Dean of Research with the Center for Visual Information Technology, International Institute of Information Technology, Hyderabad, India. His current research interests include computer vision, computer graphics, and virtual reality.

Dr. Narayanan was the Area Chair of the Indian Conference on Computer Vision, Graphics and Image Processing in 2008 and the Asian Conference on Computer Vision in 2009.