Assessment of Radiation Dose Received by Nuclear Plant Personnel through a Video-based Surveillance System

Carlos Alexandre F. Jorge\textsuperscript{1,2}, José Manoel de Seixas\textsuperscript{2b,3}, Eduardo Antônio B. Silva\textsuperscript{2c,3}, Antônio Carlos A. Mól\textsuperscript{1d,4}, José G. T. Monteiro\textsuperscript{3e}, Gabriel R. Lira\textsuperscript{3f}, Pedro H. S. Silva\textsuperscript{3h}

\textsuperscript{a}calexandre@ien.gov.br, \textsuperscript{b}seixas@lps.ufrj.br; \textsuperscript{c}eduardo@smt.ufrj.br; \textsuperscript{d}mol@ien.gov.br; \textsuperscript{e}gmonteiro@poli.ufrj.br; \textsuperscript{f}gabriellira@poli.ufrj.br; \textsuperscript{h}pedrosevenini@poli.ufrj.br

Mailing Addresses:
\textsuperscript{1}Comissão Nacional de Energia Nuclear, Instituto de Engenharia Nuclear
Rua Hélio de Almeida,75, Cidade Universitária, Ilha do Fundão, P.O. Box 68550, ZIP 21941-906, Rio de Janeiro, RJ, Brazil
\textsuperscript{2}Programa de Engenharia Elétrica, COPPE, Universidade Federal do Rio de Janeiro
Cidade Universitária, Ilha do Fundão, ZIP 21941-972, Rio de Janeiro, RJ, Brazil
\textsuperscript{3}Escola Politécnica, Universidade Federal do Rio de Janeiro
Cidade Universitária, Ilha do Fundão, ZIP 21945-970, Rio de Janeiro, RJ, Brazil
\textsuperscript{4}Instituto Nacional de Ciência e Tecnologia de Reatores Nucleares Inovadores/CNPq, Brazil

Abstract
This work aims at developing a video-based surveillance system for safety purposes in nuclear plants. The objective is to assess the radiation dose received by nuclear plant personnel, while they execute daily tasks, by means of computer vision methods. The system is conceived to provide some redundancy to the radioprotection means currently in use, being independent and complementary to them. After evaluating some methods from the literature for automatic target detection and tracking, a novel system is developed to correctly detect, track and identify people, so that the radiation dose received by each person is reliably computed. The video data are supplied by cameras installed in the nuclear plant room. Radiation dose rate mapping is combined with the tracking results to account for the received doses. We provide experimental results from a research reactor room, which show that the proposed system achieves radiation dose estimates that are in general similar to the ones of the ground truth. The database developed in this work for performance evaluation has been made publicly available for the research community.

Keywords: Radiation monitoring, Radiological protection, Safety in nuclear plants, Video-based surveillance, Multi-target detection and tracking, Target re-identification.
1. Introduction

Working in the nuclear field unavoidably involves receiving radiation doses. Thus, it is important to seek ways of minimizing dose exposure, following the ALARA (As Low As Reasonably Achievable) principle (Valentin 2007).

Traditional approaches for radiological protection involve the use of radiation monitors: fixed, portable, individual and/or personal ones (Knoll 1999). The fixed and portable pieces of equipment are used to monitor ambient radiation dose rate levels. The individual/personal ones are used by each worker, in order to assess the total radiation dose received during the times each one stays within the plant. Some of the latter lack precision in the results, while others need a long measurement time to give an output.

In this work, a video-based surveillance system is developed to track nuclear plant personnel, in order to assess the radiation doses received by each worker while they execute their tasks. This is planned to be a redundant safety system: the video-based system may be used simultaneously to the more traditional monitoring approaches, in order to supply further means to assess received radiation doses. In addition, it can be used as the single monitoring strategy in case of failure of the traditional ones. The proposed system can fill the gaps of traditional approaches for supplying the received doses.

The experimental data used here come from a nuclear research reactor (Argonauta), which is located at the Nuclear Engineering Institute (IEN - Instituto de Engenharia Nuclear), in Rio de Janeiro, Brazil and became operational in 1965 (Mól et al. 2009). A very important use of Argonauta is non-destructive evaluation of different types of materials. One example of activity within Argonauta’s room may comprise somebody entering it to put a sample to be irradiated inside or in front of its output radiation channel (named “J9”) and removing it when irradiation finishes (see Fig. 1). Some activities require the operator to run some experiments in front of “J9”, as in spectrography. For this reactor, a surveillance system must detect people entering the plant, tracking each one individually. Assessment of the received radiation dose is based on their tracked positions and on radiation dose rate mapping within Argonauta’s room. Such a system has to reliably track each person, even when occlusions occur, so that a correct received dose value can be assigned to each worker.
A large amount of work has been done for detecting targets and tracking them along videos (Yilmaz et al. 2006, Maggio and Cavallaro 2011). We have implemented state-of-the-art methods and tested them on Argonauta data. None of them was able to adequately solve our radiation dose estimation problem. Many of these methods tend to lose tracking or to drift. Often, when recovering from loss of tracking, they are not able to re-identify targets correctly. These tend to have a strong impact on the assessment of radiation dose received by each worker. Motivated by this scenario, in this paper we propose a new tracking system for estimation of individual radiation doses in nuclear plants. Results from the early stages of its development have been described in (Jorge et al. 2013).

Among the evaluated methods, the best performing one was the method we refer to here as Führ-Jung (see Section 3). It was able to track targets for longer times before drift or loss of tracking occur, even in occlusion situations. However, it can not re-identify targets when re-initializing tracking. This was the method selected for comparison with the proposed system.

This paper is organized as follows. Section 2 comments on radiological protection means and Section 3 reviews video-based surveillance methods. Section 4 describes the proposed system for which results are given in Section 5. A comparative analysis to a state-of-the-art method and a discussion are also provided for the database acquired at Argonauta room, available in (Jorge et al. 2016). Conclusions are provided in Section 6.

2. Radiological protection

Currently, the approaches used for radiological protection in IEN are based on the use of fixed, portable and individual/personal radiation monitors. Although the traditional approaches used for radiological protection described here focus in the Argonauta reactor, they are general purpose, having a broader range of applications. The individual/personal monitors used are: (i) Thermoluminiscent detector (TLD) personal monitors and (ii) Ionization chambers (pen dosimeters), (Knoll 1999). The former provides dose results after some time of use (typically a month). The latter gives readable result at the end of operation,
but it lacks precision. Digital radiation monitors are barely used due to their high cost (only a few of them are available at IEN). Therefore, mainly due to low precision or long time to know the results, there are usually gaps in the determination of the radiation dose received by personnel.

From another perspective, virtual environments (VE) have been used as nuclear plants mock-ups, where dose rate distribution is accounted for, and avatars (virtual persons) may navigate within (Mól et al. 2009, Sebok et al. 2002, Ródenas et al. 2004). Each avatar is assigned the dose received during a simulation.

Some authors (Rindahl et al. 2005, Ishii et al. 2006) mention the need of tracking people within nuclear plants for radiological safety purposes, by possibly using technologies based on head-mounted displays (HMD), ultrasound transducers and radio-frequency identification (RFID). However, they comment on the difficulties of implementing such systems, as the need of users porting those HMDs and laptops, problems with ultrasound reflection and the fact that RFID tags would enable measurements only at certain locations.

Therefore, up to the knowledge of the authors, there is no work in the literature as the one proposed here, which makes use of a markerless computer vision-based system totally independent of users’ behaviour, and not requiring them to carry any piece of equipment. The only need is the installation of cameras in the room (we used two in our experiments) and a map of the radiation doses inside the room.

3. People detection and tracking in videos

To track a target in a video, it must be first detected in the scene. For the whole task, some classes of methods may be distinguished: (i) background scene removal to detect candidate targets, resulting in regions of interest (ROIs), (Stauffer and Grimson 2000, Haritaoglu et al. 2000, Barnich and Droogenbroeck 2011); (ii) methods specifically developed for detecting persons in the scene (Dollár et al. 2010); and (iii) ROI tracking along the video, based on extracted parameters (Maggio and Cavallaro 2011). For a practical system, often more than one class of methods may be combined. For example, background removal methods and/or a method trained specifically to detect persons may supply the initialization of ROI tracking.
For background removal, the Gaussian Mixture Model (GMM), (Stauffer and Grimson 2000) was considered in this work. This method uses multiple Gaussian functions to model each pixel distribution and decide whether a given pixel from a new incoming frame belongs or not to the background. In general, GMM is able to deal with complex backgrounds that might cause false alarms. In (KaewTraKulPong and Bowden 2001, Zivkovic 2004, Zivkovic and van der Heijden 2004), GMM has been improved to enable shadow discrimination and provide a more computationally efficient implementation. This version was evaluated in stand-alone mode and was also used to initialize the tracking methods. The GMM is one of the most widely used methods for this purpose.

As for the tracking methods, the ones evaluated in this work were: (i) Continuously-adapted Mean-Shift (Camshift), (Bradski 1998, Comaniciu et al. 2003); (ii) Fragments-based Tracking (FragTrack), (Adam et al. 2006); and Tracking-Learning-Detection (TLD), (Kalal et al. 2012). The first is a popular method for target tracking, and is based on the color distribution obtained from a supplied ROI. It can adapt the latter to accommodate scale variations.

The FragTrack method is also widely used. It deals with occlusions by splitting the target into patches. It tends to be more robust than Camshift for tracking targets when partial occlusions occur.

The TLD method combines target detection in each frame with an inferred target trajectory constructed along time by a tracker. The learning stage adapts the detector with new target observations, in order to accommodate target variations. This latter stage comprises two subsystems: (i) one to infer false alarms and (ii) another to infer missed targets.

The Camshift, FragTrack and TLD methods, however, require an ROI to be supplied.

A recently developed method was also evaluated: Combining Patch Matching and Detection for Robust Pedestrian Tracking in Monocular Calibrated Cameras (Führ and Jung 2014), referred to here as Führ-Jung. This comprises a full pipeline for detecting and tracking pedestrians in vertical poses in the scene, including re-initialization of lost tracks. It also follows the approach of patch-based tracking. As opposed to the aforementioned methods, which were initially developed for tracking one target at a time, the Führ-Jung method supports multi-target tracking. It requires a calibrated camera to estimate target positions in
the world coordinate system (WCS). Also, it initializes ROIs by combining a pedestrian detection method, the Fastest Pedestrian Detection in the West (FPDW), (Dollár et al. 2010), with background removal using the Visual Background Extractor (ViBE), (Barnich and Droogenbroeck 2011). The latter is an alternative to GMM.

4. The developed system

Besides correct tracking, an additional requirement from the application is the correct identification of targets, because if the target identities are lost or switched, this will result in errors in dose computation. Target detection and tracking was performed on a frame by frame basis with the application of data association along frames. This type of tracking belongs to the class of tracking-by-detection methods, for which representative works are found in (Kalal et al. 2012, Breitenstein et al. 2011, Pirsiavash et al. 2011, Benfold and Reid 2011). The building blocks for the proposed system (see Fig. 2) are described in the next subsections.

![Figure 2. Block diagram for the proposed system.](image)

4.1. Initialization

The Initialization block comprises two steps: (i) people detection by FPDW and (ii) background removal by GMM. Although a very good method for people detection, FPDW may produce false alarms. Thus, a background removal method can be used to validate FPDW detections: in case a foreground is found that comprises a large enough percentage of the ROI it belongs to, the detection is accepted, otherwise it is rejected as a false alarm (Führ and Jung 2014). In the proposed system, a minimum threshold of 40 percent of foreground area within a given ROI was applied. Further validations were also adopted, similarly to in (Führ and Jung 2014): (i) aspect ratio of the ROI in the image plane should be in the range 0.2 - 0.6; and (ii) persons’ heights in the WCS should be in the range from 1.5 to 2.5 m.

The foregrounds obtained with GMM were improved by using mathematical morphological operations (opening and closing), (González and Woods 2007). The former filters background noise, while the latter connects split parts and fills holes within foregrounds.
Background removal was also employed to improve the ROIs obtained through FPDW: the produced ROIs may extend beyond the foreground. Thus, foreground limits may be used to shrink a ROI to better enclose the foreground. This effect is illustrated in Fig. 3.

Figure 3. ROI supplied by FPDW (dashed lines), along with the one improved by GMM (continuous lines).

4.2. People re-identification

4.2.1. Using previous positions

When there are multiple targets, the correct target association to their identities is an issue. This is especially relevant when a target is missing for a period of time, or when the paths of two targets cross each other.

Data association by shortest distance in nearby frames was performed when targets were sufficiently apart from each other in the WCS. A minimum distance of two meters was chosen experimentally. Furthermore, additional strategies for data association were adopted, as described next.

4.2.2. Using face/head parameters

Since in the Argonauta reactor people usually wear clothes of the same color, the colors of peoples’ bodies are not useful for re-identification. Thus, face/head colors are considered instead. Target association is performed between adjacent frames whenever possible. Tests showed that the face/head ROI parameters tend to present higher similarity for the same person in nearby frames when compared to parameters from the other persons, as abrupt 3D pose changes between adjacent frames are rare at 30 frames per second.

In the proposed system, the face/head ROI is found based on the body’s ROI supplied by FPDW. Experimentally, a good candidate region comprises the intersection of the upper 1/6th
part and the central third part taken in the horizontal direction of the body ROIs. This is similar to what is done in the work of Zhang et al. (2009), although the approach used is different.

The system should be robust against typical problems that may occur due to occlusions. Face/head localization relatively to the body’s ROI was improved considering also the foreground in the upper 1/6th part of the body ROI, in order to allow the face/head ROI to slide horizontally, as shown in Fig. 4.

Figure 4. Solution adopted to handle bad face/head location.

When people are severely occluded by one another, two situations may arise: (i) the occluded person is considered to be in a prohibitive condition, meaning he/she can not be correctly identified, and is given a false temporary identity; or (ii) the lost target is assigned to the last detected WCS position until it is detected again.

4.2.3. Using body parameters

This alternative uses Speed-Up Robust Features (SURF) parameters (Bay et al. 2008). These parameters are descriptors related to saliences in the image taken within a given ROI, thus not depending on color parameters. SURF parameters are very stable and discriminative, enabling matches between different views of a given scene or object, even when subject to affine transformations.

First, for each target, the SURF parameters are obtained within the ROIs that define them. These ROIs are supplied by FPDW. Then, for an incoming frame, SURF parameters are computed from the new ROIs supplied by FPDW, and matched to the previous ones. It was verified that even for persons with very similar appearances and experiencing partial occlusion, accurate matching could be achieved.

4.3. Multi-layer strategy for re-identification
When targets are missed, a multi-layer strategy for matching parameters is used. Four layers are implemented, by considering comparisons of the current frame to past ones. The first layer comprises the frame immediately before the current one; the second comprises the interval from the past 2nd to the 30th frames; the third comprises the interval from the past 31st to the 150th frame; and the last layer comprises the interval from the past frame number 151 to the frame number 600. The system verifies all layers. This increases robustness against identity changes. The fourth layer serves as a long term history for each target.

For the matching of face/head color parameters, the Battacharyya distances (Battacharyya 1943) are used. For the case of SURF parameters, a sum of the percentage of matched keypoints is used.

4.4. Further analysis to re-identify targets

A post-processing stage was applied for further corrections, comprising the heuristics described in what follows. The first two use contextual modelling (Maggio and Cavallaro 2011) as a criterion.

1) No one disappears or appears suddenly within Argonauta’s room, unless in some specific places, as the entrance door, for example.

2) If somebody disappears near the output radiation channel labelled as “J9” (see Fig. 1) it means the person crouched to put or pick up some irradiated sample there, for example. The last detected target’s WCS position is used for dose computation, until the person is detected again.

3) When someone is in a condition that does not allow for re-identification, correct identities may be recovered from the shortest distances in the WCS computed from past frames.

4.5. Camera switching

The system may switch camera views, to get the best one. The decision for switching the cameras also uses contextual information. Regions are defined in the WCS where specific situations are supposed to occur, as occlusion caused by the spectrograph, or when a person
becomes severely occluded by another one. Switching is performed independently for each person. Therefore, the doses are computed from the best possible view.

5. Experimental results

5.1. Databases generation and pre-processing

The development of the proposed system required obtaining both dose rate and video databases. A radiation measurement campaign was run in order to obtain the dose rate distribution in the part of Argonauta’s room that presents higher dose rate values during operations, near “J9” (see Fig. 1). Measurements were collected for gamma radiation with a portable monitor. This piece of equipment has an intrinsic relative error in the range −20 to +20%. Measurements were taken over a grid of 0.5-m resolution within Argonauta’s room, and then interpolated to lead to a smooth dose rate surface. The interpolated resulting dose rate distribution is shown in Fig. 5. The red dots correspond to the measured points. As expected, the peak dose occurs in front of “J9”.

Figure 5. The measured dose rate distribution resulting from the campaign.

The generated video database comprises videos of people performing typical movements according to Argonata operation. Two cameras were installed within Argonauta’s room, pointing towards “J9” from different perspectives. Videos were recorded with a resolution of 1,280 x 1,024 pixels. The reason for that was to avoid artefacts that could have influence on the performance of the methods. The frame rate adopted was 30 fps. The database is available for reference and open for further work at IEN (Jorge et al. 2016).

Camera calibration was performed to obtain their intrinsic and extrinsic parameters (Hartley and Zisserman 2004). The former can be used to optionally rectify the frames, correcting for typical distortions that occur during data acquisition. The extrinsic parameters are used to perform the conversions between the image plane and the WCS (Hartley and Zisserman 2004). For this work, the WCS positioning is necessary to estimate each worker coordinates.
on the floor, that enable computation of his/her received doses from the map of radiation dose rates. Equation 1 shows the projection transformation between the image plane and the WCS.

\[ wX = PX \]  \hspace{1cm} (1)

where:
- \( x \): coordinates in the image plane;
- \( X \): coordinates in the WCS;
- \( P \): projection matrix \((3 \times 4)\);
- \( w \): a scaling factor.

The projection matrix is, in turn, given by:

\[ P = K[Rt] \]  \hspace{1cm} (2)

where:
- \( K \): intrinsic parameters \((3 \times 3)\);
- \( R \): rotation matrix \((3 \times 3)\);
- \( t \): translation vector \((3 \times 1)\).

The intrinsic parameters were estimated by using a rectangular pattern similar to a chessboard \((9 \times 7 \text{ squares of } 120 \text{ mm each})\) and acquiring a number of observations at different positions and a variety of poses (Bouguet 2013). The Random Sample Consensus (RANSAC) method (Fishler and Bolles 1981) was selected due to its robustness against outliers. RANSAC was also used to estimate the extrinsic parameters (Hartley and Zisserman 2004).

The system computes doses by using the projection between the image plane and the floor. This projection reduces to a \(3 \times 3\) homography (Hartley and Zisserman 2004). However, the full \(3 \times 4\) projection matrix is still required by the proposed system in order to estimate the persons’ head positions in the WCS. Note that the same applies to the Führ-Jung method.

5.2. Ground truth
The position ground truth (GT) was also obtained to allow a proper assessment of the proposed system’s performance. It comprises manual markings of each person in the videos. The GT also includes bounding boxes of a person’s body. This was performed by marking persons only on a subset of the frames (e.g., one out of \( N \) frames), and interpolating linearly between any two of them (Silva et al. 2014). Besides the body, the GT of the feet position was also marked, wherever they were visible. This GT also accompanies the aforementioned repository.

5.3. The proposed method

Results are given in terms of estimated dose received by personnel, compared to the GT. Both are computed by correlating the workers’ positions in the WCS with the radiation dose map, and are computed by summing up the doses received during the time span of each frame. The measurement cones that are presented in Figs. 6 to 9 consider the intrinsic error of the piece of equipment used (in the range \(-20\) to \(+20\) \%). The deviations between estimated and GT doses are due to the errors in people’s position as estimated by the proposed tracking method, as both use the same interpolated dose rate distribution.

The right camera is used preferentially, since it supplies the best view of the region under analysis. The left one is used when the right one is in an occlusion situation.

Figure 6 concerns the first example, a spectrometry experiment executed by a single person. He crouched just behind the spectrograph, occluding his feet from the right camera point of view from frames 3839 to 4949. If only the right camera was considered, this would cause an error in the estimation of the feet position. The detected bounding box could only enclose the person’s body from his knees up. When projected to the WCS, this induces an error on the person’s estimated position relative to the GT. As this location is near the peak in the dose rate distribution (see Fig. 5), a one-meter deviation in the position estimation would cause a large variation in the dose rate value, and consequently in the final received dose. However, this problem was solved by resorting to the left camera, where the feet are free from occlusion, and the person’s position and dose are well estimated.
Figure 6. First example: dose estimation obtained using both cameras; tracked by the proposed system (in blue) and ground truth (in red). The lilac color indicates the superposition of the blue and red regions. The same applies to Figs. 7, 8 and 9.

Next, we consider the second example, when two persons move within Argonauta’s room. Both enter just after one another. While one person stands between the entrance door and “J9”, the other goes to a computer in the right side (see Fig. 4), which is part of the data acquisition apparatus. Then, this latter comes back to a position near the other. Fig. 7 shows the dose results for each analysed person. Fig. 7a shows the dose received by the person who stood between the entrance door and “J9”, labelled as Target A. Fig. 7b shows the dose received by the other person who moved towards the computer, labelled as Target B. Despite the occlusion situations involving the two persons, at the beginning and at the end of the scene, the system was able to track each one correctly by switching between the two cameras.

a)

b)

Figure 7. Second example: dose estimation for a two-person situation: each item corresponds to a different target. Tracked result by the proposed system (in blue) and ground truth (GT – in red).

In the third example, both persons enter almost together in Argonauta’s room, and go to the right side of the reactor, disappearing from both cameras. This is a challenging situation, because the correct dose computation demands that target association is correct when they return to the scene. Fig. 8 shows this was achieved.

a)

b)

Figure 8. Third example: dose estimation for a two-person situation: each item corresponds to a different target. Tracked result by the proposed system (in blue) and ground truth (GT – in red). The gap in the graphs occur when both people are not visible by neither camera.
In all examples, we show that by using information from both the right and the left cameras one can achieve measurements that are accurate enough even in challenging situations.

5.4. Comparative analysis

Simulations on our database were performed with methods from the literature. We have observed that Camshift tends to lose tracking after just few frames, resulting in unrealistic bounding boxes early in video sequences. The TLD method tends to lock in the background when persons change pose by turning from front to side views, for example. FragTrack was one of the best performing methods, but requires long processing time due to the size of the search region. The Führ-Jung method resulted superior to the others, even when considerable occlusion occurs between people. However, it suffers from drifts and tracking losses in severe occlusion situations. In addition, targets are re-initialized with new identities after a tracking loss. As more frames are processed, the number of identity labels tends to increase without bound as tracking losses occur along the frames.

Therefore, considering the superior performance of the Führ-Jung method relatively to the others in the literature when applied to our problem, it was used for a comparative assessment of the proposed method. Among the examples shown in Subsection 5.3, the two that comprise two-person experiments were chosen for this purpose.

Since the Führ-Jung method was designed to be used with a single camera, the right camera was used in the experiments, because it supplies the best view. Camera switching was not performed for the Führ-Jung method. To use more camera views with this method, it would result in simulations of a modification over the Führ-Jung method, not using the original one. In counterpart, camera switching is inherent to the system proposed here.

For the second example shown above (Fig. 7), the Führ-Jung method stopped tracking in the beginning of the simulation. Both persons use clothes of light color, quite similar to the floor. The Führ-Jung method initially detected them, however these were not validated by the background removal stage, and were thus eliminated. It is worth explaining that the Führ-Jung method makes use of ViBE (Barnich and Droogenbroeck 2011) for the background removal, while the proposed system uses GMM followed by mathematical morphological operations
(see Subsection 4.1). These, among other improvements, fill the holes within the foregrounds, while ViBE does not. Due to the color similarity between person’s clothes and the background, the ViBE method in Führ-Jung led to poor foregrounds. This, in turn, resulted in foreground-to-bounding boxes area ratios below the minimum limit considered for detection validation.

Fig. 9 shows results obtained for the third example discussed above. The Führ-Jung method lost tracking many times during the scene, due to occlusions. It also lost tracking even in nonocclusion situations. Whenever the method recovered tracking, it initialized the lost targets with new identities, which led to poor radiation dose estimates. However, for the results in Fig. 9, the re-identification was performed manually for Führ-Jung, in order to allow a fairer comparison of the performances of the two methods.

5.5. Discussion

The experimental results indicate that the developed system is able to adequately assess the radiation doses received by nuclear plant personnel while executing their tasks.

For each person, in each case, the doses estimated by the developed system are equivalent to the ones of the GT when one considers the measurement uncertainties. This is so because, most of the time, as can be observed in Figs. 6 to 9, the dose cones due to the intrinsic measurement error have strong intersections (an exception is Target A in the third example above). Furthermore, as can be noticed especially in Figs. 7a, 7b and 8b, the doses estimated by the proposed method do not present a bias relative to GT, as both curves cross each other.

It is important to note that the Führ-Jung results were shown in Fig. 9 just for providing a reference to the performance of the proposed method. The comparisons are not strictly fair
because, on one hand, the person re-identification in Führ-Jung is performed manually, which gives it an advantage in occlusion situations. On the other hand, the Führ-Jung method does not work with camera switching, which gives it a disadvantage in occlusion handling.

A summary of the performances of the proposed method and Führ-Jung for accumulated radiation doses estimation is provided in Table 1. The doses are given in uSv. The corresponding GTs are also shown.

Table 1. Final doses (in uSv) for the third example.

The deviation between estimated and GT dose values shown in Fig. 8a was due to severe occlusion occurring with Target A relatively to both cameras, in this part of the scene. This can be solved in the future by using more camera views adequately.

Relatively to the processing time, the proposed system currently runs at approximately 16.5 frames per minute, meaning a processing time of approximately 3.63 s per frame. This was measured in a personal computer with the following characteristics: Intel Core i7 CPU 960 3.20 GHz, Quadcore. The operations that run in Argonauta take some minutes, typically up to 5 min each, meaning around 9,000 frames. Considering the processing time of each frame given above, this would mean approximately 9:06 h, per camera. At most, two operations are run in Argonauta each day. Taking into account that more and faster computers can be used for processing, even not considering parallel processing possibilities, the system can be implemented. Experiments can be processed during the night after being recorded, giving the results the day after. This is an improvement over what is performed currently with the traditional radiological protection means.

6. Conclusions and perspectives

In this paper we have proposed a method for radiation dose estimation in nuclear plants based on visual tracking-by-detection. In order to develop it, we have built a video database comprising several situations that are typical of the work performed in a nuclear plant. This database was made publicly available (Jorge et al. 2016). The developed system has shown
good dose estimation performance, equivalent to the GT when the measurement errors of the
radiation monitors used for estimating the radiation distribution inside the region are
considered.

Results show that occlusions may be solved by switching views from one camera to the other.

The overall results are encouraging. The system is easy to deploy since all it needs is a
previous mapping of the radiation in the environment and off-the-shelf video cameras. The
system can be improved to work in real-time, although it may contribute to radiological
protection even operating offline. Doses may be estimated just after the tasks are executed.
This represents significant gains in comparison to the current radiation dose estimation
procedures based in personal radiation monitors.

Acknowledgements

This research was sponsored by FAPERJ and CNPq, Brazil. The authors would also like to
thank IEN, CNEN in Brazil, for the support for the experimental runs.

References

Adam, A., Rivlin, E., Shimshoni, I., 2006. “Robust fragments-based tracking using the
integral histogram”. IEEE Conference on Computer Vision and Pattern Recognition (CVPR),
June 17-22, V. 1, pp. 798-805.

1709-1724.

Battacharyya, A., 1943. “On measure of divergence between two statistical populations
defined by their probability distributions”. Bulletin of Calcutta Mathematical Association, V.

Computer Vision and Image Understanding, V. 110, N. 3, pp. 346-359.
*IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3457-3464.

http://www.vision.caltech.edu/bouguetj.

Bradski, G.R., 1998. “Real time face and object tracking as a component of a perceptual user 

multiperson tracking-by-detection from a single, uncalibrated camera”. *IEEE Transactions on 
Pattern Analysis and Machine Intelligence*, V. 33, N. 9, pp. 1820-1833.


with applications to image analysis and automated cartography”. *Communications of the 


809-830.


INITIALIZATION
Detecting people in the scene
- People detection, combined with
  Background removal

RE-IDENTIFYING PEOPLE
Approaches
- Distance in WCS (<2 m)
- Parameter extraction
  - Face/head ROI colors
  - Bodies ROI SURF keypoints

MULTI-LEVEL DATA ASSOCIATION
Tracking-by-Detection along frames
- Parameter matchings of people with ROIs in
  previous frames
- Voting scheme to solve for people identities

FURTHER RE-IDENTIFICATION
Solving Re-identification mismatches
- Re-identifying people not identified due to
  occlusions
- Recovering identities of lost targets

CAMERA SWITCHING
Switch Cameras in Case of
- Person-to-person occlusions
- Occlusion between person and
  objects (in pre-defined regions)
Figure 6
Click here to download high resolution image
<table>
<thead>
<tr>
<th>Method</th>
<th>Target A</th>
<th>Target B</th>
</tr>
</thead>
<tbody>
<tr>
<td>GT</td>
<td>0.039</td>
<td>0.036</td>
</tr>
<tr>
<td>Fuhr-Jung</td>
<td>0.022</td>
<td>0.023</td>
</tr>
<tr>
<td>Proposed</td>
<td>0.022</td>
<td>0.032</td>
</tr>
</tbody>
</table>