Person De-identification in Activity Videos

M. Ivasic-Kos
Department of Informatics
University of Rijeka
Rijeka, Croatia
marinai@uniri.hr

A. Iosifidis, A. Tefas, I. Pitas
Department of Informatics
Aristotle University of Thessaloniki
Thessaloniki, Greece
{aiosif, tefas,pitas}@aiia.csd.auth.gr

***Abstract*— Person identification based on gait recognition has been extensively studied in the last two decades, while information appearing in different action types (like bend) has been recently exploited to this end. However, in most application scenarios it is sufficient to recognize the performed activity, whereas the ID of persons performing activities is not important. Since the same human body representations, e.g., body silhouettes, can be employed for both tasks, there is a need to automatically create privacy preserving representations. We have applied 2D Gaussian filtering to obfuscate the human body silhouettes that implicate information about the person ID. We have experimentally showed how the use of filtering affects the person identification and action recognition performance in different camera setups formed by an arbitrary number of cameras. In addition, the discriminative ability of different activities is examined and discussed in order to detect cases in which it is possible to apply Gaussian filter with a greater variance.**

Keywords—de-identification of human body silhouette; action recognition; Gaussian filter

I. INTRODUCTION

An increasing number of video cameras observe public spaces, like streets, airports, railway and bus stations, shops, schools and other educational institutions. In some use-case scenarios, like video surveillance, there are justified reasons for capturing and sharing acquired multimedia data to authorize personnel, due to security reasons. In most scenarios it is sufficient to detect the activity, whereas data on persons engaged in these activities do not matter. Therefore, there is a strong need for protecting the privacy of persons captured in such multimedia content.

To address this privacy issue, a process of concealing identifying the ID of persons appearing in a given set of data, referred as person de-identification, should be done. Since the performed activity may be of particular interest, the goal in this problem is to protect the privacy of individuals without compromising on the performed activity and other contextual content. Identification is a process opposite to the de-identification, with the former making use of all possible features such as face, silhouette, and gait and body posture to identify persons. In order to automatically obscure such features to prevent identification, appropriate computer vision methods should be devised.

Since humans usually identify persons by observing their faces, it is not surprising the fact that the vast majority of de-identification methods deal with face de-identification. Methods proposed to this end can be categorized in two groups: the ones exploiting image distortion algorithms [1, 2, 3, 4] and those employing the k-Same family algorithms [5, 6, 7, 8]. Methods belonging to the first group alter the facial image regions using data suppression techniques (e.g., by covering part of the face), or some kind of obfuscation, such as blurring [4] or pixelation (i.e., image sub-sampling). Implementation of the k-Same algorithm replaces the face of the person under consideration with a-priori known one belonging to a set of k generic faces, like in [8].

Another approaches deal with the de-identification of the whole human body, instead of just the face, taking into account that gait recognition has been widely used to identify persons at a distance in security applications. In [9, 10], the complete human body is masked using different types of blur functions. In [11], persons on a street scene are replaced by other (a-priori known) persons appearing in a training set of images containing similar scenes. Another way to manipulate the human body regions of an image is to replace them with background [12].

Recently, it has been shown that persons depicted in videos can be identified by the manner in which they perform several activities, such as run, bend, jump, wave one hand, etc. [13, 14]. Moreover, execution style variations among individuals for several activities, like jump, wave one hand, etc., may contain enhanced discriminative information for person identification, when compared to gait. In addition, as has been shown in [14], the combination of information appearing in several different activities may lead to enhanced identification performance. Finally, the observation angle plays an important role on the performance of activity-based person identification. These facts are usually neglected by person identification approaches, and, thus, by person de-identification ones, too.

In this paper, we aim to address the privacy issue of persons performing several activities. The goal was to protect the privacy by automatically obfuscating the human body shape information, so as to drop the performance of person identification approaches. In addition, since the information of the performed activity may be important in some application scenarios, performance of activity recognition approaches should be preserved as much as possible. We are interested in

silhouette-based person de-identification in applications employing multi-camera setups. To change the human body posture information while preserving activity information we have employed 2D Gaussian low pass filters, which are applied to the image regions corresponding to the human body. We have employed the state-of-the-art activity-based person identification method proposed in [13] in order to observe how the adopted approach influences the person identification and action recognition performance on the i3DPost multi-view action recognition database [16]. The obtained results are compared to person identification and action recognition rates obtained without applying the filtering. In addition, performance on action recognition and person identification is compared with respect to the number of arbitrary chosen cameras and different observation angles. Finally, the discriminative ability of specific activities and observation angles is discussed. The paper ends with conclusion and with guidelines for future research.

II. Person Identification Pipeline

In this Section, we briefly describe the method in [13], which has been used as the base method for our activity-based person de-identification framework.

A. Preprocessing Phase

In the preprocessing phase, elementary videos depicting one activity instance (a single movement period) are manually created by splitting multi-period videos. Elementary videos are used for training. In the test phase, elementary or multi-period videos can be used. Generally, the number of frames in elementary videos may vary for different activity types, since activity instances may vary in duration. Moreover, even different instances of the same activity type performed by the same person may vary in duration, e.g., due to mood changes.

By applying appropriate video frame segmentation techniques, like background subtraction, or chroma keying, human body silhouettes are detected in video frames and binary images (masks) encompassing the human body Region Of Interest (ROI) are determined. Fig. 1 shows an example video frame of the i3DPost database depicting an instance of walking activity, the corresponding human body ROI and mask.

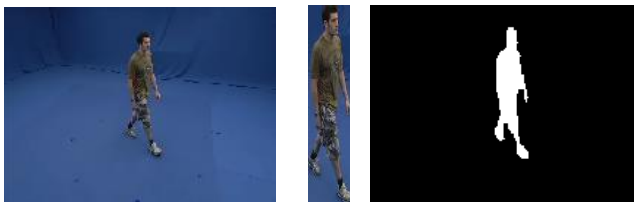

Fig. 1. A video frame depicting a person walking (left), human body image regions (middle) and the corresponding mask (right)

Each of the obtained masks is centered to the human body center of mass and rescaled to a fix size image (experimentally determined to 32x32 pixels in [13]) to obtain a scale-invariant human body posture representation. Example posture images appearing in five activities captured from various viewing angles are shown in Fig. 2.

Fig. 2. Posture images depicting five activities observed from various viewing angles. From left to right: walk, run, jump in place, jump forward, and wave one hand.

The resulting binary posture images are vectorized column-wise, in order to produce the so-called posture vectors.

B. Video Representation and Classification

In the training phase, posture vectors obtained by applying the above described process on the training videos are clustered using the k-means algorithm in order to produce action-independent posture vector prototypes, so called dynemes. Dynemes preserve human body shape, observation angle and activity information. After dynemes calculation, posture vectors of training and test videos can be represented in the dyeneme space by applying fuzzy vector quantization. Training and test videos are finally represented by the so-called action vectors, which are the mean vector of the corresponding posture vectors, represented in the dyneme space.

In order to increase the discriminative information relating to the action class, person ID and observation angle, Linear Discriminant Analysis (LDA) is applied on the training action vectors. Subsequently, each action class, person ID and observation angle is represented by the mean class vector in the corresponding discriminant subspace. Finally, test action vectors are mapped to the discriminant subspaces and classified to the class of the closest class centroid.

III. Proposed Person De-Identification Method

Since the shape of human body silhouettes is crucial for person identification in the above described framework, it should be modified to prevent person identification from activities. To this end, we will modify the color (RGB) values of the video frame pixels corresponding to the human body. In order to do this in a structured way and not to introduce artifacts on the resulted video frame, we employ a zero-mean, discrete two-dimensional Gaussian filter of size h, defined by:

$$h_g(n_1, n_2) = e^{-\left(n_1^2 + n_2^2\right)/2\sigma^2}$$

$$h(n_1, n_2) = \frac{h_g(n_1, n_2)}{\sum_{n_1}\sum_{n_2} h_g}$$

where n_1 and n_2 denote the indices in the filter window of size h and σ is the standard deviation of the Gaussian distribution.

Applied Gaussian filter replaces each pixel in the ROI with a weighted average of the neighboring pixels such that the weight given to a neighbor decreases monotonically with respect to its lateral distance from the central pixel. In this way the effect of distortion is applied locally, which is useful for

keeping the key information relating to the performed activity. By changing the filter parameters, i.e., the value of σ, the effect of the distortion can be appropriately adjusted.

The degree of blurring of a Gaussian filter is parameterized by σ, and the relationship between σ and the degree of smoothing is proportional. A larger σ value implies a larger smoothing and excessive blur of the image features. To determine the appropriate filer size h and adjust the degree of blurring, we have experimented with different filter sizes (h values ranging from 3 to 25 pixels) and with Gaussian distributions of different standard deviation (σ values ranging from 3 to 10). Different h and σ values influence shape of the extracted human body silhouettes and, thus, the action recognition and person identification performance. Since we are focused on lowering the person identification performance, while keeping the action recognition performance high, we have chosen the values providing the maximal p_a/p_I ratio, where p_a, p_I denote the obtained action recognition and person identification rates, respectively (Fig. 3).

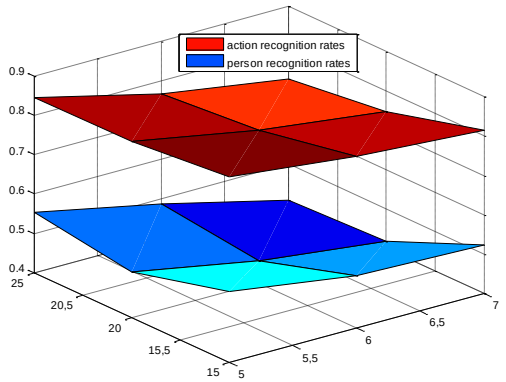

Fig. 3. Person identification and action recognition rates obtained for different Gaussian filter parameter (h, σ) values

After applying the above described process, we have experimentally chosen the values of h=20 and σ=6.4. The corresponding Gaussian filter is shown in Fig. 4.

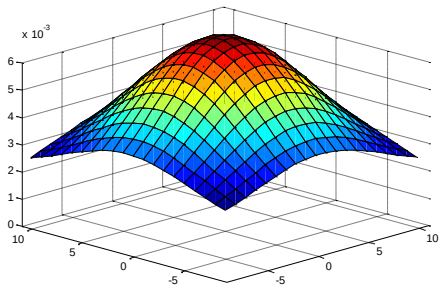

Fig. 4. Gaussian filter with h=20 and σ=6.4 chosen for blurring of the ROI

Since 2-dimensional Gaussian functions are rotationally symmetric, the amount of blurring performed by the filter will be the same in all directions. This property implies that a Gaussian filter will not bias subsequent edge detection in any particular direction and that edges in the resulted image will not be oriented in some particular direction that is known in advance.

We apply the Gaussian filter to image ROIs centered to the human body ROI and having size equal to s times the size of the human body ROI. We have experimentally found that a value of s=1.1 gives the best results with respect to our goal. Fig. 5 shows example video frames and the corresponding human body silhouettes for actions walking and jumping in place after applying the Gaussian filter using the value *s*=1.1. As can be seen in this Figure, the human body silhouettes obtained by using the blurred video frames are coarser, compared to the ones obtained by using the original video frame. This will affect the person identification performance. In addition, it can be seen that the global action information, e.g., opened legs for the case of walking, is preserved. Thus, action recognition performance should not be affected so much.

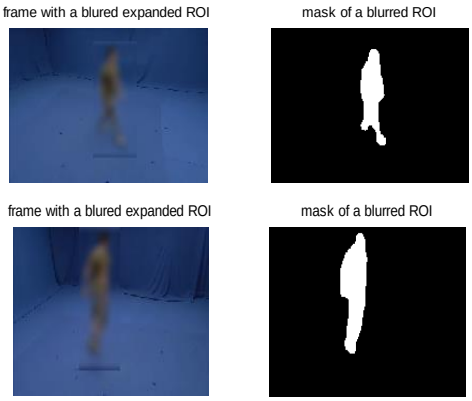

Fig. 5. Frames and corresponding masks depicting walking (top) and jumping in place (bottom) actions after 2D Gaussian filtering using the values h=20, σ=6.4 and s=1.1.

IV. Experimental Results

In this Section we describe experiments conducted on the i3DPost database in order to evaluate how the proposed approach influences the person identification performance and action recognition. The adopted database and experimental setup are described in the following subsection. Experimental results for several experimental scenarios are subsequently provided

A. The i3DPost database

The experiments are performed on the i3DPost multi-view activity database [16] containing 64 high-resolution (1920×1080 pixel) videos depicting eight persons (six males and two females) performing eight simple actions, which may be periodic, e.g., walking, or not, e.g., bend. Actions differ in duration and, thus, the number of video frames forming videos may differ. Since for non-periodic actions, one action instance is available per person, we have used the videos depicting periodic actions, i.e., walk, run, jump in place, jump forward, and wave one hand, in our experiments. Each person in the database is captured by eight cameras located at a height of 2 m from the studio floor, arranged at every 45º degrees of a circle with a diameter 8 m to provide 360º coverage of the capture volume (Fig. 6). Consequently, an action instance

(e.g., a walk step) performed by a person is depicted in eight videos, each captured by one of the eight observation angles. In our experiments we have used elementary videos obtained from the original videos in the database in order to train the algorithm. In the test phase we have used the entire sequences in order to evaluate the influence of blurring in both, person identification and action recognition.

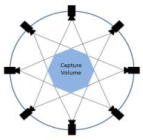

Fig. 6. Set-up of eight cameras which enables 360º coverage of the capture volume [16].

Example video fames depicting two persons walking and jumping captured from all the available viewing angles is shown in Fig. 7. Notice that each person has optionally chosen a direction to perform the action, so that its frontal view is not captured by the same camera angle in all cases.

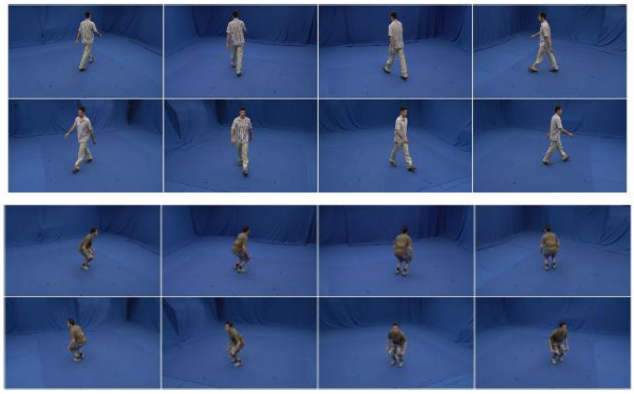

Fig. 7. Video frames depicting a person walking (above) and jumping forward (below) from all the 8 available observation angles

B. Multi-view Person De-Identification

In our first set of experiments we have applied the proposed method on the videos of the i3DPost database. Since each action instance is depicted by eight cameras, it is reasonable to fuse the classification results obtained for each video (corresponding to one observation angle) in order to increase performance in both person identification and action recognition tasks. We have used majority voting fusion to this end. In addition, we have tested the performance of the method for single-view view-independent person identification/action recognition. In the latter case, we assume that each video in the database is independent to the remaining ones. The performance obtained for both cases is shown in Table I. As can be seen, enhanced performance is obtained in the first case, verifying the conclusions drawn in [13].

To examine how Gaussian filtering affects the action recognition performance and to what extent preserves privacy, the obtained results are compared to published results when no filter is applied on the same framework [13], where a person identification rate equal to 94.37% has been reported for the case where all the eight cameras are used for identification. Action recognition performance has not been reported in [13], but the obtained performance (equal to 95%) is very good. In fact we have tested the method on the original videos of the database and a performance equal to 100% has been obtained for both multi-view action recognition and person identification.

TABLE I. MEAN PERSON IDENTIFICATION AND ACTION RECOGNITION PERFORMANCE WHEN GAUSSIAN FILTER IS APPLIED TO THE HUMAN BODY ROI

Action recognition		**Person identification**	
Single-view	***Multi-view***	***Single-view***	***Multi-view***
0.8406	**0.95**	0.4875	**0.70**

Comparing obtained results before and after the application of Gaussian filter, it is obvious that application of the Gaussian filter decreases the action recognition and person identification performance. However, the decrease in action recognition performance is very low (~5%), while the decrease in person identification performance is high (~30%)

C. Person De-Identification in the Case of Total Human Body Occlusion in Some of the Cameras

In real applications, it is possible that a person performing an activity is not visible to all cameras, either because he/she is not inside their field of view, or because he/she is occluded. To simulate the person identification and de-identification scenario when a person is captured by an arbitrary number of cameras, a set of experiments was set as follows. In the training phase, we have used all the eight cameras in order to train the algorithm, while in the test phase a subset of the available cameras was used for identification. The cameras used for identification were randomly selected. It should be noted that, as the motion direction of persons differs, the test cameras do not correspond to specific observation angles. This means that a camera may depict any view of a person.

In Tables II and III we illustrate the performance obtained for different numbers of randomly chosen test cameras for the case of action recognition and person identification, respectively.

TABLE II. MEAN ACTION MULTI-VIEW RECOGNITION RATE WHEN DIFFERENT NUMBER OF RANDOMLY SELECTED TEST CAMERAS ARE USED

No of used cameras	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Mean (20 experiments)	0.77	0.79	0.86	0.89	0.905	0.909	**0.911**	0.9

TABLE III. MEAN PERSON MULTI-VIEW IDENTIFICATION RATE WHEN DIFFERENT NUMBER OF RANDOMLY SELECTED TEST CAMERAS ARE USED

No of used cameras	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Mean (20 experiments)	0.45	0.42	0.51	0.53	0.55	**0.57**	**0.57**	0.53

As can be seen, the more cameras are used, the better the performance tends to be. This is reasonable, since by using more cameras the probability of including a "good observation angle" in the identification process is higher. In addition, it

can be seen that there is a plateau in performance with respect to the number of adopted cameras. That is, the highest action recognition performance is obtained for 7 cameras, while the highest person identification performance is obtained for 6 or 7 cameras.

The relation in performance between action recognition and person identification, for different number of test cameras, is shown in the Fig. 8. It is important to note that difference in action recognition and person identification performance exceeds 30% in most cases.

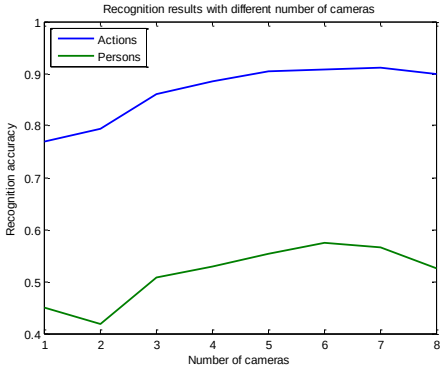

Fig. 8. Relation between activities recognition rates and person identification rates when different number of cameras is used in the test phase

In order to graphically illustrate the impact of Gaussian filter application on action recognition and identification performance, we have also performed experiments for varying number of test cameras by using the original videos in the database. Comparison results can be seen in Fig. 9 and 10. Given results refer to randomly chosen cameras for a corresponding subset (number) of the available cameras. Since, each camera does not correspond to a specific observation angle obtained results may differ in various iterations of experiment. Fig. 9 and 10 show results of arbitrary selected iteration, while Tables II and III show the mean value obtained after 20 iterations of experiment are performed.

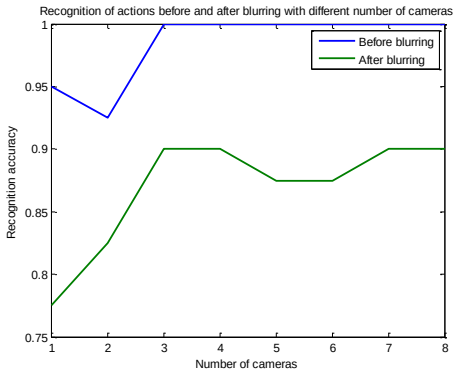

Fig. 9. Comparison of activities recognition rates before and after application of Gaussian filter when different number of cameras is used in the test phase

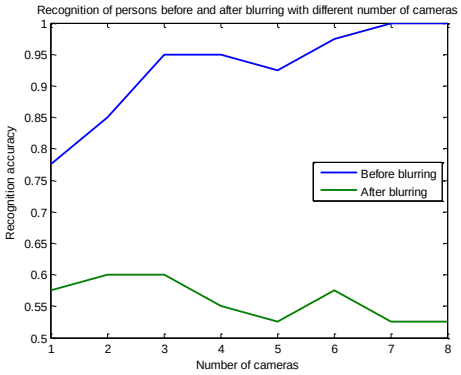

Fig. 10. Comparison of person identification rates before and after application of Gaussian filter when different number of cameras is used in the test phase

D. Discriminant Ability of Different Observation Angles

As has been described in the previous subsection, different observation angles have different discriminative ability, in terms of action recognition and person identification. This is due to the fact that the human body shape during action execution differs significantly, when the person is observed by different viewing angles. After applying the Gaussian filters it is possible that some human body shapes will become less discriminate, i.e. that body shapes of different persons performing various actions will more closely resemble. We have tested the discriminant ability of all the available observation angles after applying Gaussian blurring and illustrate the corresponding classification rates in Tables IV, V and Figure 11, for the case of person identification and action recognition.

TABLE IV. MEAN AND MEDIAN IDENTIFICATION RATES FOR ALL PERSONS IN THE DATABASE OBTAINED FROM A SPECIFIC OBSERVATION ANGLE

	0º	**45º**	**90º**	**135º**	**180º**	**225º**	**270º**	**315º**
Mean	0.19	0.34	0.39	0.37	0.19	0.39	**0.41**	0.4
Median	0.21	0.42	0.38	0.38	0.21	0.38	0.42	0.43

TABLE V. MEAN AND MEDIAN ACTION RECOGNITION RATES OBTAINED FROM A SPECIFIC OBSERVATION ANGLE FOR ALL PERSONS IN THE DATABASE

	0º	**45º**	**90º**	**135º**	**180º**	**225º**	**270º**	**315º**
Mean	0.64	0.63	**0.82**	0.59	0.53	0.76	**0.84**	0.69
Median	0.63	0.69	0.88	0.63	0.54	0.79	0.88	0.75

We denote, by identification ability of a viewing angle the mean value of identification rates obtained for all persons in and for all activities in database. Similarly, action recognition ability of a viewing angle corresponds to the mean value of action recognition rates obtained for all actions and for all persons in database. It turned out that the side views are more discriminative for person identification, since identification performance equal to 40% and 41% has been obtained for observation angles of 270^o and 315^o (with respect to the frontal human body direction). Median of identification rates are not significantly different from the mean ones, indicating that there are no major differences in the identification rates when looking discriminate ability of an viewing angle for each individual action. In the case of actions, the observation angles

equal to 270^o and 90^o (with respect to the frontal human body direction) have been found to provide the best performance.

Taking into account the obtained results, specific viewing angles that achieve high identification and recognition accuracy (like angle of 270 º in this case) could be blurred with Gaussian filter with larger variance in order to improve the results of de-identification.

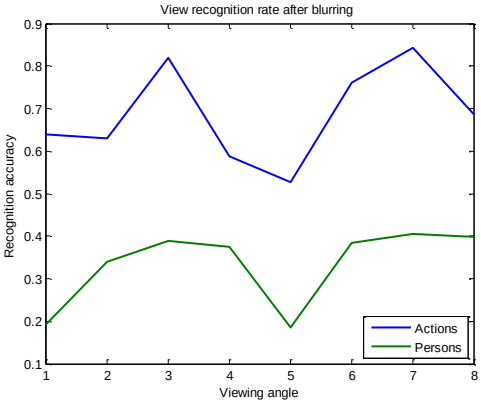

Fig. 11. Identification and action recognition rates obtained on the videos depicting the person from a specific view angle (1 - 0º, 1 - 45º, continues in a clockwise direction until the number 8 - 315º)

E. Discriminant ability of a specific actions

In [14] it has been shown that the discriminative information of different actions is not the same. In addition, depending on the performed action, the identification performance may differ. We have also tested the discriminative ability of different actions after Gaussian blurring. The obtained results are shown in Table VI. As can be seen, the most discriminative action was found to be jump in place, providing a person identification rate equal to 38.75%, followed by activities wave one hand and jump forward. Differences between the median and mean values of person identification rates indicate that there is no action that is most discriminate for each person.

TABLE VI. MEAN AND MEDIAN IDENTIFICATION RATES FOR ALL PERSONS IN DATABASE WITH RESPECT TO EACH ACTION

Identification	**walk**	**run**	**jump in place**	**Jump forward**	**Wave one hand**
Mean	0.3482	0.3494	**0.3875**	0.3678	**0.3756**
Median	0.375	0.3818	0.4062	0.3844	0.4125

Taking into account the obtained results, specific activities that contain high discriminant information for identification (like wave one hand in this case) could be blurred with Gaussian filter with larger variance in order to improve the results of de-identification.

V. CONCLUSION AND FUTURE WORK

In this paper, we have proposed a pipeline for person de-identification based on activities. We have employed Gaussian blurring in order to change the obtained human body silhouettes, so as to discard identity information, while preserving action information. Preliminary results indicate that such an approach is able to lead to a significant decrease in person identification performance, while achieving a relatively high action classification performance. In order to further improve the de-identification results, the discriminative ability of different observation angles and of specific activities can also be taken into account in order to apply additional blurring steps. In our future work we will investigate the robustness of the proposed approach against reverse de-identification processes, like Wiener deconvolution.

ACKNOWLEDGMENT

The authors are grateful to COST project, Action IC1206 which has supported and facilitated this collaboration.

REFERENCES

[1] M. Boyle, C. Edwards, and S. Greenberg. The effects of filtered video on awareness and privacy. In *ACM Conference on Computer Supported Cooperative Work*, pages 1–10, Philadelphia, PA, December 2000.

[2] J. Crowley, J. Coutaz, and F. Berard. Things that see. *Communications of the ACM*, 43(3):54–64, 2000.

[3] C. Neustaedter, and S. Greenberg. Balancing privacy and awareness in home media spaces. In *Workshop on Ubicomp Communities: Privacy as Boundary Negotiation*, 2003.

[4] M. Nishiyama, H. Takeshima, J. Shotton, T. Kozakaya, and O. Yamaguchi, "Facial deblur inference to improve recognition of blurred faces," in CVPR, 2009

[5] R. Gross, E. Airoldi, B. Malin, and L. Sweeney. Integrating utility into face de-identification. In *Workshop on Privacy Enhancing Technologies (PET)*, June 2005.

[6] R. Gross, L. Sweeney, T. de la Torre, and S. Baker. Semi-supervised learning of multi-factor models for face de-identification. In *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2008.

[7] E. Newton, L. Sweeney, and B. Malin. Preserving privacy by de-identifying facial images. *IEEE Transactions on Knowledge and Data Engineering*, 17(2):232–243, 2005.

[8] E. Newton, L. Sweeney, B. Malin. Preserving privacy by de-identifying facial images. *IEEE Transactions on Knowledge and Data Engineering, 17:232–243, 2003.*

[9] P. Agrawal, P. J. Narayanan, "Person De-identification in Videos", *The IEEE Trans. on Circuits and Systems for Video Technology (TCSVT), 2010.*

[10] P. Agrawal, P.J. Narayanan, Person De-Identification in Videos, IEEE Transactions on Biometrics Compendium, Volume:21, Issue: 3, 2011

[11] K. Chinomi, N. Nitta, Y. Ito, N. Babaguchi, "Prisurv: Privacy protected video surveillance system using adaptive visual abstraction," in MMM, 2008, pp. 144–154.

[12] A. Nodari, M. Vanetti, I. Gallo; Digital Privacy: Replacing Pedestrians from Google Street View Images, 21st Int. Conference of Pattern Recognition (ICPR), 2012

[13] A. Iosifidis, A. Tefas, I. Pitas, Activity based Person Identification using Fuzzy Representation and Discriminant Learning, IEEE Transactions on Information Forensics and Security, Vol. 7, Issue: 2, 2012, pp. 530 – 542

[14] A. Iosifidis, A. Tefas, N. Nikolaidis, I. Pitas, Learning Human Identity using View-Invariant Multi-view Moverment Representation, Biometrics & ID Management: COST 2101 International Workshop (BioID), 2011

[15] R. Haralick and L. Shapiro Computer and Robot Vision, Addison-Wesley Publishing Company, 1992.

[16] N. Gkalelis, H. Kim, A. Hilton, N. Nikolaidis, and I. Pitas, The i3DPost multiview and 3D human action/interaction database, in: 6th Conference on Visual Media Production, 159–168, 2009