RECEIVED RADIATION DOSE ASSESSMENT FOR NUCLEAR PLANTS' PERSONNEL BY VIDEO-BASED SURVEILLANCE

Carlos Alexandre Fructuoso Jorge

Tese de Doutorado apresentada ao Programa de Pós-graduação em Engenharia Elétrica, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia Elétrica.

Orientadores: José Manoel de Seixas
Eduardo Antônio Barros da Silva
Antônio Carlos de Abreu Mól

Rio de Janeiro
Junho de 2015

RECEIVED RADIATION DOSE ASSESSMENT FOR NUCLEAR PLANTS' PERSONNEL BY VIDEO-BASED SURVEILLANCE

Carlos Alexandre Fructuoso Jorge

TESE SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA (COPPE) DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS EM ENGENHARIA ELÉTRICA.

Examinada por:

Prof. José Manoel de Seixas, D.Sc.

Prof. Eduardo Antônio Barros da Silva, Ph.D.

Dr. Antônio Carlos de Abreu Mól, D.Sc.

Prof. Cláudio Rosito Jung, D.Sc.

Prof. José Gabriel Rodriguez Carneiro Gomes, Ph.D.

Dr. Cláudio Márcio do Nascimento Abreu Pereira, D.Sc.

RIO DE JANEIRO, RJ – BRASIL
JUNHO DE 2015

Jorge, Carlos Alexandre Fructuoso

Received radiation dose assessment for nuclear plants' personnel by video-based surveillance/Carlos Alexandre Fructuoso Jorge. – Rio de Janeiro: UFRJ/COPPE, 2015.

XX, 163 p.: il.; 29, 7cm.

Orientadores: José Manoel de Seixas
Eduardo Antônio Barros da Silva
Antônio Carlos de Abreu Mól

Tese (doutorado) – UFRJ/COPPE/Programa de Engenharia Elétrica, 2015.

Referências Bibliográficas: p. 139 – 159.

1. Radiological Protection. 2. Safety and Security in Nuclear Plants. 3. Video-based Surveillance. 4. Multi-target Detection and Tracking. 5. Target Re-identification. I. Seixas, José Manoel de *et al.* II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia Elétrica. III. Título.

A Deus.

À minha família; em particular:
à minha esposa Simone;
à minha mãe Maria Marcolina,
pela superação dos graves
problemas de saúde,
o que lhe permitiu ver
a conclusão desta tese;
à minha tia Maria da Glória
(in memorian),
que partiu desta vida durante o
desenvolvimento da tese (+2011);
e ao meu afilhado Lucas,
*nascido durante a tese (*2012)!*

Ao dr. Luiz Osório B. Aghina
(in memorian), cuja carreira se
confunde com a história do
Instituto de Engenharia Nuclear:
desde o início da década de 1960,
até recentemente (+2015).

Agradecimentos

A Deus, que me chamou à vida e me sustenta nesta caminhada; em particular, nesta última difícil tarefa. Às intercessões da Virgem Maria e de São José.

Aos meus pais, pela vida. À minha mãe, pelos valores transmitidos e incentivo aos estudos desde a infância. À minha esposa, presente a esta jornada desde o início do curso. De ambas, o incentivo e o apoio incondicionais, acreditando que eu poderia seguir adiante, quando eu achava que não conseguiria. À memória de minha avó Elza, que sempre se alegrava com minhas conquistas, até o início do mestrado... À memória de meu avô Antelmo, que não cheguei a conhecer, mas cuja retidão moral sempre me serviu de exemplo *"(...)eterna será a memória do justo"*, *(Sl 111,6).*

Aos meus orientadores: profs. José Manoel de Seixas, Eduardo Antônio Barros da Silva e Antônio Carlos de Abreu Mól, que me guiaram e incentivaram. Ter mais de um orientador enriquece a nossa formação acadêmica.

Aos outros professores membros da Banca: profs. José Gabriel Rodriguez Carneiro Gomes, Cláudio Jung e Cláudio Márcio do Nascimento Abreu Pereira, por aceitarem o convite, e pelas sugestões que tornaram esta tese mais clara e completa.

A todos os professores por quem passei durante os cursos das disciplinas: além dos profs. Seixas e Eduardo, também aos profs. Luiz P. Calôba, Mariane R. Petraglia, José Gabriel R. C. Gomes e Alexandre P. A. Silva.

À Universidade Federal do Rio de Janeiro, pela oportunidade de retornar a esta casa, e buscar galgar mais um degrau na formação acadêmica.

À Comissão Nacional de Energia Nuclear/Instituto de Engenharia Nuclear, pela liberação parcial a fim de permitir o desenvolvimento desta pesquisa de doutorado, em um tema proposto pela própria Instituição de origem, e que encontrou acolhida na Universidade.

Ao servidor dr. Maurício Alves Aghina, por sugestões devido à sua experiência com projeto de instrumentação para a área nuclear, incluindo monitores de radiação que foram usados nesta tese.

Ao servidor dr. Francisco José Ferreira, do Reator Argonauta, Instituto de Engenharia Nuclear, pela colaboração para a realização da parte experimental, fundamental para esta pesquisa; também por ter me esclarecido quanto às operações típicas realizadas no Argonauta, para a elaboração dos roteiros de filmagens. Ao servidor João Carlos Pereira. Ao servidor dr. Dante L. Voi, pesquisador do Argonauta, por me explicar suas rotinas de experimentos de espectrometria, e ter se disposto a participar das filmagens, reproduzindo toda a movimentação que costuma realizar durante a execução de seus trabalhos!

Ao servidor Victor Lassance Cunha, que se predispôs a executar modelagem e simulações computacionais de distribuição de taxa de dose no salão do Argonauta, usando o método Monte Carlo. Faço votos por seu pronto restabelecimento.

Aos servidores do Serviço de Proteção Radiológica deste Instituto, que orientaram e acompanharam o levantamento de medidas de taxas de dose de radiação: Eara Oliveira e Oswaldir Santos. E àqueles que operaram o Reator durante este levantamento.

Aos bolsistas de Iniciação Científica que cooperaram diretamente com esta pesquisa; dentre os quais destaco, dos mais recentes para os mais antigos: Gabriel R. Lira, José Guilherme T. Monteiro, Fábio Waintraub, Raphael Cota, Bruno L. Ramos; e outros bolsistas que tenham eventualmente ajudado em alguma tarefa. Ao bolsista graduado Daniel R. Chelles, pela participação nas filmagens para a geração da base de dados de vídeo. Ainda ao bolsista José, pela colaboração no projeto e na compra de dispositivos eletrônicos para o circuito de *flash* para sincronização dos vídeos; por se dispor a ajudar na compra das fitas, e na colocação das mesmas para a marcação do chão do salão do Argonauta. Ao bolsista Gabriel, pela árdua tarefa de realizar as marcações manuais de *ground truth*. A ambos, por sua dedicação na "reta final" desta tese.

Ao Conselho Nacional de Pesquisa e Desenvolvimento Tecnológico e à Comissão Nacional de Energia Nuclear, pelas bolsas concedidas a estes bolsistas graduandos ou graduados.

Aos servidores Douglas S. Sales, Roberto Cláudio Nunes e Valério Martins, pela colaboração no projeto e/ou montagem do circuito de sincronização.

Ao servidor dr. Marcel Waintraub, do setor de Informática, pelo suporte com as câmeras usadas nesta pesquisa. À servidora Edione Vasconcelos, por tê-las cedido temporariamente. Ao servidor Valdeci Maurílio, dentre outros, pelo auxílio na sua instalação e desinstalação. À servidora Mônica Vianna, pelo auxílio na disponibilização de plantas e medidas do Argonauta.

A outros servidores que eventualmente tenham colaborado de alguma forma à pesquisa.

Aos colegas de turma da COPPE/Elétrica, à época das disciplinas, pelas vezes em que nos reunimos para estudar.

Aos colegas de curso que me ajudaram cedendo códigos, ou partilhando dicas: Mateus T. Nakahata, Allan F. Silva, Lucas A. Thomaz, dr. Andreas Ellmauthaler, Jonathan N. Góis, Felipe M. Ribeiro, dr. Gabriel M. Araújo, Camila Gussen, Anderson Oliveira,... Aos três primeiros, por cederem seu código para auxílio à marcação manual de *ground truth*. Ao Andreas e ao Jonathan, pelas dicas sobre calibração de câmeras. Sou grato aos colegas que também me ajudaram com o uso do TeX - em particular, ao Lucas, que me preparou um modelo para imediato uso e adaptação. Ao pós-doutorando José Fernando Oliveira, pelas conversas e dicas. Ao dr. Natanael N. Moura, pelas dicas e incentivo.

À Daniele, da Secretaria do Programa de Engenharia Elétrica, pela ajuda em diversos momentos. À Ana Cristina e à Isabene, pelo suporte. Ao Marcos de Paula, pela boa vontade com a revisão complementar do inglês.

Fazer agradecimentos é uma tarefa arriscada, uma vez que corre-se o risco de omitir o nome de alguém... Porém, não acho que por isso deva deixar de citar aqueles que estiveram mais próximos. A partir destes, estendo o agradecimento a todas as outras pessoas – nos âmbitos profissional ou pessoal – que de alguma forma incentivaram ou colaboraram com esta pesquisa.

"Narram os céus a glória de Deus, e o firmamento anuncia a obra de suas mãos." (Sl 18,2)

"A fé e a razão (fides et ratio) constituem como que as duas asas pelas quais o espírito humano se eleva para a contemplação da verdade. (...)", São João Paulo II, Carta Encíclica Fides et Ratio (FR), 1998.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

AVALIAÇÃO DE DOSE DE RADIAÇÃO RECEBIDA POR TRABALHADORES DE PLANTAS NUCLEARES BASEADO EM VIGILÂNCIA POR VÍDEO

Carlos Alexandre Fructuoso Jorge

Junho/2015

Orientadores: José Manoel de Seixas
Eduardo Antônio Barros da Silva
Antônio Carlos de Abreu Mól

Programa: Engenharia Elétrica

Este trabalho propõe o desenvolvimento de um sistema para avaliação de dose de radiação recebida por trabalhadores em plantas nucleares. O sistema é concebido para operar de forma complementar às abordagens existentes para proteção radiológica, de modo a oferecer redundância, o que é desejável em operações de plantas críticas. O sistema proposto deve operar de forma indepedente das ações a serem tomadas pelos operadores sob avaliação. Para tanto, foi decidido que seria baseado em métodos usados em vigilância por vídeo. A planta nuclear usada como exemplo é o Reator Nuclear de Pesquisa Argonauta, do Instituto de Engenharia Nuclear, Comissão Nacional de Energia Nuclear. Durante a pesquisa desta tese, foram geradas bases de dados tanto de distribuição de taxa de dose, como de vídeos. Métodos disponíveis na literatura, para detecção e/ou rastreamento de alvos, foram avaliados para esta base. A partir destes resultados, um novo sistema foi proposto, com o objetivo de atender aos requisitos desta aplicação em particular. Dadas as posições rastreadas de cada trabalhador, a dose de radiação por eles recebida durante a execução de tarefas operacionais é estimada, podendo servir como parte de um sistema de apoio a decisão.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

RECEIVED RADIATION DOSE ASSESSMENT FOR NUCLEAR PLANTS' PERSONNEL BY VIDEO-BASED SURVEILLANCE

Carlos Alexandre Fructuoso Jorge

June/2015

Advisors: José Manoel de Seixas
Eduardo Antônio Barros da Silva
Antônio Carlos de Abreu Mól

Department: Electrical Engineering

This work proposes the development of a system to evaluate received radiation dose for nuclear plants personnel. The system is conceived to operate in a complementary form to the existing approaches for radiological protection, thus offering redundancy, what is desirable for critical plants operation. The proposed system must operate in an independent form on the actions to be performed by the operators under evaluation. Therefore, it was decided it would be based on methods used for video surveillance. The nuclear plant used as example is Argonauta Nuclear Research Reactor, belonging to *Instituto de Engenharia Nuclear*, *Comissão Nacional de Energia Nuclear* (Nuclear Engineering Institute, National Nuclear Energy Commission). During this thesis research, both radiation dose rate distribution and video databases were obtained. Methods available in the literature, for targets detection and/or tracking, were evaluated for this database. From these results, a new system was proposed, with the purpose of meeting the requisites for this particular application. Given the tracked positions of each worker, the radiation dose received by each one during tasks execution is estimated, and may serve as part of a decision support system.

Contents

List of Figures

List of Abbreviations

2D	Two-dimensional, p. 36
3D	Three-dimensional, p. 35
ALARA	As Low as Reasonably Achievable, p. 2
Argonauta	Argonauta Nuclear Research Reactor, p. 2
BG	Background radiation, p. 62
BSS	Blind Signal Source Separation, p. 32
CCD	Charge-coupled Device, p. 82
CDTN	*Centro de Desenvolvimento da Tecnologia Nuclear*, p. 9
CERN	European Organization for Nuclear Research, p. 25
CNEN	*Comissão Nacional de Energia Nuclear* - National Nuclear Energy Commission, p. 2
CPD	Coherent Patch Displacement, p. 38
CPMD-RPT	Combining Patch Matching and Detection for Robust Pedestrian Tracking, p. 38
CSPR	*Coordenação de Segurança e Proteção Radiológica* - IEN's safety and radiological protection coordination, p. 2
Camshift	Continuously Adaptive Mean-Shift, p. 35
DES	Double Exponential Smoothing, p. 38
DLT	Direct Linear Transformation, p. 79
DP+NMS	Dynamic Programming - Non-max Suppression, p. 39
EMC-EMI	Electromagnetic Compatibility - Electromagnetic Interference, p. 25

EMD	Earth's Mover Distance, p. 37
EM	Expectation-Maximization, p. 51
FPDW	Fastest Pedestrian Detector in the West, p. 37
Führ-Jung	The same as CPMD-RPT, p. 38
FragTrack	Fragments-based Tracking, p. 37
GMM	Gaussian Mixture Model, p. 32
GPS	Global Positioning System, p. 24
GPU	Graphical Processing Units, p. 136
GSM	Global System for Mobile Communication, p. 25
GT	ground truth, p. 6
HD	High definition, p. 71
HMD	Head Mounted Display, p. 23
HOG	Histogram of Oriented Gradients, p. 36
HSV	Hue-Saturation-Value color space, p. 43
Halden	Halden Virtual Reality Centre, p. 22
IAEA	International Atomic Energy Agency, p. 16
IAE	*Instituto de Energia Atômica*, p. 9
IBERINCO	IBERDROLA Engineering and Construction, p. 24
ICA	Independent Component Analysis, p. 33
ICRP	International Commission on Radiological Protection, p. 16
ICRU	International Commission on Radiation Units and Measurements, p. 16
IC	Integrated Circuit, p. 73
ID	Identity, p. 7
IEN	*Instituto de Engenharia Nuclear* - Nuclear Engineering Institute, p. 2

IME	*Instituto Militar de Engenharia*, p. 13
IPCA	Incremental PCA, p. 32
IPEN	*Instituto de Pesquisas Energéticas e Nucleares*, p. 9
IPR	*Instituto de Pesquisas Radioativas*, p. 9
IRD	*Instituto de Radioproteção e Dosimetria*, p. 17
IVT	Incremental Learning for Visual Tracking, p. 40
J9	J9 output radiation channel, p. 10
JAEA	Japan Atomic Energy Agency, p. 23
JAERI	Japan Atomic Energy Research Institute, p. 22
JNC	Japan Nuclear Cycle Development Institute, p. 22
KDE	Kernel Density Estimation, p. 32
KLT	Kanade-Lucas-Tomasi, p. 35
KL	Kanade-Lucas, p. 35
LoG	Laplacian of Gaussians, p. 36
MCMC	Markov-chain Monte Carlo, p. 136
MILTrack	Multiple Instance Learning, p. 39
MPEG	Moving Picture Experts Group, p. 71
MS	Mean-Shift, p. 35
NCC	Normalized cross-correlation, p. 48
ND	no data, p. 12
NN 3.01	*Diretrizes Básicas de Proteção Radiológica*, p. 16
NN	Nearest-neighbour, p. 49
NTSC	National Television System Committee, p. 71
OpenCV	Open Source Computer Vision, p. 90
PCA	Principal Component Analysis, p. 32

Pfinder	Person Finder, p. 31
RANSAC	Random Sample Consensus, p. 79
RFID	Radio Frequency Identification, p. 23
RGB	Red-Green-Blue color space, p. 43
ROI	Region of Interest, p. 31
RU	Research Unit, p. 2
SCK•CEN	Belgian Nuclear Research Centre, p. 24
SIFT	Scale Invariant Feature Transform, p. 36
SURF	Speed-Up Robust Features, p. 37
SXGA	Super Extended Graphics Array, p. 71
TEPKO	Tokyio Electric Power Co., Inc., p. 23
TEPSYS	TEPKO Systems Coorporation, p. 23
TLD	Thermoluminiscent Detector, p. 17
TLD	Tracking-Learning-Detection, p. 39
UFRJ	*Universidade Federal do Rio de Janeiro* - Federal University of Rio de Janeiro, p. 13
VJ	Viola-Jones, p. 35
VTD	Visual Tracking Decomposition, p. 39
ViBE	Visual Background Extractor, p. 32
W4	Who? When? Where? What?, p. 31
WCS	World Coordinate System, p. 4
WVMF	Weighted Vector Median Filter, p. 38
fps	frames per second, p. 7
led	lighting-emitting diode, p. 73
pixel	picture element, p. 31

Chapter 1

Introduction

Nuclear power benefits society in many ways. Probably, the most known application is nuclear-based electric power generation. But besides this, there is a number of other applications, among which some can be highlighted:

- Nuclear medicine – of great importance for both diagnosis and treatment of many diseases;
- Agriculture and food conservation;
- Radiotracers for industrial use - as for monitoring the efficiency of industrial effluents separation, or the quality of industrial mixtures;
- Multiphase flow measurements - that finds important use in the nuclear and petrochemical industries;
- Non-destructive evaluation of diverse objects and equipment - to detect internal fissures, for example;
- Materials processing to improve their strength or other characteristics; and
- Books and arts conservation – in the former case, an advantage over traditional chemical-based treatments.

Thus, nuclear power usage is present in the society in a diversity of situations and applications. It requires, however, that the involved personnel follow safety procedures, in order to protect themselves and the general public.

1.1 Motivation

Working with nuclear materials or radioactive sources involves, unavoidably, exposure to radiation, mainly for personnel, which deal with it frequently - and, sometimes, almost daily. The goal is, thus, to minimize the received dose to what is strictly needed, following the As Low as Reasonably Achievable (ALARA) principle [1]; its application involves three recommendations:

- Minimizing elapsed time to radiation exposure;
- Increasing distance to the radiation sources;
- The possible use of shielding.

Important means for ALARA implementation is radiation monitoring - both the ambient radiation dose rate distribution and the radiation dose received by each worker. These are performed by radiation monitors, both fixed and portable - for the ambient dose measurement -, and personal/individual[1] ones - for individual radiation protection.

In this context, this thesis was developed as a contribution to radiological protection, aiming to improve safety for nuclear plants personnel, in a redundant form to the approaches currently in use, and in an independent form - not depending on workers' behaviour. Another important characteristic of the developed system is that it can detect personnel standing still in a place subject to high radiation dose rate, and may generate an alarm.

Another motivation is nuclear security. Of course, video-based surveillance is of great importance for this security. Thus, this research can bring insight into this field, for future usage. For instance, the developed system contributes readily for security, as long as it detects people entering a nuclear plant's room.

1.2 Overview

The experimental part of this thesis was carried out at Argonauta Nuclear Research Reactor (Argonauta), at *Instituto de Engenharia Nuclear* (IEN) [2], *Comissão Nacional de Energia Nuclear* (CNEN) [3] - respectively Nuclear Engineering Institute, and National Nuclear Energy Commission. IEN is one Research Unit (RU) of CNEN.

During Argonauta operations, personnel from *Coordenação de Segurança e Proteção Radiológica* (CSPR) - the safety and radiological protection coordination of IEN -, assess radiation dose rate levels at some key locations within Argonauta's Room (and sometimes also at some other locations outside it), generating radiation dose rate distribution databases. These distributions are rather scarce, serving for current radiological protection routines, but insufficient for more detailed research.

[1]Personal and individual monitors might seem to mean the same. But there is a type of monitor that is really personal - labelled with users' names -, while there are other types which are for individual use, but not labelled: anyone may use them, by registering just before and after use. For simplicity, though, they may be referred to as as individual or personal monitors, interchangeably, in this thesis.

Our staff has been carrying out developments related to virtual simulations, to support training and planning in critical plants. The application most related to this thesis is named Virtual Argonauta, in which this Reactor had been virtually modelled, considering both architectural and dose rate databases, to support monitoring, personnel training and activities planning. The advantages of this approach are that they can reduce elapsed times, and/or optimize personnel's locations within the nuclear plant, before they enter it to execute their tasks; in summary, they enable better tasks and paths planning, thus reducing their received radiation dose. Readers may refer to [4–12] for more information on the historical evolution of IEN developments - among which some may be picked up as very representative ones [6], [8], [11]. Relatively to nuclear security, refer to [12–14], for some details on IEN developments.

1.3 Research Objectives

This thesis was planned for developing a system to monitor and assess received radiation dose for Argonauta personnel, while they execute operational and/or maintenance tasks, aiming to contribute for improving ALARA principle implementation. The proposed system must detect one or more workers entering Argonauta's Room, and track them, accounting for the radiation doses received by each one, even when they are not using individual radiation monitors, for any reason. The doses are estimated by persons' detected and tracked positions within the environment, by using computer vision-based methods. It may be used to compose a decision support system, in the following ways, to mention some typical situations:

- First of all, Argonauta and/or CSPR heads can be alerted if anyone enters Argonauta's Room - specially when there is no scheduled operation, maintenance or visits;
- During scheduled tasks, Argonauta and/or CSPR heads can decide whether a worker can or can not continue doing his tasks, or must be substituted by another one, according to the accumulated received dose by him;
- Argonauta and/or CSPR heads can alert any worker that remains unintentionally standing still in a "hot area"[2], while executing his task;
- In case of failure of any personal/individual radiation detector, the developed system can still estimate the received radiation dose for the respective worker;

[2] Hot area: region with not negligible radiation dose rate level; typically, with higher dose rate levels comparatively to other locations in the environment.

- Also, in case a worker or a visitor forgets using individual monitor, the system will be able to estimate the received radiation dose for him;
- Tracking workers' trajectories may be used to improve the radiation dose level database, as these values are saved together with the corresponding positions.[3];
- Important to mention: the system is planned to operate concurrently and in a complementary form to the existing approaches for radiological protection, therefore supplying a redundancy that is important for critical plants operation, such as nuclear ones;
- The approaches currently in use for individual radiation monitoring, either take some weeks to give responses, or have low accuracy; thus, this developed system can fill this gap by supplying faster responses.

The first item contributes to improve security, while the other ones contribute to improve safety for personnel.

To achieve these objectives, two video cameras were installed within Argonauta's Room. The processing is based on monocular vision - the cameras were not intended to operate automatically in conjunction, simultaneously, for two-view processing. The cameras were installed in such a way they could cover a region within the Argonauta's Room that presents higher dose rate levels during operation. They were installed approximately - but not exactly - pointing towards each other, to cover possible regions that might not be covered by any one of them.

Since the architectural data of Argonauta is available, it is possible to estimate each worker's position in the world coordinate system (WCS) - more specifically, in the floor -, by means of projective geometry [15]. Given these estimated positions and the dose rate distribution mapping, it is possible to estimate the doses received by each worker during the elapsed times they stay within this environment. Alternatively, the dose rate at each one's position can be monitored.

The nuclear plant used as example in this thesis - Argonauta - is a research nuclear reactor, which produces lower power levels comparatively to power nuclear reactors for electricity generation; but is comparable to many similar research nuclear plants. Therefore, this system can be ported these other nuclear or radioactive plants[4] - as cyclotron accelerators -, or laboratories where nuclear materials are

[3]This topic will be explained later, in Chapter 3, Section 3.3.

[4]Definition of nuclear plant: one where there is nuclear - fissile - materials, such as Uranium-based nuclear fuels - nuclear reactors are within this type of plants. Definition of radioactive plant: one in which there are no nuclear sources, but radiation is rather produced temporarily by accelerating particles - accelerators and X-ray equipment fall into this category -; although in these latter ones radiation is produced temporarily, it is common that some parts of these plants remain radioactively activated after operation.

dealt with[5].

1.4 Thesis Organization

This thesis is organized as follows:

Chapter 2 describes Argonauta and related topics, to contextualize this research. Section 2.1 comments on some of its characteristics - both historical and operational -, and also about its physical environment. Section 2.2 shows and briefly comments on the measurements routinely taken by CSPR during operations. Section 2.3 briefly comments on typical applications of this Reactor.

Chapter 3 reviews the problem of safety in nuclear plants. Section 3.1 treats this nuclear-related subject focusing in the more traditional, well established, approaches adopted for radiological protection - including those currently in use in Brazil and, more specifically, at Argonauta. This section also mentions some guidelines that must be followed to achieve radiological protection in nuclear plants. Section 3.2, in turn, treats more recent approaches that make use of computer-based simulations to support training, monitoring and planning for the nuclear field. Different groups have been adopting these approaches; IEN staff related achievements are also commented. Section 3.3 gives remarks about this chapter.

Chapter 4 reviews topics from the computer vision field that are of interest for this thesis. In Section 4.1, some methods for targets segmentation based on background modelling and removal are commented - and this is done by following approximately their historical development. Section 4.2 comments, in a general form, on methods for targets detection and/or tracking in video. Section 4.3 finishes this chapter with remarks, mentioning the methods then chosen for further evaluation.

Chapter 5 describes the chosen methods in more details. Each Section, from 5.2 to 5.8, describes one of them. Section 5.9 gives some remarks about each method and some possible combinations among them.

Chapter 6 describes the databases generation. Dose rate database generation is treated in Section 6.1. Subsection 6.1.1 comments on measurements carried out in Argonauta's Room, the procedures and equipment used, and shows resulting radiation dose rate distribution. Subsection 6.1.2 gives some comments on alternative approaches for dose rate calculation. The video database generation is treated in Section 6.2. Subsection 6.2.1 comments somewhat the operational routines usually carried out in Argonauta, which guided the screenplay to be executed within this Reactor's Room to evaluate the computer-vision based methods selected for use, and also comments on some possible problems that might arise with this specific

[5] There are other research nuclear reactors, cyclotrons and many laboratories operational at CNEN RUs.

video database. Subsection 6.2.2 explains the setup, comments on the equipment used, and about cameras installation and the video acquisition process. Section 6.3 finishes this chapter with remarks.

Chapter 7 treats projective geometry and its application to the video database. Section 7.1 explains basic concepts of projective geometry. Subsection 7.1.1 explains the cameras calibration, while Subsection 7.1.2 explains the projection matrices computation. Section 7.2 explains how ground truths (GT) were obtained. Section 7.3 gives remarks about all these processing.

Chapter 8 shows results and developments carried out to achieve this thesis objective. Section 8.1 shows results obtained with the previously selected methods. Then, Section 8.2 explains the proposed and developed system, with the strategies and heuristics adopted, mainly those used for targets re-identification - data association - along the videos, with the purpose of estimating correctly the radiation doses received by each one worker. The system is detailed from Subections 8.2.1 to 8.2.5. Section 8.3 closes this Chapter with remarks.

Chapter 9 shows the results achieved with the developed system. Section 9.1 shows the final results thus obtained and encompasses discussion about them. Section 9.2 concludes the achievements with some remarks.

Chapter 10 concludes this thesis. Section 10.1 gives the conclusions, while Suggestions for possible future works are given in Section 10.2.

1.5 Contributions

The contributions of this thesis are:

- Development of a system for estimating the received radiation dose by each worker in a nuclear plant, during operational and/or maintenance tasks, by video-based surveillance approach (specifically, by using computer vision-based methods for detecting and tracking) by re-identification/data association of people from videos taken within a nuclear plant. The approach adopted was markerless, to result in an independent form of tracking, meaning not depending on workers' actions - as otherwise by using markers, head-up apparatus, and similar means. To the extent of the knowledge of the author (based on the bibliographical review performed in a period during the development of this thesis), there is no other system proposal such as this; although some authors mention tracking in a general form, or by other means or for other purposes.

- The computer vision-based methods chosen to be evaluated - among some belonging to the state-of-the-art -, did not solve completely, each one alone, the specific problem in hand, failing to track or to re-identify persons, due to this particular video database particularities.

- Some strategies were then proposed to try solving persons re-identification:

  - First, trajectory-based data association was used as much as possible, when persons were sufficiently distant among themselves, in the WCS.

  - Another strategy was to discriminate among persons from color-based features extracted from their faces/heads positions. As long as discriminating persons from their faces/heads in pose-free variation conditions (in front, side or back views), is a quite difficult task, comparisons were planned to be made as much as possible from frame to frame. The reason for this is that no-one changes pose so abruptly between adjacent frames, in a frame rate of 30 frames per second (fps).

  - Another further strategy was to discriminate among persons from interest-points extracted from their bodies regions, and also trying to match them from frame to frame as much as possible, also due to the fact that the chosen parameters are very discriminative and present low variation between subsequent frames.

  - For both of the previously mentioned approaches, comparisons were also made considering non-sequential frames, for the cases where the system loses someone's detection or tracking, re-initializing it later. The reason for that was the need to achieve more robustness in re-identification, for the cases where a person's Identity (ID) changed only from the immediate or some few previous frames to the current one, against the previous ID history. This might also aid to re-identify persons when they get out and return to the environment in different poses.

  - Other strategies adopted made use of heuristics specific to the problem in hand, since people can only appear (targets birth) or disappear (targets death) at specific places, thus minimizing the false alarms and/or missed targets. This was done in a post-processing stage.

- The system was developed to operate concurrently and in a complimentary form to the existing approaches for radiological protection. As these latter do not supply much precise results, or the more precise ones usually take some weeks, the proposed system can fill this gap with faster results. But one should not discard completely the results obtained by the other means, as redundancy is important for critical systems as nuclear plants.

- Generation of a video database specific for the problem in hand, which presents some particularities as: (i) the use of same color uniforms; (ii) people standing still for long time very near to each others; and (iii) people crouching to execute some tasks. General-purpose video databases are available in the Internet, and methods have been developed by using them, but those did not seem to match the needs for the current problem. The generated video database can be used in the future for further evaluations and developments, and can also be extended, mainly when the system is operational, by recording more real operational and maintenance tasks.

- Generation of GT for (part of) the video database[6].
- Generation of a dose rate distribution database, which can be improved in the future by different means.

[6]Obtaining GTs is a hard and tedious task, and it demands long working times (or a great number of people dedicated to it). It can be complemented in the future.

Chapter 2

Argonauta

2.1 Main Characteristics and Environment

Argonauta, of IEN, CNEN, is dedicated to a diversity of research, development and teaching activities. It is a pool type reactor, with typical and maximum operation power levels of 500 W and 5 kW, respectively [16], [17]; the former limit is the licensed one, while the latter is the maximum enabled. But typical operation levels are: 1.7 W, 3.4 W, 17 W, 34 W, 85 W, 170 W and 340 W [17]. These are far below the typical power levels achieved by nuclear power plants for electricity generation - take as examples the Brazilian ones, named Angra 1 and Angra 2, with operating power levels of 640 MW and 1,350 MW, respectively [18].

IEN was founded in 1962, with the objective of having the first nuclear research reactor of Rio de Janeiro State - Argonauta -, following the tendency at that time for constructing and installating nuclear research reactors in Brazil. Previously to Argonauta, one had been installed in Minas Gerais State - the TRIGA Nuclear Research Reactor -, at *Instituto de Pesquisas Radioativas* (IPR), now renamed to *Centro de Desenvolvimento da Tecnologia Nuclear* (CDTN) [19], another CNEN RU. Another one had also been installed in São Paulo State - the IEA-R1 Nuclear Research Reactor -, at *Instituto de Energia Atômica* (IEA), now renamed to *Instituto de Pesquisas Energéticas e Nucleares* (IPEN) [20], also another CNEN RU.

Argonauta achieved his first criticality[1] in 1965, and has since then been using for the already mentioned activities.

It occupies an area comprised by thick concrete walls that serve as shielding - besides the actual shielding of the Reactor itself. Its Room is classified as a controlled area. Its Control Room and another room near its entrance are considered as supervised areas [2].

Argonauta has some different output radiation channels, but currently, the nor-

[1] Criticality: operational condition, achieved by controlled fission.

[2] The definitions of controlled, supervised and free areas are given in Chapter 3, Section 3.1.1.

mally used is the one named J9 output radiation channel (J9), which outputs neutron radiation from the Reactor's core during operation. Figure 2.1 shows a photographic view of Argonauta within its room, with its J9 channel indicated.

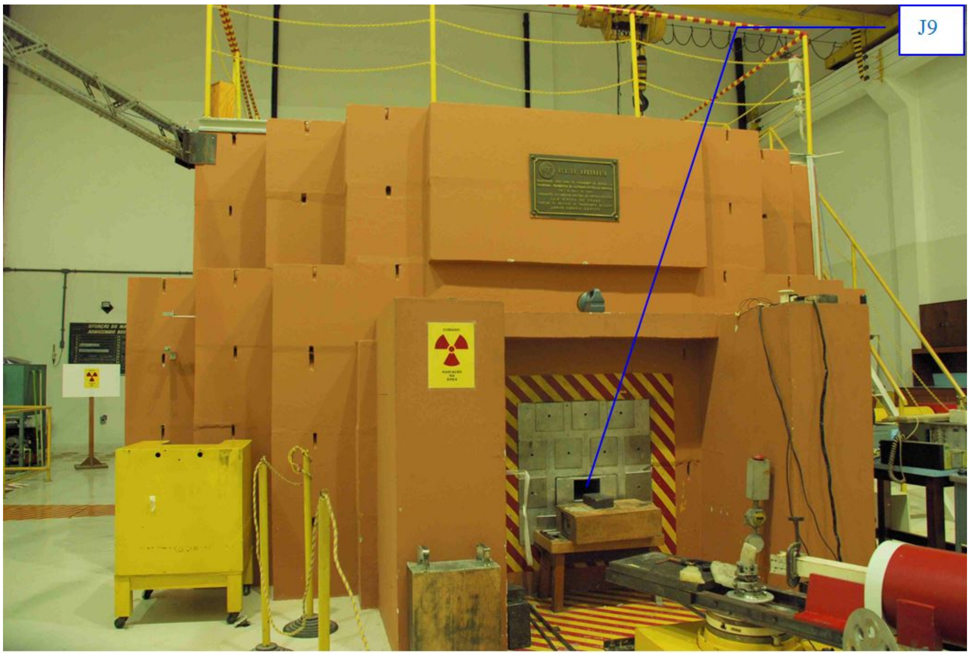

Figure 2.1: Argonauta with its J9 output radiation channel.

Figure 2.2 shows a partial architectural view of Argonauta and its Room, with some key indications, as: the Argonauta itself, its entrance, its Control Room and, again, J9. In this figure, the "room near its entrance" (as mentioned two paragraphs before), classified as a supervised area (as the Control Room), is the one at bottom center of this figure.

2.2 Radiation Measurements Routine

CSPR routinely collects dose rate measurements at some key points within Argonauta's Room; some measurements are also taken outside it when operating at 340 W [17], as in front of a metallic door, which is thinner than the whole wall[3]. As the number of measured locations is few, and they are taken somewhat away from each other, the resulting database is quite scarce. This has been the current procedure adopted by CSPR up to the present, although the resulting dose rate database is quite poor for further analyses.

Some measurements are taken by simply letting portable radiation monitors stood still at some position; then a CSPR worker comes later to take it away and

[3]This can be noticed in Figure 2.2, at its top center.

read the measurement value, and so on. In other situations, CSPR personnel remain within the Argonauta's Room carrying portable radiation monitors from place to place, according to the pre-defined key points, and have to stay stood still for some elapsed time at each one, to acquire the measurements. This is even rising questions about the real needs for CSPR workers remain within Argonauta's room during operations [17].

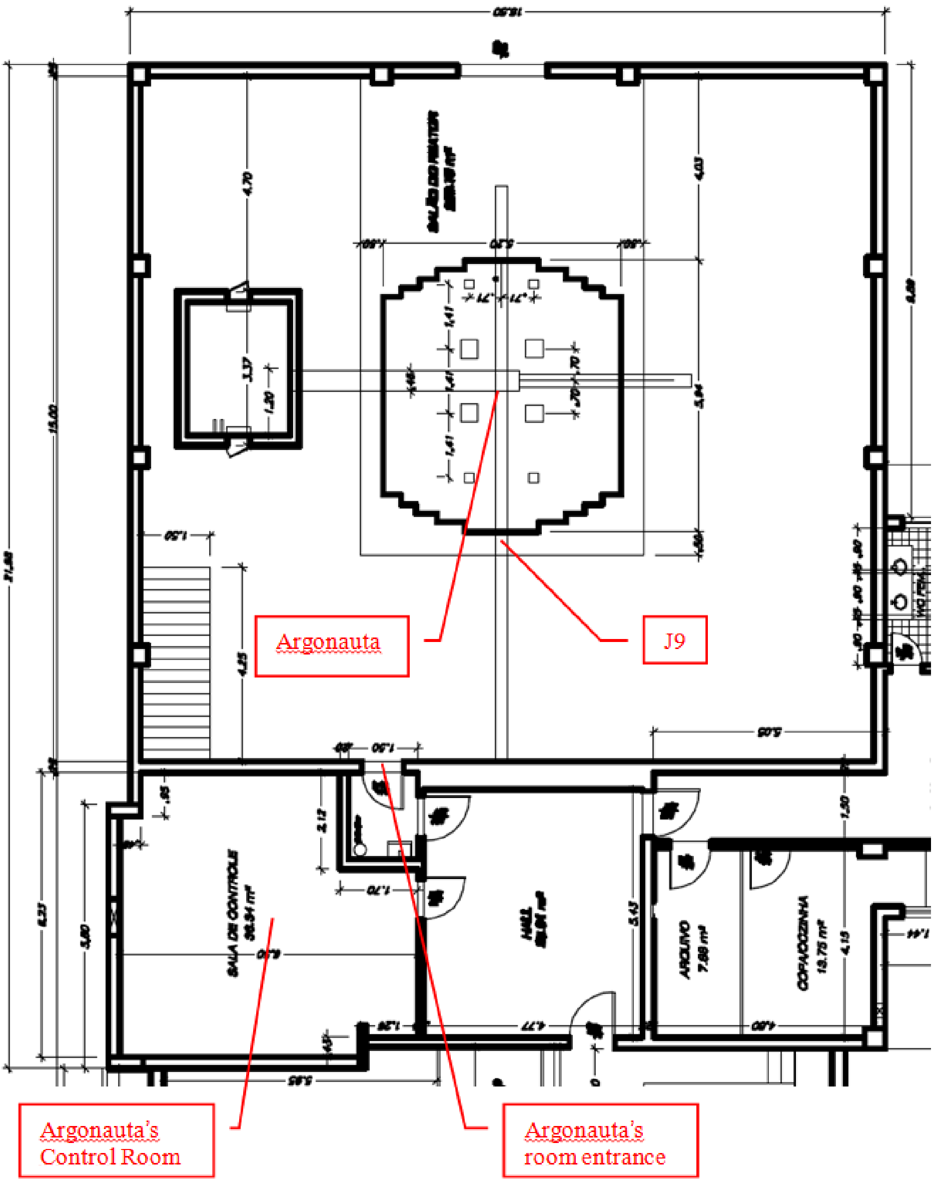

Figure 2.2: Partial architectural view of Argonauta, with some key indications.

All these matters show the needs for improvements in the measurements and radiation monitoring processes, in general.

Both gamma- and neutron-specific radiation monitors are used. Figure 2.3 shows an example of the coarse map where CSPR personnel register dose rate values, for both type of radiations. Some areas have no measures, resulting in no data (ND) indications.

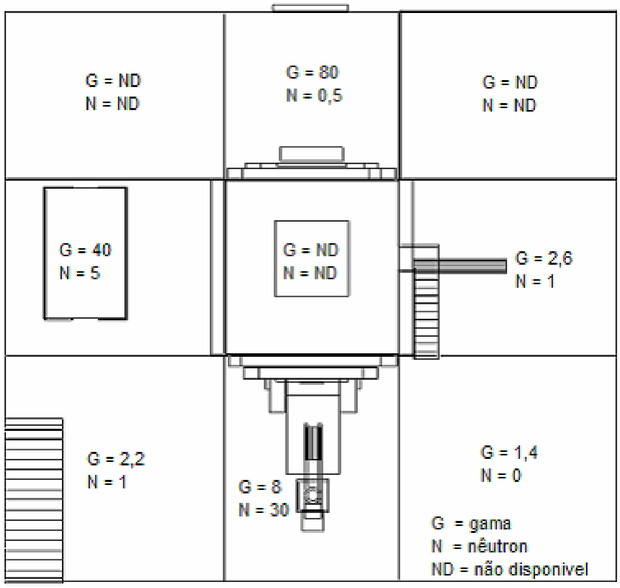

Figure 2.3: Measurements routinely collected by CSPR within Argonauta's Room.

Some measurements are taken with the J9 channel closed, some others with it open. This is due to the fact that some operations make use of J9 closed, while other ones need it to remain open - resulting in even higher dose rates, specially in front of J9 and its neighbouring regions.

Besides all these routinely monitoring processes, there are some fixed radiation monitors installed within Argonauta's Room, to monitor ambient dose and give alarms if dose rate levels increase above some threshold in their locations.

2.3 Typical Uses of Argonauta

Argonauta has been used for a somewhat broad range of research, development and teaching, as:

- Irradiation of materials for different purposes;
- Non-destructive evaluation of a diversity of materials samples, by means of radiation;
- Spectrometry experiments;
- Radioisotope production for industrial applications;

- Argonauta use during the experimental parts of undergraduate and graduate subjects offered by the IEN Graduate Program in Nuclear Science and Technology, and by other universities and colleges located at Rio de Janeiro, such as *Universidade Federal do Rio de Janeiro* (UFRJ) [21] - the Federal University of Rio de Janeiro -, and *Instituto Militar de Engenharia* (IME) - the Brazilian Army Engineering School [22].

Chapter 3

Safety in Nuclear Plants

This chapter presents a review concerning safety-related topics in nuclear plants, focusing on radiological protection. This has been addressed along the years, since dealing with radiation (both natural and artificially generated) stands for many decades. There are international guidelines that must be followed to achieve safety for people (both personnel working in this field and the general public). Different countries have also their own guidelines that, however, are based on the international ones.

Traditionally, radiological protection - or radiation protection[1] - is based on the use of measurement instruments - radiation monitors -, to assess both ambient dose and individual doses received by each worker of the field.

Radiological protection supports monitoring during routine operations and maintenance tasks. Great attentions are given to those interventions that cause greater impacts on personnel health, as: (i) fuelling/refuelling; (ii) decommissioning; and (iii) outage planning. Workers receive huge amounts of dose in such operations, and the issues of training, planning and monitoring play then even more important roles.

In recent years, some groups from different countries have been using computer-based technology to aid radiological protection[2]. This approach enables achieving safety for personnel both in routine tasks and in those more severe interventions mentioned in the last paragraph.

Some material in this chapter is partially based on [23][3], which contains a quite good review about radiological protection; this was done to clarify more this topic. Other references were also used and are cited along this Chapter - including the international and Brazilian guidelines.

---

[1] "Radiological protection", "Radiation protection" or "Radioprotection" are equivalent expressions and may be used interchangeably in this thesis.

[2] The reason for this topic is explained in Section 3.2.

[3] Hyperlinks are supplied for some of these works, to enable the readers see some of their very interesting snapshots, showing the use of these computer-based technologies to nuclear safety and ambient dose representation, since reproduction here would need previous authorization.

3.1 Radiological Protection

Dealing with radiation requires precaution, since this phenomenon is invisible for us, nor can we feel the radiation fields in which we may be immersed. A very important issue is related to human exposition to low radiation levels. The consequences of exposing to high radiation doses - as occurs typically during accidents -, are well known: the effects are deterministic, meaning the relations between high dose levels and biological consequences are straightly related. But the consequences of human exposition to low radiation dose levels - as those typically occurring in occupational exposition - do not have such straight relation: they are rather not so well understood, and are said to be stochastic, meaning exposition *may* lead to consequences... but not surely, nor how much [24, 25].

Therefore, the conduct adopted has to be conservative, justifying the need for exposition. This is what ALARA principle [1] is all about: it requires (i) justification; (ii) optimization; and (iii) limitation of doses received by people - especially of those who work directly with radiation. As already mentioned in Section 1.1, practical implementation of this principle requires basically those three recommendations, repeated her for a clearer understanding:

- Minimizing elapsed time to radiation exposure;
- Increasing distance to the radiation sources;
- The possible use of shielding.

All these three recommendations - especially the first two - require, in turn, a good planning of working tasks to be executed by personnel, and are directly related to those other requisites: optimization and limitation.

Justification means that the choice of using radiation results in benefits to the society, in such a way that its use is justified against other alternative techniques.

Optimization is related to reducing the radiation dose received by people. This requires the evaluation of the dose received by each worker, the number of workers required to execute tasks, and which radioprotection means need to be adopted, to reduce the dose received by each one. It therefore involves: (i) optimizing paths (to increase distance from radiation sources); (ii) optimizing elapsed times (to reduce exposition times); and (iii) the possible use of shielding. In a sentence: it requires good activities planning and adequate personnel training, besides environments monitoring.

Limitation is rather related to follow legal dose limits established by the responsible organisations - both the international and the local ones. Among the international organisations responsible for establishing recommendation guidelines for working and dealing with radiation, a main one is the International Commission on Radiological Protection (ICRP) [26]; together with the International Commission on Radiation Units and Measurements (ICRU) [27]. And, last but not least, there is the International Atomic Energy Agency (IAEA) [28], the main organisation responsible for nuclear energy usage worldwide.

3.1.1 Radiological Protection in Brazil

In Brazil, CNEN is the organization responsible for establishing and publishing guidelines and also for supervising their correct implementation. Its recommendations are based on the international ones, and in general changes accordingly to them, to cope with any modifications implemented in order to improve safety for people. CNEN main document related to radiological protection is *Diretrizes Básicas de Proteção Radiológica* (NN 3.01) [29] - meaning basic recommendations for radiological protection -, which establishes, among other topics, the annual dose limit of 20 mSv for each worker of the nuclear field, and of 1 mSv for people from the general public. Another important document is one of its complimentary documents named *Posição Regulatória 3.01/004:2011* [30] - that, in turn, follows the IAEA guideline [31] - and gives some definitions of areas relatively to radiation exposition, as already mentioned: (i) controlled area; (ii) supervised area; and (iii) free area. These classifications are explained in the following, as detailed also in [25]:

- Controlled area: subject to special protection and safety rules, to control normal expositions, to prevent dissemination of radioactive contamination and prevent or limit the potential expositions; are defined as those areas subject to an average exposition greater than 30 percent of the maximum limit established by CNEN.

- Supervised area: kept under supervision, even if specific safety protection means are not required; are defined as those in which dose rates are above 1 mSv in a year.

- Free area: those in which dose rates fall below 1 mSv in a year.

Those annual dose limits mentioned in the first paragraph of this subsection, correspond to averages taken in a period of time of five consecutive years. The dose limit in a given year is, in fact, higher, requiring that the averages are kept under those mean limit values. For example, for personnel working in the nuclear field, the dose received in a given year must not exceed the limit of 50 mSv, requiring that the average received annual dose taken at five consecutive years do not exceed the 20 mSv limit. This means that, if some worker exceeds the average limit of 20 mSv in a year - since still below the 50 mSv annual limit -, he must compensate for it, working less exposed to radiation in the subsequent years, to keep within the five-year average limit.

But remembering that stochastic radiation effects for humans can not be well predicted, and that, for this reason, people need to be more conservative about exposing to radiation, the most the received dose is kept below 20 mSv every year, the better.

For practical reasons, it is also common to use "secondary" limits derived from the annual one. The radiological protection services of nuclear or radioactive plants usually derive monthly, daily, and per-hour limits to protect personnel when executing their tasks exposed to radiation.

3.1.2 Current Radiological Protection at Argonauta

This subsection deals with the means that are traditionally and routinely used for radiological protection worldwide - at each plant according to their particularities and available means. But it focuses on the means used at IEN, and more specifically at Argonauta.

Radiological protection is performed by the use of radiation monitors, to evaluate both the ambient dose, and the doses received by each worker. In the former case, monitors may be fixed or portable, and in the latter case, individual or personal (as already explained in Chapter 2). The personal monitors commonly used at Argonauta (and also at other IEN installations subject to radiation) though, in general require time lapses of some weeks to be analysed and give the responses of the received dose during the period in that they had been used by workers. In counterpart, the most commonly used individual monitors do not have good accuracy in dose indication.

The Argonauta's Room is a controlled area, meaning only authorized people may enter it. Argonauta workers - the operators, who also execute maintenance tasks - have each one their respective personal radiation monitor that registers the dose received during a period of some weeks - typically a month[4] -, after which the response is known. These monitors are Thermoluminiscent Detectors (TLD) [25], which contain in their interior a thermoluminiscent material - a crystalline dielectric - that accumulates the received dose. After the period of use, they are sent to be analysed at *Instituto de Radioproteção e Dosimetria* (IRD) [32] - another CNEN RU. Figure 3.1 shows one of such monitors. Each worker uses it attached to his clothes.

[4]Sometimes it may be used for two months, in which case this information is also registered along with the accumulated received dose.

Figure 3.1: The personal TLD monitor.

There is an issue with this type of detector, related to the fact that only a portion of absorbed radiation deposited in the material is converted to light for analysis [25], but the authors of this cited reference say that with an adequate control, it is possible to achieve a good measure of accumulated dose.

A main disadvantage of this type of monitor is exactly the long time lapse every worker has to wait for until knowing the accumulated received dose information. This can cause some difficulties. Imagine some worker receives dose beyond what was expected in a given period: he will only know about it after at least a month. This may be a problem specially during some maintenance tasks - as those which require opening the Argonauta's core, for example, when higher ambient dose rates arise.

Besides this, there is an important detail to mention. For each worker who routinely executes tasks in controlled areas such as nuclear or radioactive plants, a radiologic history report is generated, mentioning his accumulated received dose per year - from the monthly-based results. This is based on TLD monitors results[5]. The minimum level for which dose is registered is 0.20 mSv (during a month). Thus, dose is only computed to compose the annual radiologic history report when it reaches a value equal to or greater than this minimum threshold in a month. If, at each month, a worker receives doses always below this threshold, it is registered in that report as null, despite the worker is receiving some dose.

[5] Besides the TLD detector showed in Figure 3.1, there are other similar types as the Albedo and right hand ring - sometimes also left hand one. For instance, that Figure also shows such a ring monitor attached the the bigger yellow one that in turn is used attached to the worker's clothes.

There is another monitor type, individual, but not assigned to any specific person: the ionization chamber of the type commonly referred to as "dosimetric pen" [25], due to its format. Any person that does not belong to the Argonauta operators staff and needs to enter Argonauta's Room must be assigned temporarily one such monitor, by registering his name just before and after entering there, besides the initial and final hours, along with initial and final dose values. Sometimes, a worker of the field must also use such monitors, when their TLD ones are sent for analysis at IRD. The received dose can be readily seen, but the disadvantage of this type of monitor is related to its low indicator accuracy: the scale is gross, and there is the risk of not noting small received doses - also because of possible parallax errors. Figure 3.2 shows one of these monitors.

Figure 3.2: The individual pen dosimeter.

There is a new type of individual monitor: a digital dosimeter by Polimaster [33], model PM1621 [34], which indicates either the dose rate at its location or the dose integrated over the elapsed time in use, and can also sound an alarm above some thresholds. Although more versatile, they have not so good precision: plus or minus 15 percent simply due to measurement statistical errors, besides other additional errors due to ambient conditions[6]. They also have not been currently used at Argonauta until recently - and even so, only in some very special situations. There are not so much of them available for use. Figure 3.3 shows one of these monitors.

A general issue related to all these personal and/or individual monitors is that they need calibration from time to time. Near their expiration date limits, measurements are not so confident. Sometimes a number of them are beyond their expiration date limits and need to be sent for calibration at IRD[7]. When personal and/or in-

[6]Temperature variations, relative humidity, exposition to magnetic fields and/or radio frequency fields, to mention some, according to its Operating Manual [34].

[7]For instance, by the time we needed to perform the radiation dose rate measurements, all of the few available digital monitors were beyond their expiration date, needing calibration; and we were informed that they would not be sent soon to IRD for it. We had to use another monitor type: a portable one.

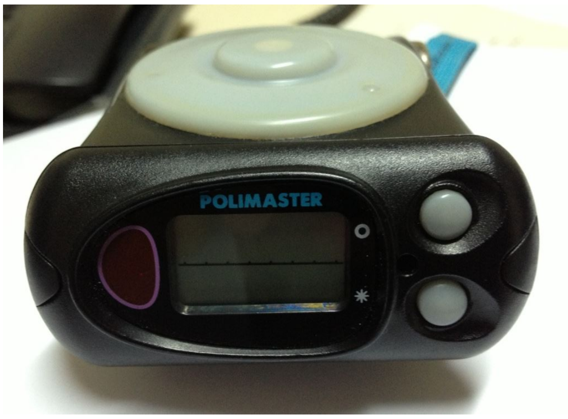

Figure 3.3: The individual digital monitor.

dividual monitors are sent for calibration, operators must use the dosimetric pens, as already mentioned (that have lower indication accuracy).

Relatively to the fixed monitors, they are installed at few places within Argonauta's Room, supplying thus scarce measurements. They give an alarm above some pre-defined dose rate threshold. The fixed monitors are the so-called MRA 7027 [35, 36], formerly developed at IEN. This monitor can be coupled to different probe types, but in Agonauta they are coupled to a Geiger-Müller type to measure gammas, the Geiger-Müller SGM 7027 [37], also previously developed at IEN. Figure 3.4 shows this monitor.

Figure 3.4: The fixed radiation monitor MRA 7027.

The portable monitors help in monitoring the Argonauta's Room during operations, but also only at some specific points.

Thus, there are all these issues related to practical usage in radiological protection. Therefore, there is a gap that encourages other developments aiming to improve safety for personnel - the reason why this thesis was proposed.

From all these considerations, it is clear the importance of developing methodologies to support better planning, monitoring and training for personnel working in the nuclear field[8]. In the next Section, some approaches dealing with novel technologies[9] applied to this purpose will be commented.

3.2 Computer-based Simulation for Safety in Nuclear Plants

This topic may, at a first sight, seem not to be directly related to this thesis, but it is treated for some reasons:

- It shows the importance of adopting novel computer-based technology to aid in radiological protection, besides the well-established more traditional means used;
- The way researchers obtain and deal with radiation dose rate distributions is of great importance for this thesis, since it makes use of it in a similar way, and improvements and developments achieved by other research groups could be added to this thesis in the future;
- Also very important, some authors mention proposals for tracking portable radiation monitors and/or workers within nuclear plants, but using quite different approaches from ours, and for somewhat different purposes.

The first item shows the tendency of people from the nuclear field for adopting novel computer-based technologies, besides traditional existing ones. This, thus, encourages the use of further computer-based methods such as the one proposed in this thesis.

Relatively to the second item, although in this thesis we are using fixed-level radiation dose rate database, it is clear that more realistic time-varying ones may also be used in the future.

The third item also gives a context in which this thesis was developed: the possibility of tracking people within nuclear plants. But we propose markerless tracking, differently from these other authors, as they mention some limitations relatively to those other approaches.

[8]Not forgetting that people of the general public also need to be protected, for example, when visiting Argonauta; or could be people who do no belong to Argonauta operators staff.

[9]For "novel technologies", it is not meant that these are novel in fact, but may be technologies already in use for other applications. The novelty is on their application to the nuclear field.

Some research groups from different countries have been adopting approaches of using computer-based simulation - particularly, virtual simulation - to improve safety for nuclear plants personnel. These somewhat recent types of technologies - specially virtual reality - have experienced lowering cost and becoming more and more accessible (and even popular, as in the case of computer games). Researchers had noticed their potential use towards more serious simulation - a field known as "serious games" [38–41][10].

Computer-based approaches for simulation are so promising to the nuclear field, that regulatory organizations such as IAEA have mentioned them by publishing some example uses for improving nuclear safety [42, 43].

In this field, two main applications have been explored: (i) simulation of environments subject to radiation (as nuclear or radioactive plants); and (ii) development of virtual control desks and rooms. The former is of interest for this thesis. Some main examples are commented in the following.

The Institute for Energy Technology (IFE) [44], in Halden, Norway, has a laboratory named Halden Virtual Reality Centre (Halden) [45], dedicated to the development and application of virtual, augmented and mixed reality-based technologies for nuclear plants simulation. Among different applications, they mention training for operations and maintenance tasks, and planning for outage and decommissioning interventions. Their research has been developed for years, since the late 1990's up to the present, and has evolved to a product: HVRC VRdose [46], which can be used for both planning and post-work analysis. This software incorporates dose rate data from computation based on radiation transport methods such as Monte Carlo or point-kernel - the latter, an alternative deterministic approach. There is a great number of references related to this development, among which some representative ones may be cited, some former and other more recent ones [47–51]. A good overview is given in [52] - specially on its Section 18.6.

Their software makes use of real-time dosimetric models (those mentioned before, based on computations) that are able to compute radiation distributions fast, and coping with changing scenarios - as when somebody changes the environment geometry by introducing shielding, for example[11]. Post-work analyses include real dose rate measurements for comparative analysis between computed distributions and real measurements.

[10] These references do not refer to applications for the nuclear field, but to using game engines towards more serious applications in general. For examples of other critical uses, including the nuclear field, readers can refer to [8], where we cite some of them.

[11] The possibility of real-time computed dosimetry is a very important topic, and has great interest for future developments over this thesis; the reason why this is highlighted in this text.

VRdose development began in a cooperation between IFE with the Japan Nuclear Cycle Development Institute[12], between 1999 and 2003, for a specific objective [54]: to support the decommissioning of the Fugen Nuclear Power Station in Tsuruga, Japan.

IFE also mentions cooperation with Tokyio Electric Power Co., Inc. (TEPKO) and TEPKO Systems Coorporation (TEPSYS), which would include also tracking radiation monitors positions within environments subject to radiation, to take into account dose rate distribution from measurements associated with corresponding tracked positions [55]; this would be a means of data gathering. All these for the objective of dose reduction and work planning. Proposal then mentioned the use of ultrasound-based positioning and of Radio Frequency Identification (RFID). Both would be used on the tip of measurement devices such as Teletectors[13]. Ultrasound would be used to locate the monitor tip position within the environment; this requires that ultrasound detectors should be placed in various locations within the environment, to detect emitting signal to locate the probe - and optionally the worker. Also, many RFID chips should be pasted onto various locations in the environment. Thus, once the monitor approached sufficiently an RFID location, it would be identified, and an action such a measurement collection could be performed. Important to mention is that measurements would be taken only at specific locations in the environment, such as near specific equipments or plant parts where the measurements were to be performed. This could not assess dose rates through workers' trajectories.

Another interesting work proposes marker-based tracking of workers, but in a rather different form: by using cameras attached to workers' Head Mounted Displays (HMDs), and visual markers pasted in various locations within the environment [56]. Interesting to mention is that in this article review, authors comment on other alternative tracking methods, explaining their advantages and disadvantages. Relatively to ultrasound tracking, they criticize this approach due to some difficulties: (i) relatively short covered ranges, needing many sensors; (ii) problems due to ultrasound reflections in the environment; and (iii) high cost of sensors. Relatively to magnetic-based tracking, they comment about problems due to: (i) low range; and that (ii) they could be affected by metal parts in the environment. Relatively to inertial-based tracking, they mention that: (i) tracking is relative to initial position and pose; and that (ii) errors are cumulative.

[12] Japan Nuclear Cycle Development Institute (JNC) and Japan Atomic Energy Research Institute (JAERI) have been unified in 2005 into a new Institute named Japan Atomic Energy Agency (JAEA) [53].

[13] Teletector: a radiation monitoring device in that its probe can be drawn out so as the worker does not need to approach too much a hot area to proceed measurements.

This paper finishes its review by establishing some needs for tracking methods suitable for nuclear plants: (i) they have to be suited for use within a building; (ii) need to have accuracy; (iii) have to avoid errors as due to metallic objects interference or ambient reflections; (iv) it have to avoid the need of previously installing many devices in different positions within the environment.

We add a comment relatively to the first requirement: Global Positioning System (GPS) would be a very good option, if in outdoors environments; but not feasible for use in indoors ones, such as nuclear plants, with their typical thick walls and roof tops made of concrete and/or steel parts, which blocks GPS signals.

The authors in [56] propose using visual markers of circular and line types, arguing that these would lengthen the distances to which the system could measure workers' positions, and that the line ones could be easily placed in pipes. But they also propose that workers use HMDs-mounted cameras, and port laptops while moving through the environment.

Here we make a question: if this would be adequate for workers, executing their own operational and maintenance tasks while porting all those equipment.

These authors also propose similar variations but all of them based on markers and the use of HMD cameras.

The Belgian Nuclear Research Centre (SCK•CEN) [57] has also developed a software platform named VISIPLAN ALARA Planning Tool, for which some representative references can be cited, some former and other more recent [58–62], by using also virtual simulations for training and planning. Doses are computed by point-kernel method, enabling to recompute dose rate distribution in pre-job planning for different scenarios, for changing geometries and placement of shielding in the simulated environment. They perform various preliminary simulations and, as long as a good configuration is achieved, more detailed dose rate distribution computations are performed for final simulation. Post-job analysis compare dose rate predictions with measurements taken at the real environment, to make adjustments for future planning and training.

Universitat Politécnica de Valencia [63] developed, in a cooperation with IBERDROLA Engineering and Construction (IBERINCO) [64], a software platform also using virtual simulations for similar objectives to the formerly exposed ones: training and planning before real interventions [65, 66]. This group focused on taking dose rates into account mainly by measurements. They report some issues in obtaining dose rate data this way, because not all locations planned to be measured are accessible, besides the problem of workers' exposure to collect the measurements. They rather collected measurements where possible, and then performed linear interpolation up to a 0.20-m resolution grid. They collected dose rates at different physical height levels within the plant room, due to their particular case; interpolations were performed separately for each height level. Dose computation based on Monte Carlo was also performed for some particular points within the environment, to complement the dose rate database preparation.

Another example is a software developed by Hitachi [67, 68], for similar purposes. In this case, dose rate was only computed, not measured. These authors report an approach for identifying the regions affected by modifications performed in the scenario, for recomputing radiation dose rate distribution only for limited areas, reducing thus the computational cost.

There is an interesting work briefly cited before [23] that develops similar approaches of using computer-based simulations for improving radiological protection, but in the European Organization for Nuclear Research (CERN) environment. Although there are some particularities related to the environment in that it was developed, it deals basically with the same objectives. Its author mentions some interesting points regarding received dose estimation for personnel that applies directly to this thesis, and are also related to the other previously cited works.

It justifies, for example, the simplifications adopted for considering only a single position for each person: although there are models for the human body considering approximately typical bodies and organs localizations, to account for more detailed radiation effects in different bodies' parts and organs, there is no one - as commented by that author, to his knowledge - that accounts for a moving body taking in consideration all the complex parts (including internal ones) movements. Also, it has to be considered that, with the monitoring approaches currently in use - by radiation monitors which can be ported by personnel -, only point measurements are taken.

This author also mentions the use of computer-based approach for obtaining radiation dose rate distributions. Another observation he does is about the difficulty of associating dose rate measurements to absolute positions within a plant, due to lack of a positioning system, and cites some tentative works that make use of radio Global System for Mobile Communication (GSM)-based systems [69–71].

But here we make an alert about issues related to Electromagnetic Compatibility and Interference (EMC-EMI) in nuclear plants, which must be considered for this type of application [72–74].

There are other examples of using virtual simulations to support monitoring, training and activities planning in nuclear plants, but these cited ones are very representative of what has been done in this field, as all of them make use of dose rate computation and/or measurements - with even recomputing functionalities, in some cases.

Our staff at IEN has also been working by following similar approaches. A virtual reality-based platform has been adapted for simulating IEN facilities, among which the main interest for this thesis is the application named Virtual Argonauta. Dose rate was initially considered only by using the scarce measurements distribution supplied by IEN radiological protection service for a proof of concept [4–6]. As soon as more detailed dose database could be obtained, it would be just a case of inserting this new one into the simulation platform.

Then, the dose rates collected by the existing radiation monitors installed within Argonauta's Room were inserted into the simulation, in their corresponding neighbouring areas [7, 8]. This brought the possibility of representing time-varying dose rates in the simulations, exactly the ones measured in real-time within Argonauta's Room.

Later, a further dose rate database was obtained in a finer grid of points in this environment, at different operating power levels, and also interpolated between these levels, fed by the monitors data, by using neural networks [9–11]. This brought a very interesting functionality for this platform, as it would now be able to approach real-time dose rate distribution within Argonauta's Room, from the previous measurements campaign, adjusted by the online data measured by the fixed radiation monitors. Why would this approach be useful for simulation? Let's comment on it.

Our staff has chosen the approach of obtaining dose rate distribution from measurement campaigns - similarly to the one adopted by [65, 66]. Of course the dose rate distributions measured this way result stationary. But, from our previous experience in collecting online dose rates from the Argonauta's Room during operations, by the fixed radiation monitors installed within its room, it was clear that radiation levels were time-varying, showing some fluctuations around different levels. Figure 3.5 gives an idea of how dose rate fluctuates even for pre-set operating power levels of Argonauta. It shows measurements collected by two radiation monitors of the model MRA 7027 installed at different locations, for three operating power levels: 17 W, 170 W and 340 W. It is possible to notice the three main power levels from the more abrupt values bumps, around which dose rates[14] vary.

Radiation transport is a rather complex phenomenon, in which many effects occur, such as reflection, transmission and absorption. This is the reason why researchers use, in general, a stochastic approach to compute it - the Monte Carlo-based computation. Alternative deterministic methods exist - as the point kernel one, mentioned before. But Monte Carlo is a typically computational-consuming method, and predicts stationary dose rate levels. Our staff idea was to consider somehow those fluctuations that occurs in real-time, to adjust the dose rate distributions accordingly.

Thus, dose rate measurements were collected at different Argonauta operating power levels. This was a time-consuming task, and involved some dose received by

[14]Dose rates are shown in mR/h, which can be converted to the now most commonly unit used mSv/h.

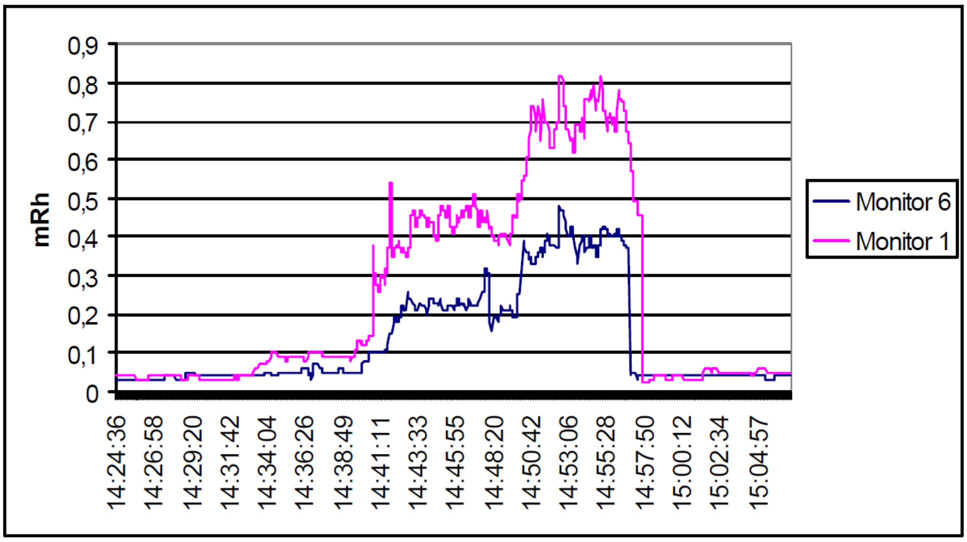

Figure 3.5: Dose rate fluctuations observed along three pre-set Argonauta operating power levels: 17 W, 170 W and 340 W.

the ones who performed them. Therefore, a number of people took this work, so as to divide dose among themselves. ALARA was used because justification criterion was met, since all this work could benefit personnel safety in the future.

The measurements were performed at different operating power levels, but the dose indications collected by the fixed monitors show the real-time fluctuations. Our staff planned to use these fluctuations to adjust online the radiation dose rate levels that are not stationary at all. As the phenomenon is complex, possibly with non-linear variations among the supposedly fixed levels for each operating nominal powers, our staff made use of a neural network-based interpolation to account for those variations. Thus, it might result in a more realistic representation of dose rate distributions in real-time, to insert them into the virtual simulation environment - Virtual Argonauta.

And, the most important point to mention here, all this development may be ported to the objective of this thesis, in the future.

3.3 Remarks

Before commenting on the computer-based approaches for radiological protection, it is important to mention the traditional means, as explained in Section 3.1. First of all, it is clear that the personal and individual dosimeters have their own limitations that can mislead received radiation dose registration for personnel of the nuclear field; fixed and portable monitors have also their limitations.

- Main drawbacks of TLD monitors are:

- They only supply responses about accumulated received doses typically a month after all the period in that they were used, what can not give an alert if someone receives higher doses levels than expected during this period;

- The responses are registered only if equal to or greater than the 0.20 mSv minimum threshold; received doses below this, despite being in fact received by workers, are not registered in the annual radiologic history report.

- Main drawback of pen dosimeters is their low accuracy for reading received dose, due to its gross scale that may lead to readings errors - despite supplying information just after use.
- The newly available digital individual monitors are more versatile, but error margin is at least 30 percent around the measured value (more or less 15 percent), and they had been made available for use too much recently, and their use is not already a routine in Argonauta;
- Both the fixed and portable radiation monitors perform measurements only in some scarce points, thus not accounting for precise measurements along workers' paths.

Following the developments cited in Section 3.2, the idea of using visual tracking of personnel to estimate received radiation doses by nuclear plants personnel has arisen. This would follow a rather opposite direction of that had been followed up to then: we were initially following the steps of other groups worldwide, for using virtual simulations for training and planning of tasks before their execution by personnel in nuclear or radioactive plants. Now, the objective would be to "measure" doses while personnel execute their tasks, without not necessarily using real monitors.

Let's comment on other more or less related works:

- Relatively to other works that mention tracking somehow, it was clear that they use different approaches from ours and, in general form, for different purposes; even for the visual approaches, authors propose using visual markers and HMD-mounted cameras.
- About the use of previously measured or computed radiation dose rates, the following possibilities can be considered for future use with this thesis developments:

  - The approaches of computing online radiation dose rate distribution by numerical computation methods, as mentioned by some other research groups, to account for modifications in the environment (such as moving objects or the introduction of shielding);

- The other approach of adapting radiation dose levels by some type of interpolation such as our staff does, among the different operation power levels, by taking into account the real dose rate variations detected by monitors installed within the plant, online;

- The possibility of combining both approaches mentioned just before, to achieve the most realistic dose rate accounting as possible.

Since the radiation dose level database can be adjusted online by any of the mentioned means, sudden variations and peaks could be taken into account along with the tracked positions; even if the final computer vision-based processing is not online.

Of course, as stated in the Introduction, all this development would not be meant to discard the traditional radiological protection means currently used, but rather to complement them, since redundancy is good for dealing with critical plants such as nuclear ones. This new approach could help in the following cases:

- In case of failure of the monitors, for example;
- For those previously mentioned cases when personal monitors are out of date and need calibration, and personnel have to use the pen dosimeters that have low accuracy;
- For coping with the minimum threshold of 0.20 mSv in a month, to register the received dose.

Therefore, taking into account all these arguments: it would be better to get such an estimate of received radiation dose, as supplied by the current developed system, than estimating no received dose at all.

Some research groups also mention post-work analysis, in which real measurements are compared to predictions, so as to improve future works. Thus, the system developed in this thesis could be used not only to supply estimates in the cases of failures, but could also record workers' positions along with measurements acquired by real monitors, for this type of analysis. Real measurements associated to tracked positions could lead to smoother and more realistic time-averaged databases in the future.

Chapter 4

Video-based Surveillance

This chapter presents a bibliographical review of computer vision-based methods for targets[1] detection and tracking.

Video-based surveillance has been receiving great attention, due to its importance for safety and security. Typical applications range from scene change detection, to targets tracking and even people behaviours analysis. The scope of this thesis involves multi-targets detection and/or tracking, not behaviour analysis. Therefore, the following sections deal with both of these topics: detection and tracking. This bibliographical research was based mainly on papers from periodicals related to image or video processing, pattern analysis, and similar ones. But some specific references can be cited as good reviews of this field, as [75–82]. From this bibliographical review, some topics can be highlighted as interesting methods:

- Targets detection and segmentation based on background modelling and subtraction;
- Targets detection and/or tracking based on parameters of interest extracted from them;
- Targets tracking based on their movements;
- Targets parts-based tracking, which involves tracking each part[2] and their integration for the whole targets tracking;
- Targets segmentation based on active contours;
- Targets detection and/or segmentation based on image salience; or
- Possible combination of different methods.

[1]Target: a general term which may mean: a person (whole body, face, upper limb,...), vehicles or any type of object of interest. In this thesis, the interest is on persons. Therefore, from then on, whenever the text refers to target, it means person - unless otherwise stated. The bibliographical research focused on persons detection and/or tracking.

[2]From this point on, targets' parts will be referred to as patches.

4.1 Background Modelling and Subtraction

Background modelling and subtraction - or background removal, for short - comprises a class of methods which can be used to detect targets in the scene, segmented from a background - or, in the most general case, from a background model. They can be used to detect and segment targets from frame to frame, but a common usage is as a first stage to initialize a region of interest (ROI)[3] - or more then one ROI -, for subsequent tracking. The initialization for tracking can be done alternatively by using detection methods based on parameters of interest.

This class of methods range from the most simple ones, as frame differences, to adaptive methods that model backgrounds in a statistical form, among other approaches. Frame differences can be performed basically in two different ways: (i) between the current frame and one previously defined as reference, or (ii) between successive frames; each one with its own advantages and disadvantages. In the former case, the fixed reference does not enable to account for any background variation, while in the latter, the background rapidly changes from frame to frame, possibly leading to failure in detecting targets that remain standing still - or almost standing still - in the scene.

New proposals emerged to enable background statistical modelling, which are thus adaptive. These may simply involve the computation of moving averages or medians, among other parameters [83]. Among these methods, $W4$ [84, 85] can be highlighted. $W4$, originally, did not deal with shadow discrimination, but there is a further work dealing with this matter, as cited by [84]. But, as the authors say, it had high computational cost. Even so $W4$ became a popular method for background removal. Further, new proposals to deal with shadow discrimination emerged for use with $W4$ [86–88]. These latter deal with gray-level images, but there are other color-based approaches.

Shadow discrimination is a very important topic, since shadows also move together with the true targets, resulting in targets location errors. Shadow discrimination is achieved basically by considering that shadow regions contain the same color distribution of background, but with less intensity values.

Another proposal is a parametric modelling considering a Gaussian distribution per pixel, named Person Finder (Pfinder) [89], by considering each picture element (pixel) has unimodal distribution. This method, as $W4$, is capable of dealing with controlled environments, where background does not present complex variations - for example, it deals well with soft and gradual illumination variations, common in indoor areas.

[3]In this work, the terms "ROI" and "bounding box" are used interchangeably to describe a region including (at least approximately) a target.

Soon after, another parametric modelling proposal arose, now considering that pixels may have multimodal distributions: the Gaussian Mixture Model (GMM) [90, 91]. It can deal better with more complex backgrounds, as ones with less interest moving parts which must not be detected as targets - as for example, moving water in shore or rivers, or moving trees leafs, commonly occurring in outdoor areas.

Later, GMM was improved, among other modifications, to consider also shadow discrimination [92–95]. GMM has higher computational costs when compared to simpler methods as $W4$, for example, but has also interesting advantages over them.

There is another proposal that performs non-parametric background modelling, alternatively to what GMM does: Kernel Density Estimation (KDE) [96–98]. It compares the pixels distributions by using histograms directly, instead of trying to model them by compositions of parametric models (as Gaussians). Its results are reported as being similar to those of GMM, and it is also cited approximately as frequently as GMM by many works.

Besides parametric or non-parametric modelling, there are other approaches for background removal. Eigenbackground [99], for example, makes use of Principal Component Analysis (PCA) [100] to generate an eigen-space, from N frames containing only background. Each new frame is then projected onto this eigen-space, possibly resulting in foreground detection - if there is any. Another difference from this method to GMM and KDE is that it is not pixel-based, but considers whole frames.

Originally, Eigenbackgorund did not cope with time-varying backgrounds. But a further proposal [101] enabled dealing with this. Model update is based on eigen-spaces fusion, where a new one is computed from time to time, based in turn on Incremental PCA (IPCA) [102].

Another interesting method, named Visual Background Extractor (ViBE) [103, 104], evaluates each pixel value, but without any explicit modelling (parametric or non-parametric). It evaluates a new pixel value with the most similar ones within its neighbour region, defined by a pre-established radius: this pixel is considered as belonging to the background if the number of similar neighbouring pixels is above some threshold. This is done in an adaptive fashion, so as to accommodate some temporal background variations. Its authors compare it to GMM favourably, as giving similar results in indoor areas, and even better in outdoors ones, able to coping with bad climatic conditions such as as wind and rain.

Blind Signal Source Separation (BSS) methods aim at separating signals (the original statistically independent signal sources) that had been mixed, without any *a priori* knowledge of neither what those source signals were, nor about the way they had been mixed - this is why they are referred to as "blind". Some points have to be considered here:

- Relatively to the sources: they may contain only second order statistics, or also superior order statistics;
- Relatively to the way they are mixed: mixtures may be linear or non-linear, and may be further instantaneous or convolutive.

These points guide the appropriate method choice for the given problem solution.

Independent Component Analysis (ICA) [105] is the basic and most commonly used method for solving BSS problems. It has been successfully applied to many diverse classes of problems, including image processing in that it can be used to separate objects from given images. This obviously leads towards the idea of separating foregrounds from the background. For example, the authors of [106] make use of ICA considering that foreground and background are statistically independent signal sources, and that given video frames containing both are the available mixtures. Therefore, foregrounds can be estimated from the given frames by this method. There are other works related to applying ICA or similar methods for targets segmentation in video [107–115].

There are possibly other methods for background removal, but GMM and KDE are among the most cited ones. ViBE and $W4$ are also very promising ones. Therefore, here finishes this part of the bibliographical research.

4.2 Detection and/or Tracking

Detection methods serve basically to find a target in the scene. They could be applied frame by frame; but only this, without considering any explicit tracking or recognition approach, may lead to switching the detections among different targets, for multi-target tracking - particularly when different targets approaches or crosses among themselves.

Tracking can aid following each target individually and discriminating among themselves by different means, as examples:

- Considering their individual trajectories;
- Discriminating them from some parameters of interest (as their colors, among others).

Tracking deals with the task of following a given target - or most commonly, a given ROI containing a target - along video frames, but requires initialization. Issues related to tracking are:

- Tracking may drift along frames, leading to targets loss - possibly due to targets appearance changes, or to distraction among different targets or between a target and background containing objects with similar characteristics;

- Tracking does not ever deal well with targets termination, requiring re-initialization; as it similarly does not ever also deal well, alone, with targets appearing in the scene.

Targets changing appearance can also cause detection errors. In some approaches, detectors have to be previously trained and, if not adaptive, they may fail to match a target that changes its appearance.

Thus, each one class of methods (detection and tracking) has different purposes and, in general, each one alone does not match the whole task of tracking a target - or targets - along a video: they should rather be combined somehow. Detection can be used to:

- Initialize a tracker in the very beginning of the tracking task;

- Re-initialize tracking when it drifts or lose tracking, which causes targets deaths.

From another point of view, tracking can be used to:

- Discriminate among multiple targets;

- Select among multiple detections in an incoming frame, eliminating false detections.

Detection or tracking can be based simply on image templates corresponding to targets' ROIs with brute data (as pixels intensity or color values), which have to be matched with candidate image regions. But detection can be alternatively based on parameters of interest extracted from these templates. Further, tracking can be based on targets' patches, instead of whole ones, to achieve robustness against occlusions.

Another important characteristic about modelling for detection or tracking, is that they may be (i) generative; or (ii) discriminative. In the former case, only targets are considered as positive examples, and modelling can be static or adaptive. In the latter case, background samples are also considered, to form false examples against which targets must be discriminated. This latter approach may improve models robustness against background clutter.

For parameter-based detection or tracking, different parameter types may be used, as: colors; textures; gradients and other derivatives for corner and edge modelling; movement; parameters derived from those ones; or combination of different parameter types. This topic is explained in more details in [75] and in [79], or in the other references cited in the beginning of this Chapter.

Some methods are commented in the following.

As template-based methods, one main example is Kanade-Lucas (KL) [116] that later evolved to Kanade-Lucas-Tomasi (KLT) [117]. The template maps pixels intensity or color values belonging to the target. The method operates in two stages:

1. An initial training one where, from different targets' realizations by affine transformations, the system learns to cope with different appearances;

2. An execution stage where the learned templates are matched with the best candidates by using distance or cross-correlation metrics.

These methods are commonly used to detect and track peoples' faces in real-time, but can be adapted to track other persons' parts or other general targets, according to its training stage. But, training for multiple poses of body's parts (as faces, for example), may not lead to good results. It is preferable to use specific pre-trained systems for each cases. They have also low robustness to occlusions, easily loosing tracking when targets cross among themselves.

There is a method commonly used for face tracking: the Viola-Jones (VJ) [118, 119], in which parameters are extracted during a training stage, and thus matched during running stage. The processing relies on shift and scale-invariant features that may be matched without the need of using pyramid-based processing, to account for multiple image scales. But this method is very sensitive to out-of-plane (three-dimensional (3D)) rotations. Also, the condition of scale invariance limits the choice of parameters to be used, as some do not present such characteristic.

A common approach of using parameters is through histograms. In the case of color-based histograms, an interesting example may be cited [120] that is referred to as capable of dealing with occlusions and background clutter.

The following methods, though, are among the most cited ones for color-based tracking: Mean-Shift (MS) and Continuously Adaptive Mean-Shift (Camshift) [121–128]. Camshift is an adaptive Mean-Shift version more suitable for video processing, partly due to its low computational cost, and also due to its adaptiveness.

MS and Camshift aim at searching for a statistical distribution mode. For targets tracking, they can be used to search for the mode of a color-based histogram extracted from a target's given ROI, from frame to frame. This search is performed by a quite simple optimization method, what favours video processing applications. Camshift is further able to adapt the ROI along time.

As occurs with GMM and KDE for background removal, MS and Camshift are among the most cited ones for their purpose of video tracking, being cited in general for comparative analyses with other proposed methods.

Another class of methods rely on derivatives extracted from the images, either gradient or Laplacian. This favours to extract corners and contours. They usually also make use of histograms. One of the most known and cited ones is the Histogram of Oriented Gradients (HOG), initially proposed by [129], and then applied for people detection in [130]. This method obtain a histogram where bins correspond to gradient orientations, weighted by gradient magnitudes, and by cumulating the number of occurrences within the orientation bins.

Other interesting methods for corners and contours detection are the Moravec [131] and Harris Corner Detector [132, 133]. Among the improvements over this latter, an interesting one [134] made use descriptors based on Gaussian derivative showing invariance to in-plane (two-dimensional (2D)) rotation. But all these approaches were single scale-based, what caused difficulties in matching targets with different target scales in the frame used for design stage and candidate frames.

To overcome this limitation, scale-invariant methods arose, as the Scale Invariant Feature Transform (SIFT) [135], later improved in [136, 137]. This method extracts descriptors based on points of interest in the images, named keypoints. As explained by its author, these are invariant to scale and in-plane rotation, and robust to illumination changes, noise, and partly robust to out-of-plane rotations; they are also robust to partial occlusion and clutter. They are simple to extract and are highly discriminative.

The SIFT method aims at identifying points in the so-called scale-space domain [138–141]. The keypoints correspond to extrema of the Difference of Gaussians (DoG) function that, in turn, is an approximation to the Laplacian of Gaussians (LoG) - for which the extrema are points of interest for scale-space image recognition.

Scale-space processing is obtained by convolving a variable-scale Gaussian filter with the original image; scale variation is achieved by varying its standard deviation. Many smoothed versions of the original image are thus generated, composing a 3D signal: with two spatial coordinates plus the scale one. The extrema thus obtained are more stable than those obtained by other methods such as gradient, Hessian and Harris Corner Detector [142]. The LoG has a pass-band filtering effect, enhancing the images corners details. SIFT includes gradient orientation computation, thus having some similarity with HOG.

SIFT is patented [143] and its code is not totally open; however, a demo version is available from [144]. But there is an alternative SIFT implementation, open source [145]. Another similar method, very popular, is the Speed-Up Robust Features (SURF) [146, 147].

There are also methods referred to as Active Contours [148, 149], to segment targets from frames and track them. This type of method is based on the minimization of an energy function that has two terms: (i) one internal; and (ii) another external. They tend to minimize when a curve approaches the foreground contour from within or without it. These methods need initialization.

Some methods consider the targets' dynamics (motion prediction) to track them. The most classic of this class of methods is the Kalman Filter [150, 151]. This method is intended for broad applications; among which targets tracking in video. It operates in two stages:

1. Estimation of the future system state;
2. Correction of estimates after collecting measurements when the systems achieves approximately the predicted state.

This method considers that measurements and transitions equations, and the initial state vector, have all Gaussian distributions, and are uncorrelated. As the filter is linear, it assumes the targets' movements are also at least approximately linear around a trajectory/state point. For video-based tracking, this method reduces the searching region to perform the prediction.

For cases where targets may present abrupt moving velocity or direction changes, or to cope with more complex distributions, it is recommended to use more sophisticated methods derived from the Kalman Filter, as the Extended Kalman Filter, among other proposals [151–156].

There is a very interesting method directed specifically towards people detection - not general targets - that, though, presents very good and fast results for this particular case: the Fastest Pedestrian Detector in the West (FPDW) [157]. This method has low computational cost and operates only for persons in vertical poses (not crouched). It can be used alternatively to, or in conjunction with, segmentation methods such as background removal, to initialize ROIs for tracking methods. It tries to match the model with the candidate frame in different scales, with is an important advantage.

There is another class of methods which split targets into patches, to cope at least with partial occlusions. These methods are able to track the visible targets' parts despite some occlusion. One main method of this class is the Fragments-based Tracking (FragTrack) [158]. This method matches patched-based histograms by the Earth's Mover Distance (EMD) metric [159], and estimates outliers, to avoid wrong-tracked patches interference. Histograms are based only on gray-level values, to reduce computational cost.

Another method with a somewhat similar approach - relatively to be also patch-based tracking - is Coherent Patch Displacement (CPD) Tracking [160, 161]; there is also a former similar method proposed [162].

Evolving from CPD, another method arose, first for single target tracking [163] that evolved next for multi-target tracking [164]. This latter may be referred to, in this thesis, by that later paper title: Combining Patch Matching and Detection for Robust Pedestrian Tracking in Monocular Calibrated Cameras (CPMD-RPT)[4], or even simpler, Fürh-Jung.

In CPD, the target is split into contiguous patches, and processing follows three stages:

1. Independent patch tracking;
2. Global target's displacement computation based on a combination among all patches tracking and on the whole target's movement prediction;
3. Parameters adaptation along time, to cope with target appearance variations.

Each patch is modelled by an average vector and a covariance. These, in turn, are based on parameters of interest extracted from their regions, maybe including: intensity, colors, textures, among others. Each patch is independently tracked by matching using the Bhattacharyya metric [165, 166]. CPD also predicts the target's movement by using Double Exponential Smoothing (DES) [167], which has a performance similar to Kalman Filter, but with less computational cost.

The global targets tracking are obtained by the Weighted Vector Median Filter (WVMF) [168] that reduce errors due to wrongly tracked patches. This filtering uses different weights, according to:

- When a patch is well tracked, based on the Bhattacharyya distance, this patch receives higher weights;
- Otherwise, higher weight is given to the prediction.

This filtering process smooths the trajectories and helps coping with partial occlusions, or short-time total occlusions.

CPD needs initialization.

The current stage reported in [164] include some more functionalities:

- Multi-target tracking;
- Automatic ROIs initialization;
- The WVMF includes, besides patch tracking and movement prediction, also frame to frame detection;

[4]This abbreviation is not used by its authors, but may be used in this thesis for easy in writing, interchangeably to Führ-Jung.

- Targets tracking in the WCS.

By using calibrated cameras and projective geometry [15], this method computes each person's position in the floor. This is an advantage for this thesis. The other evaluated methods do not implement this computation, but it is readably implementable, once the needed parameters are available - the homography, as explained later in Chapter 7, Subsection 7.1.2.

There is still a method that combines detection and tracking, also with learning, named Tracking-Learning-Detection (TLD)[5] [169–171]. The learning functionality is performed online, so as the method can adapt the targets' models to cope with changes along the video stream. This method belongs to a class of methods known as tracking-by-detection. It makes use of the pyramidal implementation of KLT [172]. It requires, though, initialization.

All the three TLD functionalities: (i) tracking; (ii) learning; and (iii) detection, operate in conjunction, correcting one another. The tracker follows a target frame by frame. The detector takes into account the different target's appearances known so far. The learning adapts the detector from new appearances, according to detection errors. The latter functionality - learning - is based on the evaluation performed by two "experts": (i) one named P-expert and another named N-expert; the former estimates targets loss, while the latter, false alarms.

Another method for detection and tracking people in video streams is Dynamic Programming - Non-max Suppression (DP+NMS) [173]. This method is directed towards multi-targets tracking, and includes targets initialization. It also deals with targets birth and death. This method, though, is non-causal - meaning it makes use of future information. Therefore, it can not be used online, or even with new streams, only with previously recorded videos; and requires a somewhat long-term training stage, according to its author.

Besides all those cited methods, there are others commonly cited in other publications for comparative analysis with new proposals, thus consisting also in good methods for tracking. It follows, thus, a brief description of some of them.

Tracking based on Multiple Instance Learning (MILTrack) [174, 175]. It makes use of an adaptive model to cope with targets appearance changing; and also belongs to the classes of tracking-by-detection methods.

The tracking method named Visual Tracking Decomposition (VTD) [176] combines the tracking task in parts by considering different parameters, different movement models and different trackers. All these information is integrated, to result in a method that is more robust to both targets' appearance and movement variations.

[5]Please, notice the difference between this and the same abbreviation used for Thermoluminiscent Detector, for radiation dose monitoring.

Another class of methods are referred to as Online Boosting [177–180]. These methods make use of classifiers to detect and track targets in video. These classifiers are not trained previously, but rather online, in an incremental form, from new incoming data; they belong also to the tracking-by-detection class of methods.

Another proposal also involves online adaptive learning models for tracking: Incremental Learning for Visual Tracking (IVT) [181–183]. This method makes also use of eigen-space projection for targets modelling.

4.3 Remarks: Methods Selection

After all this bibliographical review, some methods were given minor attention, while others were considered for further evaluation and use[6]. First, attention was directed towards some of the most cited methods in the literature, comprising three main groups:

- Background removal-based methods for targets detection and segmentation;
- Targets detection based on parameters of interest;
- Targets tracking based on parameters; in these cases, including those that consider targets dynamics, and also those that split targets (the patch-based methods).

Relatively to the first above mentioned item, the GMM, ViBE and $W4$[7] were considered for further analysis and application to the objectives of this thesis, due to the fact that they are commonly mentioned in the literature, serve very well to initialize and re-initialize ROIs for trackers, and their codes are made available - or are already implemented in computer vision software packages.

Relatively to the second item, the FPDW method is a very promising one, supplying good results and with low computational cost. It is not intended for general purpose detection, but is rather directed towards people detection. But as long as this is part of this thesis objectives, it is a good option, since specific trained detectors are supposed to work well for the applications they were developed for. Also, this method could cope with the issue of other objects ported by personnel, as monitors, which might lead to false detections, as otherwise could result by using only background removal methods.

It is important to stress the importance of methods evaluation for this thesis, especially tracking ones, to assess how well they behave facing the specific problems observed in this research, as:

[6]This selection resulted strongly based on the Qualifying Exam, occurred in June 2013.

[7]The $W4$ implementation used was a modified one, mentioned in Section 4.1: the one combined with shadow discrimination [86–88].

- People occlusion - especially by other persons, when they cross each others;
- When people approaches among themselves, even more when they stay stood still in this situation for some period of time;
- When people use clothes of the same color - what is the case in this thesis; and
- People crouching.

Some methods evaluated in this thesis are tracking methods that need initialization, and other ones used only for detection and/or segmentation. The tracking methods evaluated were:

- Camshift;
- FragTrack;
- TLD.

all of which need initialization that, in turn, may be supplied by one of the following:

- GMM;
- ViBE;
- $W4$;
- FPDW.

The latter may be used alone or in conjunction with one of the first others; or any the three firsts may be used alone, or in combination with the fourth.

Another method was considered that already combines detection and background removal with tracking:

- CPMD-RPT.

Chapter 5

Detailing the Selected Methods

This chapter treats in more details the methods selected for use with the video database generated (those commented in Section 4.3). Section 5.1 describes Camshift, Section 5.2 describes FragTrack, while Section 5.3 describes TLD. As already commented, all these methods require initialization (even TLD that has detection in its processing). Therefore, Section 5.4 describes GMM, Section 5.5 describes ViBE, Section 5.6 describes $W4$ and Section 5.7 describes FPDW. Section 5.8 describes CPMD-RPT. Section 5.9 gives some remarks on this chapter.

5.1 Camshift

Camshift is an extension of MS for tracking targets in video streams, due to its adaptive characteristics that enables it to change a ROI area.

MS is a general-purpose method for data analysis, and was theoretically derived in [184], to search for statistical distributions modes. Thus, this method had not been initially proposed for image or video processing applications, but soon these possibilities have arisen, particularly with the MS modification to the Camshift method. The history of these developments can be found in [121–128]; among which the most representative ones may be cited [123], [128].

MS and Camshift share the main basic principles, which are detailed in the following. Given an initial ROI, these methods extract color-based histograms - although other parameters may be used, as texture. Then, they converge to the distribution mode. Considering that the mode is approximately centred at the given ROI, this frame by frame convergence to the mode leads to tracking the ROI. This tracking method does not consider any explicit movement of the target, but results from its convergence to the ROI distribution mode.

The optimization method is quite simple: gradient ascendant[1]. This simple

[1]In fact, literature treats it as gradient descendent, but in this case, the objective is maximization, to converge to the distibution mode/maximum.

choice justifies itself because a target does not move much between subsequent frames. Thus, in general, the method is able to converge to the nearest extremum without getting stuck in a wrong local one that does not correspond to the mode.

Another characteristic of these methods is their somewhat robustness to outliers, since only the pixels within the ROI are evaluated - discarding bad influences of more distant clutter. Obviously, there are still pixels belonging to the background within the ROI, what may interfere and cause drift.

MS and Camshift methods make use of the Hue-Saturation-Value (HSV) color system, what has advantages over the Red-Green-Blue (RGB) one, for example. In HSV color representation, the color information itself (Hue) is separated from saturation and intensity - the latter two, with minor importance relatively to color-based pattern recognition. Another advantage of using only one color component is the resulting univariate histograms, instead of multivariate ones, for the three original color components. This contributes also to computation efficiency. In the following, the pseudo-algorithm of MS/Camshift 1 is shown:

Algorithm 1 MS/Camshift algorithm

for $FirstFrame$ **do**
  $Initial\ ROI \leftarrow Initialization$
  $Hue - based\ univariate\ histogram \leftarrow ROI\ colors$
  $2D\ probability\ map \leftarrow ROI\ colors\ evaluated\ with\ histogram$
  $Searches\ for\ ROI\ mode\ and\ moves\ towards\ it$
  $If\ Camhift,\ computes\ also\ ROI\ area$
end for
for $SubsequentFrames$ **do**
  $Computes\ mode\ for\ current\ ROI\ in\ the\ new\ frame$
  $2D\ Probability\ map \leftarrow ROI\ colors\ evaluated\ with\ histogram$
  $Searches\ for\ its\ new\ mode\ and\ moves\ ROI\ towards\ it$
  $If\ Camhift,\ computes\ also\ ROI\ area$
end for

ROI area adaptation is based on zeroth-order moment. This adaptation enables Camshift to scale according to target proximity relatively to the camera. When a target is near the camera, its position presents high variation, but then its greater ROI area compensates for this. When a target is distant from the camera, its ROI area is smaller, but as its position presents smaller variation, it compensates for this other issue. But MS, in counterpart, by using fixed ROI areas, may present more difficulties to deal with scale variation: a ROI area initialized small could present difficulties to track targets near the camera, and greater areas would include too much background pixels when targets were distant from the camera.

The moments computation for images is performed as:

Zeroth-order moment:

$$M_{00} = \sum_x \sum_y I(x, y) \tag{5.1}$$

First-order moments in each direction:

$$M_{10} = \sum_x \sum_y x I(x, y) \tag{5.2}$$

$$M_{01} = \sum_x \sum_y y I(x, y) \tag{5.3}$$

The mode results as:

$$x_c = \frac{M_{10}}{M_{00}} \tag{5.4}$$

$$y_c = \frac{M_{01}}{M_{00}} \tag{5.5}$$

where $I(x, y)$ is the pixel color (probability) at position (x, y).

5.2 FragTrack

FragTrack [158] is a patch-based method which has the purpose of improving robustness to occlusion. It splits the target's ROI in parts with no relation with human body parts, and thus can be used to track any other target type. Targets are split into rectangular patches either vertical- or horizontal-wise. It can use color or gray level-based histograms (the implementation supplied by its authors uses the latter approach to lower computational cost).

FragTrack authors mention some Camshift problems, for which they propose solutions:

- Local basin of convergence;
- Loss of pixels spatial relations;
- Occlusions.

The first mentioned problem may lead Camshift to get easily stuck on local maxima, unless the displacements between frames are really sufficiently small. The second one is due to the fact that pixels values, once split into the histograms bins, loose completely their original localization information. The third problems is due to the use of an unique ROI for the whole target: during occlusions, the color information is partially lost, maybe leading to tracking drift or target death.

FragTrack proposes to solve these problems by:

- Efficient search;
- The use of patches and their spatial relations.

First item means that the method searches for the patches around the former target's localization. This region dimensions can be modified by user. Second item tries to solve the two latter mentioned Camshift problems. In FragTrack, the use of histograms still cause loss of pixel spatial relations within each patch; but it keeps information about relative positioning among the patches. Last, the use of parts can cope with occlusions.

FragTrack obtains its gray-level histograms by using the Integral Histogram method [185], which is an efficient form of obtaining histograms from rectangular regions. This latter method is, in turn, based on the Integral Image method [118]. FragTrack extends this latter for patch-based processing.

Each patch is tracked from frame to frame, by matching with most similar candidate regions, and then combined in a robust form, by using a voting mechanism. Matching makes use of typical distance metrics for histograms comparison, since the simple norm of histograms difference, up to Kolmogorov-Smirnov statistics and the EMD metric.

Some interesting characteristics of the FragTrack method may be mentioned:

- It is robust to partial occlusions;
- It can use different histogram comparison metrics, besides Battacharyya or its equivalent Matusita;
- It deals with spatial relations through a voting mechanism;
- The efficient use of Integral Histograms turn irrelevant the way the patches are chosen.

Next, FragTrack method is detailed.

5.2.1 Patch-based Tracking

Let the whole target be designated as template T, which position is given by its center in the former frame (x_0, y_0). Let any patch in the former frame be designated as patch P_T, parametrized by (d_x, d_y, h, w), where:

- (d_x, d_y): relative distances - horizontal and vertical, respectively - from patch P_T to template T;
- (h, w): Respectively, patch P_T half height and half width.

Let *(x,y)* be the template T estimated position at the current frame *I*. Let also P_I correspond to the formerly mentioned patch P_T, but now at the current frame; its position is given by $(x + d_x, y + d_y)$. The similarity between patches P_T and P_I is given by the distance $d(P_T, P_I)$, which may be evaluated by different forms.

The target position is varied around its estimated position *(x,y)*. Each patch has its distance evaluated relatively to each one candidate, generating a voting map:

$$V_{P_T} = d(P_T, P_I) \tag{5.6}$$

5.2.2 Dissimilarity Between Patches

The metric used in principle is EMD, which evaluates dissimilarities between the histogram bins - it evaluates which probability is needed to move bins to transform one histogram into the other. Low dissimilarities may be due quantization, while higher ones mean really dissimilarities. The voting map obtained by the EMD metric is smoother than those obtained by other simpler methods, as the norm of distance between histograms that results noisier.

5.2.3 Combining the Voting Maps

Each patch corresponds to a voting map indicating its most probable candidate in the current frame. All voting maps are combined by using robust statistics to estimate the whole target position. Considering each hypothetical position *(x,y)* evaluated for a patch, results in an ordered list $[V_{P_T}|Patches P_T]$. In principle, the lower value could be chosen as corresponding to the best match, but FragTrack authors suggest choosing the *q-th* lower value to contemplate possible occlusions, where q corresponds to the non-occluded target area percentage. For example, if a 3/4 of occlusion is admitted, q - that corresponds to the non-occluded area - is chosen as 1/4 of the number of patches.

5.2.4 Integral Histogram

This method assigns point *(x,y)* the sum of all pixels values from the upper left image corner up to this point. Doing this for all four ROI corners, the sum of pixels values within the ROI is fast obtained. FragTrack uses this idea for the patches.

FragTrack also tries to minimize outliers effects by assigning lower weights for pixels near the ROI borders, maybe considering multi-level weighting. For more details, readers can refer to [158].

5.3 TLD

The authors of TLD [169–171] propose splitting the whole tracking task into three subtasks: (i) tracking; (ii) detection; and (iii) learning. They argue that only detection or tracking have their own advantages and disadvantages, and both of them are subject to failures in different ways. Therefore, their combination can improve the whole system, even more by learning the model to be detected during the video stream processing. They consider that:

- The tracker, following the target frame by frame, can supply the detector with its different realizations;
- The detector can re-initialize the tracker when the target is lost;
- The learning stage can estimate false alarms and missed detections, and this improve the model to be detected.

All this is detailed in the following.

5.3.1 *P-N* Learning

The *P-N* learning is semi-supervised, by using two data sets: one labelled and another not-labelled. The former is generated from known targets' appearances resulting in an initial estimate of them. Then, the non-labelled data is iteratively presented to the classifier. The results are assessed by independent experts:

- *P*-Expert: assesses results classified as negative, possibly detecting false negatives (missed targets);
- *N*-Expert: assesses results classified as positive, possibly detecting false positives (false alarms).

Both the detections assessed as missed targets and false alarms have their labels switched, and added to the corresponding subset: (i) the targets data subset; and (ii) the background data subset. Authors comment that both "experts" may result in errors, since there is no desired value to validate their responses, but their errors tend to cancel mutually.

The ROIs supplied to TLD in the first frame are labelled ones, the other in the following frames are non-labelled.

For the detector, there is no constraint on the number of candidate detections in a new frame, from a former ROI; maybe, more than one detection can arise. But this obviously is not desired and must be corrected. From another perspective, the tracker predicts the target's ROI in the next frame based on it position in the former one. This enables the use of a structure - a coherent trajectory for a given target along frames -, where:

- There is only one detection per target, in a new incoming frame;
- The frame-to-frame detections form a coherent trajectory.

The P-N learning explores this structure to correct both missing targets and false alarms. The P-Expert explores the temporal video structure, while N-Expert, the spatial structure, as explained in the following.

- The P-Expert receives the target predicted position given by the tracker: if the detector results in a negative output, this expert infers there was a missed target, and changes this output to a positive one;
- The N-Expert compares the detections with the position predicted by the target, choosing the best match: detections that do not intersect with this position are considered as false alarms, and thus re-label them as negative ones. The most confident (the best) match is used to re-initialize the tracker.

The learning initialization is performed by taking ten nearest regions to the given ROI. From each one of these, twenty more versions are generated by affine transformations, thus resulting in 200 positive samples. The negative ones are obtained from regions appart from the given initial ROI.

5.3.2 Detection

In TLD, a target is defined by a ROI position and scale. Internally, the aspect ratio is converted to 15 x 15 pixels, for processing. Similarity between ROIs is given by superposition:

$$S(p_i, p_j) = 0.5(NCC(p_i, p_j) + 1) \tag{5.7}$$

where:

- (p_i, p_j): patches (ROIs) for which similarity is evaluated;
- NCC: normalized cross-correlation (NCC).

The detector is implemented in three stages, in this order:

1. Patch variance evaluation;
2. Ensemble classifier;

3. Nearest-neighbour (NN) classifier.

Each stage rejects some candidates and passes the remaining to the following others. First stage rejects candidates for which gray-level variances are half of the ROI to be matched, resulting in approximately 50 percent of candidate rejection.

The second stage, composed by independent base-classifiers, where each one gives a response in terms of probability, from certain pixels comparisons; overall result is an average of these individual probabilities. A candidate passes to the next stage if this probability is over 50 percent.

In the third stage, a candidate is finally accepted as a match if a relative similarity measure is above a threshold: $S^r(p, M) > \theta_{NN}$, where:

- p: an arbitrary patch;
- M: the target's model, given by: $M = p_1^+, p_2^+, ..., p_m^+, p_1^-, p_2^-, ..., p_n^-$, where in turn: the m positive labelled patches correspond to the target, while the n negative labelled patches correspond to background;
- θ_{NN}: a user-defined threshold.

5.3.3 Tracking

The tracking stage is based on the Median-flow tracker [171], in that a tracking failure detection was added. This tracker estimates a target's movement between successive frames, based of the Pyramidal-KL tracker [172].

This tracker fails during target total occlusions and when it gets out of the scene. This is done by evaluating points in the ROI: if the $median|d_i - d_m| > 10\ pixels$, a failure is detected, where:

- d_i: an evaluated point displacement;
- d_m: the displacements median - that gives the global ROI displacement.

5.4 GMM

GMM [90, 91] performs parametric modelling of each pixel distribution by using a mixture of Gaussians. It extends, though, the one-Gaussian modelling as performed by Pfinder, for example. It is limited to fixed cameras, which is the case in this thesis.

The mixture is formed by K Gaussians - tipically in the range from three to five. It can be applied either to gray-level or color images, in the latter case, by using multivariate Gaussians. The authors in [90, 91] made a simplification of considering equal variances for each color channel in the covariance matrix. Thus, in the following, the term variance may be used sometimes, for simplicity, instead of the most general term covariance.

Simple scenes with quasi-static backgrounds can be easily modelled by one Gaussian per pixel, because there would be only a small variance around a given pixel due to acquisition fluctuations around its value. Thus, a background model is made of those simple per pixel distributions. When a target enters the scene, the variance of its corresponding pixels increase, thus not matching their respective models, and alarming for a change in the scene, and a foreground can be detached from the background.

But for scenes somewhat more complex, even the background pixels may present some variations that must not be detected as foregrounds; instead, these would result in false alarms. For a time-varying background, its (or some of the) pixels may present multimodal distributions. To cope with this, a more complex distribution model must be used - GMM proposes a mixture of simpler ones, in the case, still Gaussians. Therefore, each Gaussian can model a data cluster in a pixel distribution.

It is important to differentiate between background and foreground pixels. Pixels belonging to the background exhibit:

- Low variances σ, since background remains almost stationary;
- High persistence ω, meaning a high number of observations compose the respective distributions (the scene may remain without novelties for long term).

But pixels belonging to foreground exhibit:

- High variances, since pixels values change more abruptly when a target enters the scene;
- Low persistence, meaning a low number of observations compose the respective distributions (the change in the scene occurs for a smaller percentage of time).

Therefore, it is possible to discriminate between background and foreground pixels by considering the ratio ω/σ. The Gaussians in a mixture are sorted in decreasing order of this ratio: the higher ones compose the background model, while the bottom ones, the foreground. A heuristic is adopted for establishing the threshold.

At each new incoming frame, each pixel is compared to its model: if it matches (within 2.5 σ), the model is updated; otherwise, the pixel is detected as foreground.

Given the persistence parameter, it is noticeable that targets that remain stood still in a scene, are integrated into the background model - because variance decreases and persistence increases, and thus that ratio.

5.4.1 The Model

The probability of observing a pixel value - color-based, in general, resulting in 3D vectors - is modelled by the GMM as:

$$P(\mathbf{x}_t) = \sum\nolimits_{k=1}^{K} \omega_{k,t} N(\mathbf{x}_t \mid \boldsymbol{\mu}_{k,t} \mathbf{C}_{k,t}) \tag{5.8}$$

where:

$$N(\mathbf{x}_t \mid \boldsymbol{\mu}_{k,t} \mathbf{C}_{k,t}) = \frac{1}{(2\pi)^{2/d} \mid \mathbf{C}_{k,t} \mid^{1/2}} exp[-\frac{1}{2}(\mathbf{x}_t - \boldsymbol{\mu}_{k,t})^T \mathbf{C}_{k,t}^{-1} (\mathbf{x}_t - \boldsymbol{\mu}_{k,t})] \tag{5.9}$$

where $N(\mathbf{x}_t \mid \boldsymbol{\mu}_{k,t} \mathbf{C}_{k,t})$ denotes a multivariate Gaussian where, in turn, variables are:

- k: Gaussian index in the mixture;
- K: number of Gaussians in the mixture;
- t: current time instant;
- $\omega_{k,t}$: k-th Gaussian weight at instant t;
- $\mathbf{x}_t$: random vector with the channel colors of the given pixel at instant t;
- $\boldsymbol{\mu}_{k,t}$: $\mathbf{x}_t$ average vector, for the k-th Gaussian, at instant t;
- $\mathbf{C}_{k,t}$: $\mathbf{x}_t$ covariance matrix, for the k-th Gaussian, at instant t;
- d: the multivariate distribution dimensionality - in this case, $d = 3$, the number of color channels.

The parameter $\omega_{k,t}$ is the relative persistence of the k-th Gaussian, at instant t.

5.4.2 Model Adaptation

The Gaussians are adapted online, by a method such as k-means or Expectation-Maximization (EM) [186]. The adaptation is performed as in the following[2] algorithm 2:

[2]The gray-level (thus univariate) case is considered for simplicity, but without loss of generality.

Algorithm 2 GMM Adaptation algorithm

```
if Incoming pixel value matches at least one Gaussian of the Mixture then
    for The winning Gaussian do
        Average vector is adapted
        Covariance is adapted
    end for
else
    The least probable Gaussian substituted by new one
end if
```

The Average vector and the Covariance are adapted as:

$$\boldsymbol{\mu}_{k,t+1} = (1-\rho)\boldsymbol{\mu}_{k,t} + \rho\mathbf{x}_t \tag{5.10}$$

$$\mathbf{C}_{k,t+1} = (1-\rho)\mathbf{C}_{k,t} + \rho(\mathbf{x}_t - \boldsymbol{\mu}_{k,t})(\mathbf{x}_t - \boldsymbol{\mu}_{k,t})^T \tag{5.11}$$

where:

$$\rho = \alpha N(\mathbf{x}_{t+1} \mid \boldsymbol{\mu}_{k,t}\mathbf{C}_{k,t}) \tag{5.12}$$

where α is the learning coefficient.

The weight of the Gaussians are adapted as:

$$\omega_{k,t+1} = (1-\alpha)\omega_{k,t} + \alpha M_{k,t+1} \tag{5.13}$$

where, for the winning Gaussian:

$$M_{k,t+1} = 1 \tag{5.14}$$

and for the remaining:

$$M_{k,t+1} = 0 \tag{5.15}$$

The weights are then normalized to unity sum.

The new distribution is created for the non-match case shown in algorithm 2, has average equal to the new pixel value, high Covariance and low weight.

5.4.3 Enhancements to GMM

GMM method has been enhanced for:

- Initialization and adaptation enhancements;
- Shadow discrimination capability;

- Automatic adaptation of the number of Gaussians in a mixture.

The two firsts were proposed in [92], while the latter, in [93]. The first and the third enhancements are related to improvements in the computation, while the second one avoids bad targets localization due to false alarms caused by shadows.

Relatively to the initialization, the authors in [92] say the original algorithm [90, 91] suffered from bad convergence when a target appeared in the beginning of a video stream; the method could need around hundred frames to adapt and achieve to a background model, and one of the adaptation coefficients was too reduced, leading to slow Average vector and Covariance convergence.

These authors proposed the use of two equation systems: (i) one for the initialization; and (ii) another for medium to long term operation. First one leads to fast convergence and background model achievements, while the latter considered the last N frames for future adaptation.

Relatively to the shadow discrimination, the authors in [92] proposed to evaluate both intensity and color information, since shadows are likely to have the same color distribution of the background, but with lower intensities.

Relatively to the adaptation in the number of Gaussians per mixture, the author in [93] proposed to begin each mixture typically with a higher number of Gaussians, and evaluate each one along time, discarding the ones not supported by incoming data.

5.5 ViBE

The authors of ViBE proposed a quite different solution from that natural evolution sequence from the older simplest methods as averaging pixels values to the GMM, in which pixels values were evaluated along frames to assembly an adaptive background model. ViBE do not estimate pdf to model pixels values, but rather evaluate them in a circular neighbourhood. The method operates as in the following.

5.5.1 Background Model

No explicit pixel model is generated. Each pixel is rather compared to its neighbours to decide if it belongs or not to the background. First, a radius R is defined around each pixel. Then, for each new incoming value of a given pixel, other pixels are sampled within this circularly defined neighbourhood: the ones with the closest values to the new incoming one are chosen and counted. If the number of similar valued pixels in this region is above some threshold, the pixel is considered as belonging to the background; otherwise, it is detected as a foreground one.

ViBE authors argue that it is more reliable to compare a new incoming pixel value with a few others in its neighbour with close values to that new one, than comparing it to a great number of other pixels values. They comment that the other approaches may include outliers in a pixel model, while theirs, not; and that it would be similar to cut pixel pdf tails, taking into account only its core region.

5.5.2 Model Initialization

There seems to exist a very interesting advantage in the background model initialization process in ViBE, as compared to other methods approaches. Authors comment that other methods require a number N of frames to initialize their background model, since each pixel values must be accumulated over time to assembly a statistical model - be it parametric or not. This is, for example, the case for the GMM model. As mentioned in Section 5.4.3, proposals were made even to improve this initialization process, to deal better with the model convergence, including those cases where a target appears at the first video frames. In ViBE, the model is initialized in the first frame; that is, it does not need to wait for N initial incoming frames to model each pixel. This is due to the fact that, in ViBE, a pixel model is comprised by its spatial neighbours, not by a temporal sequence of values at its location.

ViBE authors argues that this could enable fast background model adaptation in the case of illumination changes, since only one frame is needed to adapt the model.

5.5.3 Model Adaptation

The background model generated by ViBE is also adaptive, but in a different form. When a new incoming pixel is classified as a background one, instead of eliminating older pixels from the background model, one of the previous neighbours is randomly selected to be replaced by the new one, to comprise the background model. Their authors show that the probability of a pixel belonging to the background model at time t_0 to remain in it at future time t decays monotonically - in the case, exponentially. Thus, they argue that this can deal better with different types of background variations, both fast and slow.

Another strategy is time subsampling, since the background does not need always to be adapted regularly by substituting existing background model pixels by a new incoming one classified as also belonging to the background. But this, in turn, could avoid background modelling in the case of sudden background variations which must not be detected as foregrounds, as slowly moving background objects[3]. Thus, the decision whether to substitute or not a randomly selected existing background pixel

[3]GMM also tries to deal with this issue in its own way.

by a new incoming one classified as background, is also made randomly, with a probability of 1/16, a value obtained by authors from experimentation.

Other characteristic is spatial background samples propagation. Background model propagates through neighbour pixels, and thus the local model changes also affects somewhat nearby regions. When the method decides to really substitute an existing background pixel value in the neighbourhood of a given pixel by its new incoming value - if it is classified as background -, this new pixel included in this model is also included as belonging to the models of other pixels which radius include that location.

5.6 W4

$W4$ performs in the following form:

1. In an initial training stage, based on N initial frames, the background model is obtained, per pixel, considering only those frames which pixels lay between some threshold, so as to be considered as belonging to the background; and
2. In a second stage, new frames are compared, also per pixel, to this background model, being considered as belonging to it only if laying between other threshold.

During the first stage, from the N frames, the pixels medians and standard averages are computed. Three parameters - per pixel - compose the background model: (i) the maximum pixel value; (ii) the lowest one; and (iii) the maximum difference between consecutive frames. Maximum and minimum values are considered only for those pixels for which the absolute difference between its value and its median fall within two times its standard deviation.

In the second stage, new frames pixels are considered as belonging to the background, based on those previously computed model parameters. Basically, pixels with values between models maximum and minimum thresholds are considered as background ones. Summarizing, $W4$ is a background removal method based on median filtering, and is capable of dealing with some not so complex variations in the background and has low computational cost - compared with other methods which make use of more complex statistical modelling. It also works only for gray-level images.

5.6.1 Shadow Discrimination

The $W4$ implementation chosen for use here is the one combined with shadow discrimination, as made by [86–88]. Similarly to what was explained in Subsection 5.4.3, shadow is detected basically because shadow pixels have similar intensities to those of background, but with less intensity, darker. Shadow pixels are detected by NCC, since it can identify scaled signals [86]. This is, in fact an initial estimation of shadow pixels used in this approach; a refinement is also made to avoid false background pixels. This is done by evaluating if the ratio of intensity scaling of candidate shadow pixels are approximately constant in a neighbouring region - a rectangular one. The standard deviation of these ratios are computed, and only pixels for which ratios fall within a standard deviation threshold, it is finally considered as a shadow pixel.

5.7 FPDW

FPDW [157] is probably one of the methods with best performance to detect persons in image or video frames. Its authors published also, before their actual proposed method, a review article about pedestrian detectors [187] in which they report methods that presented low false alarm rates, but run at 1 to 30 fps on 640 x 480 video frames on modern hardware of that time. Although their good results in terms of error rates, their long time consuming were critical.

FPDW authors proposed this method to be an intermediary solution between methods which propose full pyramid-based processing - in which the pyramid is formed by multiple octaves within which other multiple scales are considered -, and classifier-based detectors such as [118, 119] - even when using pyramid. The first approach deals well with detecting objects appearing on the scene at different scales, through the full pyramid-based processing. But this latter one is, by itself, time-consuming. In the second approach, objects could be detected by considering parameters that do not depend on scale, and using pyramids only considering octaves could decrease computational cost, but FPDW authors argue that this limit a lot the type of parameters to be chosen, since many of them - including gradient histograms - are not scale-invariant by themselves.

Therefore, an intermediate solution was proposed in that features are approximated within each octave, for different scales within each one, and by using such a classifier pyramid, FPDW authors argues that it reduces the time consuming by a factor between 1 to 2 orders of magnitude, while keeping detection rate within 1 to 2 percent of those other methods reported by other authors for well-known databases [157]. They approximate multi-scale gradient histograms - in fact, HOG - within an octave by only one gradient histogram computation at each one of those.

HOG, in turn, is obtained by assembling a histogram in which each bin is defined by a discrete gradient orientation, and each contribution is weighted by the corresponding gradient magnitude. Also, for each discrete orientation angle, contributions are summed to form the histogram. Considering an image $I(x, y)$, gradient orientation and magnitude are given by, respectively:

$$O(i,j) = \arctan\left(\frac{\partial I}{\partial y}(i,j) / \frac{\partial I}{\partial x}(i,j)\right) \quad (5.16)$$

$$M(i,j)^2 = \frac{\partial I}{\partial y}(i,j)^2 + \frac{\partial I}{\partial x}(i,j)^2 \quad (5.17)$$

For up-sampled image versions, no new information is added nor lost, and authors show that gradient magnitudes change proportionally with the up-sapling factor k, while the orientation is preserved. For down-sampled image versions, information is lost, but they estimated how histograms vary for both up-sampling and down-sampling, noticing that pedestrian images present similar statistics as to natural ones. Thus, their implementation is capable of detecting pedestrian images within image or video frames at different scales, by performing computations for: (i) different octaves; (ii) estimating how histograms vary within each octave for different scales, from the way they estimated how histograms vary from the experiments they made. Precision in detection is achieved with reduced computation cost. For further details on their experiments and derivations, readers can refer to [157].

5.8 CPMD-RPT

The method CPMD-RPT, in [164], is an extension of the one in [163], for multi-target tracking; and the latter, in turn, is an evolution over the CPD one [160, 161]. All of these follow the philosophy of patch-based tracking, similarly as in [158], but in a somewhat different way. In both CPMD-RPT and CPD, the targets are split in rectangular patches, stacked vertically. Each patch is independently tracked, and the global target position results from these tracks and the target new position estimation. In CPMD-RPT it is possible to consider also the target detection in the new frame.

The modifications implemented by CPMD-RPT, relatively to CPD, are:

- Multiple targets detection and tracking;
- Initializes ROIs by combining pedestrian detection by FPDW [157] and background removal by ViBE [103, 104];
- FPDW detections are also used to compose the global targets tracking;
- Gives targets positions in the WCS - at the floor.

It is composed basically by two stages: (i) detection; and (ii) tracking. Each one is detailed in the following.

5.8.1 Detection

The authors of CPMD-RPT make use of two methods for targets ROIs initialization. As we have seen in Chapter 4, initialization can be performed also directly by background removal methods. But in this case, it is performed first by FPDW. Running CPMD-RPT, it is possible to notice one advantage of FPDW over background removal methods for persons' ROIs initialization:

- Background removal methods can not discriminate between more than one person when they have some superposition; but FPDW can - if it is not so severe.

But there it is also possible to notice a disadvantage of FPDW relatively to background removal:

- Sometimes FPDW gives some false detections.

Thus, the authors of CPMD-RPT perform detections first by using FPDW; these are validated or not) by ViBE, by considering a threshold in the percentage of foreground within a ROI supplied by FPDW. Their authors chose a threshold of 20 percent below which a FPDW detection is rejected as false alarm.

For the validated ROIs, a vertical line - in the WCS - is traced to locate the target, as in the following 3:

Algorithm 3 CPMD-RPT Person Detection in WCS

1: *Sample some ROI base positions*
2: *Trace candidate lines for each sampled position*
3: *Select that with highest intersection with foreground within upper* 2/3 *of ROI*

The upper foreground point intercepted by the winning line is head estimated position, while the base sample point is feet estimated position. Considering only 2/3 upper ROI avoids possible errors if the person is walking with legs away from each other.

Detection with FPDW validated by ViBE is performed at each frame, to cope with new targets entering the scene. But full search is performed only for the first frame; for the subsequent ones, it is performed only for a margin at frame boarder.

Detections with overlaps above a threshold are rejected.

5.8.2 Tracking

Once the winning line is selected, a validated target is split into six patches, where each patch center coincides with the line.

Each patch is tracked individually. Tracking of each target is achieved by combining the components:

- Tracking results of each patch;
- Estimate of whole target position in the new frame, from past information;
- Target detection in the new frame.

Authors of [163, 164] make use of the CIELab color space, considering only the chromaticity information; for which channels a and a are considered as independent. The histograms for each channel have 128 bins and are normalized. The resulting histograms for both channels are named: h^a and h^b.

The first component type is performed by trying to match each patch to candidates in the new frame, by using the Battacharyya metric:

$$b_i = \frac{1}{2}[\sqrt{1 - BC(h_m^a, h_i^a)} + \sqrt{1 - BC(h_m^b, h_i^b)}] \tag{5.18}$$

where BC is the Battacharyya coefficient:

$$BC(h_m, h_i) = \sum_{j=1}^{N_b} \sqrt{h_m(j)h_i(j)} \tag{5.19}$$

One important detail is that searching for candidates for matching of each patch is performed not around its current position, but considering the predicted target position at the new frame, with the maximum target velocity of 1.5 m/s in the WCS. An additional relaxation is used to increase somewhat the search region.

The second component is performed by DES method [167], considering the target displacement history, as:

$$\mathbf{D}^P(t+1) = (2 + \frac{\alpha}{1-\alpha})\mathbf{D}^{'}(t) - (1 + \frac{\alpha}{1-\alpha})\mathbf{D}^{"}(t) \tag{5.20}$$

where:

- α: Smoothing factor, $\alpha = 0.08$;
- The auxiliary terms $\mathbf{D}^{'}(t)$ and $\mathbf{D}^{"}(t)$ are given by, respectively:

$$\mathbf{D}^{'}(t) = \alpha\mathbf{D}(t) + (1-\alpha)\mathbf{D}^{'}(t-1) \tag{5.21}$$

and:

$$\mathbf{D}^{"}(t) = \alpha\mathbf{D}^{'}(t) + (1-\alpha)\mathbf{D}^{"}(t-1) \tag{5.22}$$

The two components types obtained up to this point are:

- The N_P patches displacement vectors $\mathbf{D}_i$; and
- The target displacement prediction, named the (N_P+1)-th vector $\mathbf{D}_{N_P+1}$.

The whole target position can be computed by the WVMF filter [168], as:

$$\mathbf{D}_f = \frac{1}{\sum_{i=1}^{N} w_i} \sum_{i=1}^{N} w_i \mathbf{D}_i \tag{5.23}$$

where:

- $N = N_P + 1$: total number of components;
- w_i: weight of each component.

The weight w_i is important for reducing the contribution of badly matched parts, turning the method more robust to partial occlusions or even to short-term total ones.

The last component of that formerly mentioned list is detection. The method can verify if there is detection within a 1-m distance from the target position in the WCS. Two situations can occur:

1. There is only one detection in this range: thus, this detection is readily added to the WVMF filter as the (N_P+2)-th contributing vector;
2. There is more than one detection: target displacement is computed considering only up to the (N_P+1)-th component; and the nearest detection is then added for final computation.

The CPMD-RPT method can thus operate into one of four modes:

1. *Proposed-NP+ND*: uses only the patches displacement components - discarding both the prediction and the detection;
2. *Proposed-ND*: uses the patches displacement and the prediction components - discarding the detection;
3. *Proposed-NP*: uses the patches displacement and the detection components - discarding the prediction; or
4. *Proposed-All*: uses all the components.

The targets are scaled in the image plane, since the patches horizontal dimension is fixed in the WCS.

A tracking is terminated if:

- A person leaves the scene; or
- Tracking is lost - deteriorated for T_N frames.

5.9 Remarks

These methods include some of the state-of-the art ones for detection and/or tracking, and some of them are used by other authors for comparative analyses. As explained before, Camshift, FragTrack and TLD need initialization, whereas CPMD-RPT has its own initialization strategy. For the methods that need initialization, this was done by using GMM. CPMD-RPT combines FPDW with ViBE for initialization and re-initialization. GMM was run also in a stand-alone mode, just for an additional comparison, despite not able to distinguish among targets. Later, $W4$ was also used for further tests - as will be explained in Chapter 8 - since, for the generated video database, it performed well and is much more faster than ViBE.

Chapter 6

Databases Generation

6.1 Dose Rate Database

A radiation dose rate measurement campaign was performed for this thesis. This was done under the current possibilities, in which measurements would be performed manually by people standing still for some time at the desired positions to enable for radiation counting. This meant people would be exposed to radiation, mainly if it were performed with the J9 open. In the future, maybe this tedious and potentially hazardous task for humans can be performed by remotely-controlled robots carrying radiation monitors to collect the dose rates at each desired position.

A previous work by IEN staff [9–11] showed the regions within Argonauta's Room where radiation dose rate levels were important to consider during operations. Figure 6.1 illustrate what was done then. The region labelled as "Area 0" had radiation dose levels near background (BG), and was not considered for further analysis. Regions labelled from "Area 1" to "Area 3" showed levels that should not be neglected, and were subject to a further detailed measurement campaign.

For this thesis, a grid of points was defined to cover the regions from "Area 1" to "Area 3" [1]. The grid resolution was 0.5 m in both directions, and is shown with detailed information in Figure 6.2[2]. The reference system is such that:

- The x-axis is horizontal (See Figure 6.1 or, equivalently, Figure 2.3) - meaning coinciding with the internal entrance's wall (indicated in Figure 6.2);
- The y-axis is vertical (See also Figures 6.1 or 2.3) - meaning coinciding with the internal lateral left wall; and
- The reference origin coincides with the internal bottom left corner[3] - meaning

[1] This same grid was also used to obtain the projection matrices, as explained in Chapter 7, Subsection 7.1.2.

[2] This image is already rectified, as explained in Chapter 7, Subsection 7.1.1.

[3] This is not visible in Figure 6.2 because it is occluded by objects.

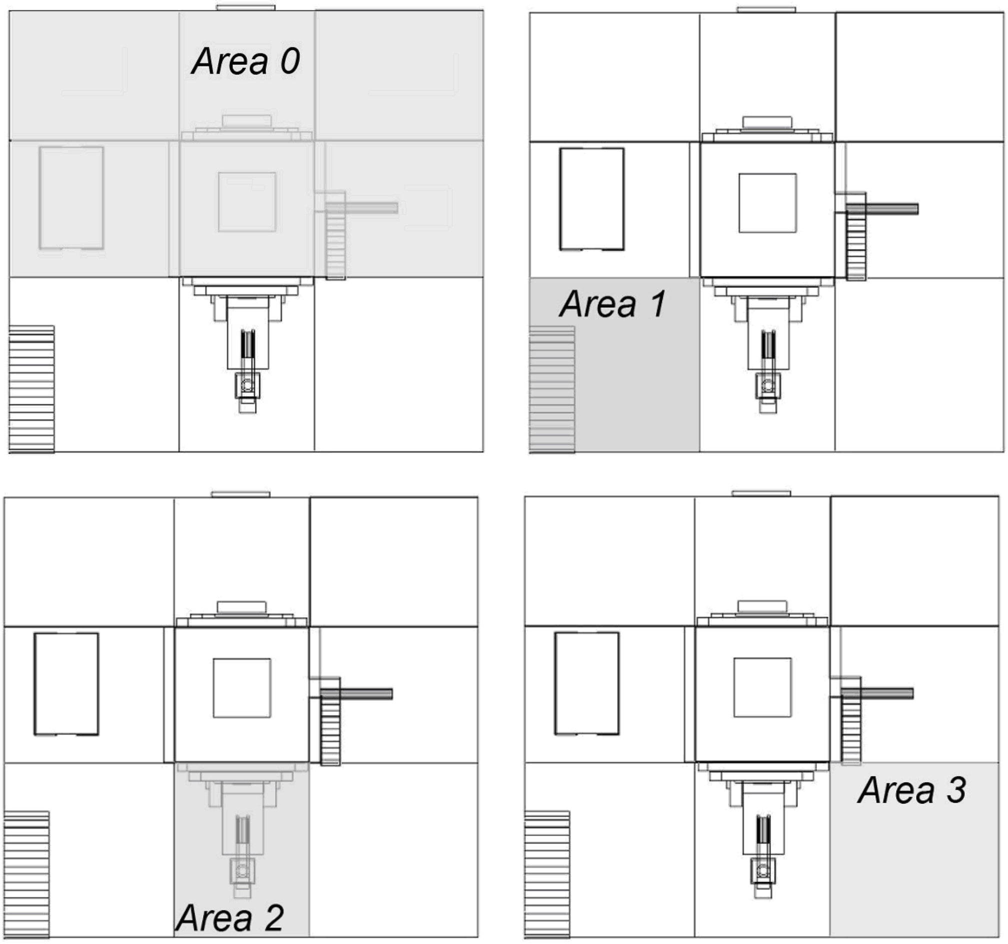

Figure 6.1: Areas evaluated to decide whether to perform or not further measurement campaign.

the internal intersection between entrance and left walls.

Referring to Figure 6.2, the distance of 3.8 m along the y-axis coincides with the first Argonauta wall[4]. Thus, the tapes were fixed beginning from this line, backwards along the y-axis, up to a minimum distance of 0.8 m from the entrance wall of the Argnauta's Room. This room has 16 m length along the x-axis, with the distance of 8 m coinciding exactly with the J9 direction. The first marked point along the x-axis was 2.0 m from the left wall, due to the lateral stair shown in the figure; up to 15 m to the right side. Some points were marked also beyond the 3.8-m distance along the y-axis.

6.1.1 Dose Rate Measurements

All the measurement campaign was oriented and accompanied by CSPR personnel. Both neutron and gamma radiations are present within Argnauta's Room during

[4]Here we mean the Reactor own concrete-made shielding wall, not its room wall.

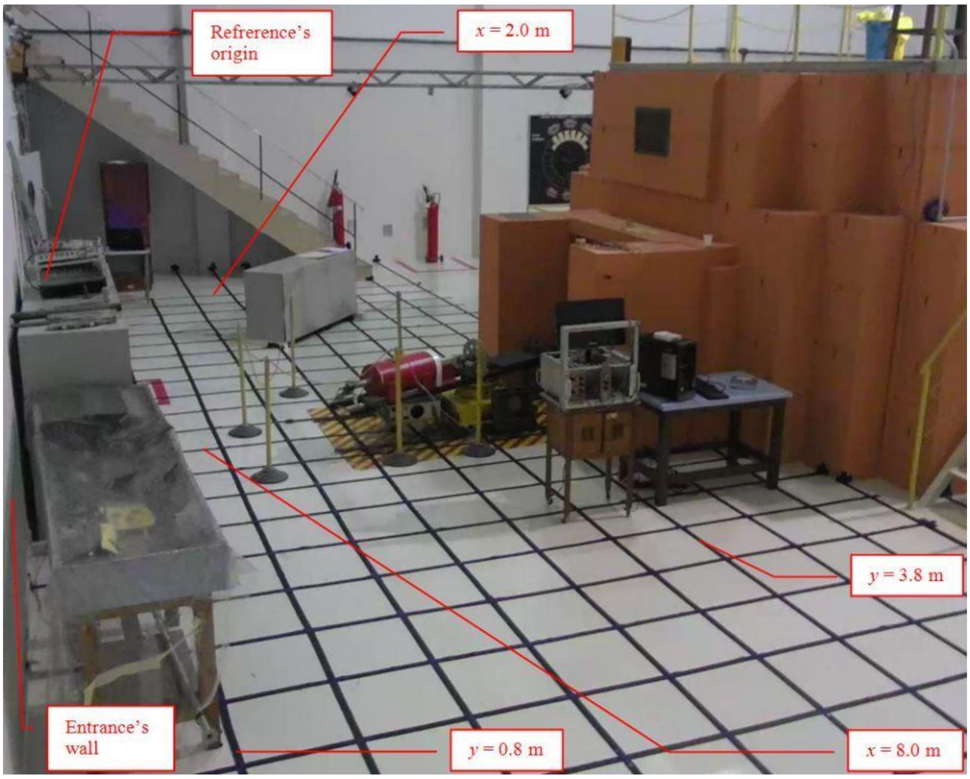

Figure 6.2: The grid used for radiation dose rate measurements.

operations; neutrons comprise the primary radiation produced, while gammas result from neutrons interaction with matter. Measurement of radiation dose rates due to neutrons are more difficult to obtain, since the monitors require longer elapsed times at fixed locations to account for measures. Gammas, though, require less time to account for them. Considering the fact that measurements would be taken manually - and worst, by only one person[5] -, CSPR did not recommend to collect neutron measurements. Besides this, as long as both radiation types are correlated, monitoring one or the other is, in general, sufficient to the routine work performed by CSPR, as the data collected by the fixed and/or individual monitors.

During some measurement routines, CSPR personnel perform both gammas and neutrons measurements, but at few specific key points [17], and in the case of neutrons, they let a monitor there at each point, coming later to take it, as commented before. That would not be the case for this thesis: rather, (i) there were much more points to be measured; (ii) Argonauta would have to operate for much longer time; (iii) the neutron monitors require longer elapsed times. Thus, this type of measurements was discarded[6].

One alert made by both CSPR and Argonauta personnel was that, with longer operation times, there is the risk of arising ionizing gases within Argonauta's Room

[5] Myself.

[6] The same difficulty arose during the older measurement campaign [9–11]

that might cause both external and internal radiological contamination for those who remains in this environment. Radiological contamination is another risk to which people may be subject, besides exposition to radiation dose - and even worse than the latter one. A person is exposed to radiation dose while he remains in a place subject to radiation fields, but the risk ceases when leaving the place. Contamination, though, occurs when someone has direct contact with nuclear or ionizing materials. In these cases, this material may remain in his body, emitting radiation for much longer time, even after leaving the place, which can cause more severe damage to his health.

Contamination may be external - when someone has contact with these materials in his skin -, or internal - when someone inhalates or ingests these materials, and it continues irradiating internally his or her body causing hazards. External contamination has an additional issue because, if not detected, people may carry this materials to other places where he or she touches or makes direct contact with. Therefore, measurements had to be taken as fast as possible[7].

Measurements were taken, for this thesis, with J9 open, since there are important experiments run at Argonauta with this condition, and thus it would be more interesting to collect them for this case. Also, with the J9 closed, results a flatter dose rate distribution, and closer BG. However, taking measurements with the J9 open posed more safety issues, and was only performed based on ALARA justification that it would be done for a research aiming to improve safety for workers.

Taking all these facts into account, and following also CSPR personnel recommendations, the operating power level of 170 W was chosen for the measurements campaign. Lower ones would result in too much measurements near BG in Areas "1" and "3" (See Figure 6.1), while the 340 W level could increase unnecessarily hazardous conditions.

The monitor used was a portable one, of model 7013, also previously developed at IEN [188, 189], coupled with the Geiger-Müller SGM 7027 probe. This monitor enables measurements in one of three time scales. Although the longer time scale would result in more precise measurements, CSPR personnel set it to shorter one, to minimize exposure time, for safety reasons.

Figure 6.3 shows the resulting measurements; red dots indicate the measured points. They were collected, along the x-axis, from 3.0 to 13.0 m; and along the the y-axis, from 0.8 to 4.3 m. It is noticeable the higher dose rates in front of J9 and nearer it (as expected), while decaying ones to both sides. The highest dose rate peak was obtained just in front of J9, and it was of 60 μSv/h, for the operating

[7]For instance, when I, who performed the measurements, finished the measurement campaign, had to pass through verification to detect possible external contamination, as alerted by CSPR personnel.

power level of 170 W.

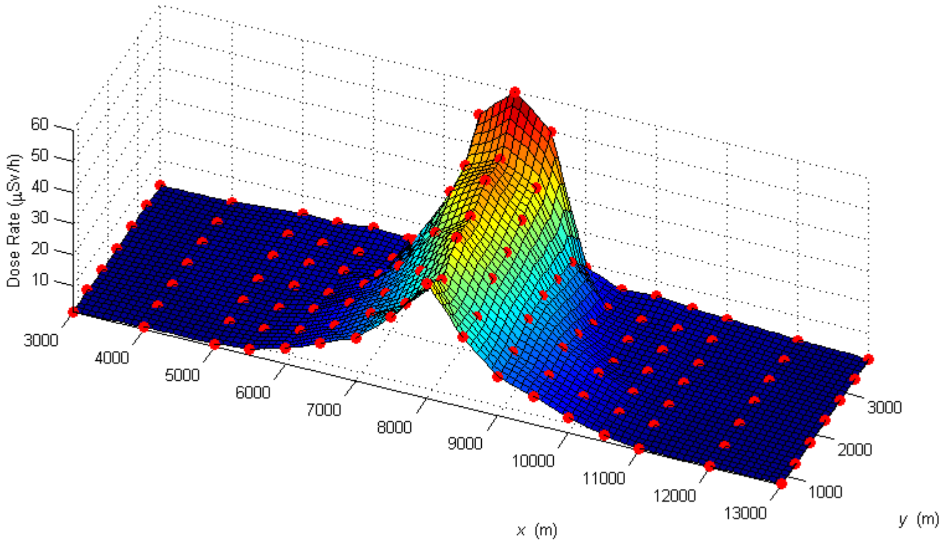

Figure 6.3: The resulting radiation dose rate measurements database.

This dose rate distribution could be interpolated to the desired resolution, by using different types of interpolation functions. But it did not seem to be so reasonable to interpolate it up to a very fine resolution - as of the order of a centimetre, for example -, since the workers' positions are approximate, spanning through broader regions, due to their bodies projected areas in the floor, and also to the intrinsic imprecisions because of legs movements. Therefore, dose rate was linearly interpolated up to a resolution of 0.10 m, from our primary resolution of 0.50 m. This is a resolution somewhat similar to that suggested, for example, by the group of Universitat Politécnica de Valencia and IBERINCO [65, 66] - they interpolated up to a 0.20-m resolution.

6.1.2 Dose Rate Computation

As already commented in Sections 3.2 and 3.3, radiation dose rates can also be computed, instead of measured. Monte Carlo N-Particle Transport method [190] is probably the most used and recommended one for this task.

A previous work was carried out for MCNP computation relatively to Argonauta [191, 192]. This work, though, encompassed only computations within Argonauta itself and its radiation channels, but also only internally. This was a very precise work in that all internal configurations and materials were taken into account, including the exact fuel rods geometries and locations. From the architectural data of Argonauta, this author began performing dose rate computations within its room; that is, extending computations to Argonauta external regions. By following this approach, dose rates can be computed for different types: (i) neutrons, gammas; (ii) for different heights, if it is necessary; (ii) for different spatial resolutions; and (iii) for different Argonauta operating power levels. This would eliminate - or at least reduce - people exposure to collect measurements.

One interesting solution is that all these computations can be combined, in the future, with the former IEN approach of taking into account the dose rate fluctuations for each operating power level, by measurements collected by fixed monitors installed within Argonauta's Room, which could be used to vary those previous computed dose rate distribution surfaces accordingly.

And, as long as this type of computations can account for modifications in the environment, all these possibilities could lead to a very interesting solution for use with both the virtual simulations - the Virtual Argonauta - and with the system developed in this thesis.

6.2 Video Database

The bibliographical review in Chapter 4 revealed the existence of some publicly available video and/or the derived frames databases for the computer vision community. But as long as this thesis would be intended for a specific application, a specific database had to be generated. Those public databases comprise many situations, some very interesting of indoors and outdoors scenes, with occlusions and other situations of interest for this community. But none of them seemed to fulfil the needs for this specific application, for the following reasons:

- It is common that Argonauta personnel use uniforms: they use dark blue upper part ones, while CSPR personnel have similar white ones;
- It is common that people stay together and standing still for some time performing some tasks;
- It is common that people crouch in some places, as in front of J9.

Thus, scenes are not so similar to general ones, in that people usually pass walking through the places, eventually experiencing occlusions generally of short term; and eventually stopping briefly to talk to others. It was not known, in advance, whether specific characteristics of the scenes within Argonauta's Room could be different enough from those publicly available ones, such that existing detection and/or tracking methods would perform well for this thesis final objective, or would need any improvement.

6.2.1 Screenplay Elaboration

Due to the facts exposed in the last paragraph, an experienced Argonauta operator explained the tasks usually performed within Argonauta's Room, to aid in composing a screenplay to be executed by people, to simulate them. There would be, in principle, the possibility of recording real operations, but sometimes they occur with more or less frequency, and the time window our staff had for recording the videos was too short, since the cameras were borrowed from an existing surveillance system already under operation at IEN at that time. Therefore, all tasks needed to be executed as fast as possible - so as not to weaken this physical security system for more time. Thus, the tasks commonly executed within Argonauta were simulated by people moving within its room by following the prepared scripts.

Let's detail some of the main tasks executed in Argonauta:

- Irradiation: is carried out with J9 closed. A material sample is inserted within J9 in a proper collimator, with a locker closing its output. After irradiation, operators come to take it out.

- Non-destructive evaluation: can be carried out by using either neutron or gamma radiations, named respectively neutrongraphy and gammagraphy. They are run with the J9 open: a radiographic plate is fixed in front of J9, to register the image of the evaluated material. A shield - named beam catcher[8] - is put behind the plate, in order to minimize radiation dose exposure, since operators have to enter just after the operation[9] - when ambient dose rate is still not negligible - to take the plate off for analysis. Sometimes, more than one plate must be used in an experiment, and an operator must enter Argonauta's Room during operation, with the J9 open, to substitute the plate. Operator is thus subject to even higher radiation dose - although hiding partially their bodies behind the beam catcher.

- Spectrometry experiments: are run also with J9 open. The researcher must stay in front of J9 for some time, to adjust manually a crystal that deflects the radiation beam to the spectrometer. The researcher tries not to stay with his whole body in the frontal J9 region, but his hands and arms still receive

[8]For instance, the beam catcher is shown in Figure 2.1: it is the yellow object at the bottom left part, mounted over wheels for smooth move.

[9]It has to be emphasized that the dose rate database obtained does not consider the movement of shielding within the environment. It would be infeasible to take it into account by using the approach chosen in this thesis: dose rate measurement. This is why the alternative approach of obtaining dose rate distributions by computer-based methods (as commented in Subsection 6.1.2) is a very interesting option, since dose rate can be recomputed at least for a limited region which would be affected by shield movement.

the beam directly. Even trying to hide his body, all that region very near to J9 is also subject to a not negligible radiation dose rate.

These basic tasks were somewhat enriched by including some variations, as: (i) in the combinations of colors of people's clothes; (ii) in the movements people could perform, to generate occlusions that could happen naturally, by people crossing each others.

Relatively to colors, it was noticed, from preliminary tests performed, the following issues:

- People using clothes made of similar colors could cause problems because the methods could switch tracking among them;
- People with some clothes made of light colors caused issues with the floor light color, leading to Camshift tracking errors, but later also leading to some failure in background removal too;
- People with some clothes made of dark colors caused issues with the background when against Argonauta walls, also made of dark red color, and with the J9 region, under shadow.

Relatively to occlusions, the following issues may arise:

- People can enter or get out of Argonauta's Room in different times, passing through each others;
- During some tasks, people usually stand still near other colleagues for some time.

The latter problem can cause an issue since tracking methods are evaluated in general for videos with people walking; then, occlusions tend to occur for short term, and trajectories estimations can be used to distinguish among them. This would not be the case for some parts of the generated video database.

Also, people usually crouch to execute some tasks, as when doing something in J9 - for example, putting or taking off samples for irradiation. This is also not common to occur in general video databases.

The generated database comprises the following scenes. The list contains brief descriptions; when there are some scenes for similar movements, they were performed by varying: (i) clothes colors combinations; (ii) times elapsed in some places or during occlusions; and (iii) some variations in the trajectories.

The following scenes just described immediately below, were performed by one person only. Two common basic movements were performed: (i) a person going to the J9 and/or to the computer; (ii) a person going to the side/back parts of Argonauta's room, where there are equipment which could be checked.

- One person, four scenes: entered room, did some movements and got out of room;
- One person, four scenes: entered room, got out of field of view, returned and got out of room;

performing a sub-total of eight scenes of these types.

The following scenes were performed by two persons, and were somewhat similar to the former ones. The first case (that of people going to the J9 and to the computer) was split basically into other two, when the first person waited and got out of Argonauta's Room together with the other, or when he got out immediately when the other passed through him, to generate somewhat different occlusion situations. The second case (that of people going to the side/back parts of Argonauta's Room) was also split into two other ones where both came back together to the cameras fields of view, and when one came first and then the other.

- Two persons, twenty nine scenes: one entered and stopped near J9, the other entered and went to the computer, came back and both got out of room;
- Two persons, eight scenes: one entered and stopped near J9, the other entered, the first got out of room immediately, the second went to the computer, came back and got out of room;
- Two persons, five scenes: both entered and got out of field of view, then returned together and got out of room;
- Two persons, seven scenes: both entered and got out of field of view, then one returned going out of room, and after the other did the same;

performing a total of 49 scenes of these types.

Some other scenes were performed with one person representing the tasks related to spectrometry experiments, also with variations. The person alternated in going to the J9 region to manipulate the crystal, and coming to the computer to see if the experiment was running with the desired quality, and took notes in a notepad. Five scenes were recorded of this type.

Spectrometry experiments were also represented by two persons dividing tasks: while one went to the J9 region to manipulate the crystal, the other remained in front of the computer with the notepad. They could alternate themselves in these tasks. Five scenes were recorded of this type.

Some scenes were recorded with two persons, to simulate lighting variations. Four scenes of this type were recorded. Finally, one scene of a real operation was recorded, when this was already running.

Summing all the recorded videos, a total of 72 scenes was produced.

6.2.2 Cameras Installation and Operation

The cameras used were from AVTech [193] - represented in Brazil by AlphaDigi [194] - of model ADIP 357 [195]. According to its manufacturer, this camera model records at a frame rate of 30 fps for the National Television System Committee (NTSC) system - which was our case. Files are saved in Moving Picture Experts Group (MPEG) version 4 - MPEG-4/H.264 format, with ".avc" file's extension. This is a rather closed format.

Our staff was authorized to borrow temporarily only two of these cameras, from the IEN existing video surveillance system, to execute the experiments for this thesis. Therefore, we had to make a decision about where to install them, to get the best coverage possible. We were also very restricted to where installing them, since Argonauta's Room had just passed then to new paintings, among other reforms; there were also very strict limitations as to where power and network cabling should be installed: this should be done near the existing conduits, to avoid loose cables through its room. All these were *sine qua non* conditions to perform the experiments.

Considering all these very strict restrictions, we decided for two places to optimize as much as possible the areas covered by the two cameras, in the regions of interest: Areas "1" to "3" (See Figure 6.1). We had also to consider that this camera model is recommended by its manufacturer for outdoors monitoring and, as such, has a bit less aperture angles than other internal models of the same manufacturer, but these ones were the only available to generate the required video database. Therefore, we installed them in opposite locations, pointing approximately towards each other - but not exactly. The reason for this choice was that regions near each one of them could not be well covered by it due to the rather less aperture angle. Thus, the other one could compensate for this, and both could cover all the regions of interest. Figure 6.4 illustrates approximately the locations where the cameras were installed.

The camera located at the right side was labelled as "Camera 1" because it was intended to be the main one, since its coverage of the area of interest seemed better. The camera at left was labelled as "Camera 2".

Figure 6.5[10] shows a photograph of Camera 1 where installed. Figure 6.6 shows a photograph of Camera 2 installed.

The highest possible resolution was chosen, in high definition (HD) video, more specifically the Super Extended Graphics Array (SXGA) format, resulting in frames of 1,280 x 1,024 pixels. This choice was done to avoid any blurring or other low quality artefacts which could lower the performance of the detectors and/or trackers.

This manufacturer supplies also a video equipment to manage up to six cameras

[10]This photograph was taken when all work had just finished. Thus, the tapes at the floor had already been taken off, and cabling was loose, as it was also being taken off.

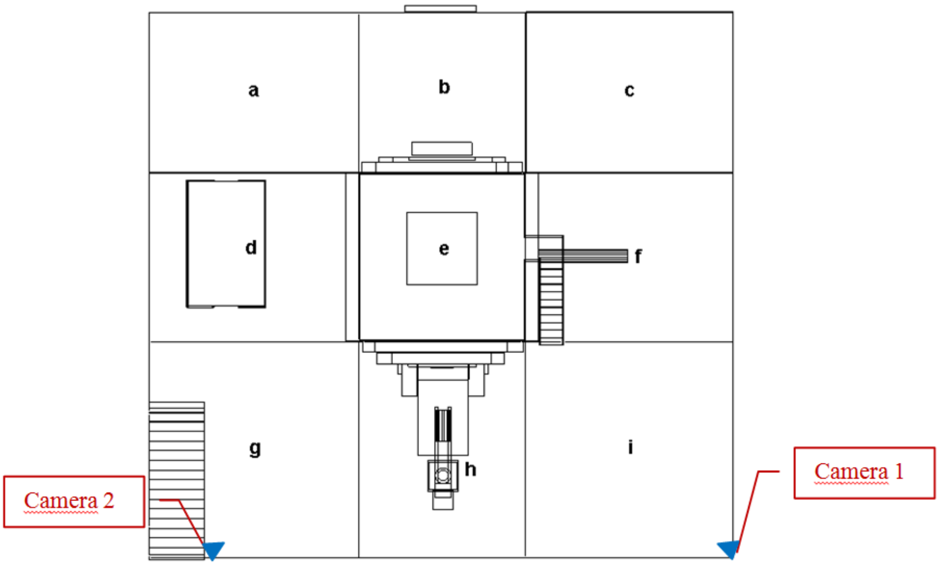

Figure 6.4: The approximate cameras locations.

Figure 6.5: Camera 1 position.

simultaneously. But, as long as the cameras were borrowed from an existing and operating surveillance-based physical security system, this was not available for use. Therefore, the two cameras were connected through a network switch to a laptop

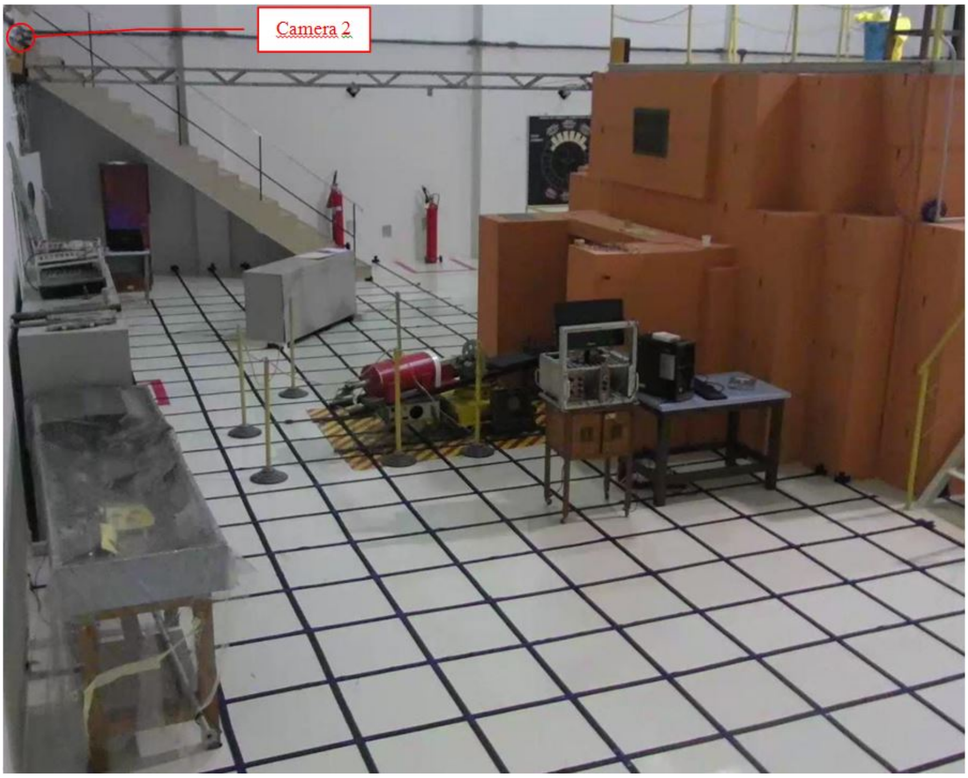

Figure 6.6: Camera 2 position.

PC running the cameras proprietary recording software.

As long as we planed to record the scenes by the two cameras, even for monocular processing, some type of synchronization was needed. This could be done by computer-based approach, as adopted by the authors in [196, 197]. But for this thesis, a hardware-based approach was adopted, by using a flash synchronization circuit developed for this purpose. It was based on the timer integrated circuit (IC) 555 in monostable operation mode, adjusted to shoot a number of power lighting-emitting diode (led) for approximately 1/30 s. This circuit is shown in Appendix A.

In the beginning of each scene recording, the cameras were started recording manually by software - and thus at different time instants. Then, someone shot the flash synchronization circuit in such a way the flash was visible to both cameras. By playing the videos after preliminary tests, it was possible to notice clearly the power leds flashing even at the distances at which cameras were installed, as well as to identify the corresponding frames.

After all the video recordings, the cameras had to be given back, and videos processed offline to prepare them for use with the detectors and trackers. First task was to convert them by using FFmpeg code [198], which is free, converts files among different formats, and usually comprises the core of other existing converters.

6.3 Remarks

Both the databases generation and processing were hard work, even more because they had to be executed under hard time pressure and few persons available to perform the screenplay.

Relatively to the video database, from all the 72 recorded videos, a number of them being left for further evaluation and processing later, if needed.

From the remaining ones, some with good representativeness were selected for further analyses. This strict selection was due to the hard work that would be taken to mark the GTs, which is a very tedious and long term demanding task, since they must be marker manually, and for all the frames - in a frame rate of 30 fps. A code obtained turned this task less time consuming, as explained in Chapter 7, Section 7.2. But even so, it still remained a hard work and took long time to be done. Therefore, it was better to concentrate in a few scenes first, than spanning through many of them. Anyway, the database can be processed for more scenes in future work.

The scenes chosen comprise:

- Two scenes with one person simulating the spectrometry experiment - one of them with two views, and the other with the right view only;
- A scene with two persons performing the above mentioned task - two views;
- A scene with two persons making some movements as going to the computer and the J9 - two views;
- A scene with two persons getting out of view then returning - two views;
- A scene with two persons with the lighting variation situation - two views;
- The real operation scene - only left view.

All these scenes with corresponding views would need to be processed for:

- Rectification, from the calibration procedure[11];
- Marking GTs, for the whole bodies and possibly the feet ROIs;
- Testing them with the selected computer vision-based methods for comparative analyses;
- besides all the processing for conversion with the FFmpeg, first of all.

[11]Calibration procedure is explained in Chapter 7, Subsection 7.1.1.

Relatively to the dose rate database, as commented in Sections 3.2 and 3.3, they can be performed by different means, basically from either measurements or computations, or by a combination of the two. In the case of measurements, it can be automated in the future by using robots coupled with radiation monitors [199–201], to cite some examples. Also, interpolation such as that IEN staff had proposed before [9–11] could cope with real-time dose rate fluctuations. For dose rate computation, some authors report different approaches that have risen mature development stages which enable fast recomputing to consider changes in the environments.

As stated in Section 1.5, the video database generated is a totally new one, than other more general existing ones. This is an important contribution to the current application and similar ones.

Chapter 7

Video Database Processing

This chapter introduces, in Section 7.1, concepts of projective geometry [15] that are necessary both for cameras calibration - which is described in Subsection 7.1.1 -, and for obtaining the projections and homographies matrices - which is described in Subsection 7.1.2. Section 7.2 describes how the GTs were obtained. Section 7.3 concludes this chapter with remarks.

7.1 Projective Geometry

In the literature, some works only evaluate detection and/or tracking methods in the image plane - the oblique plane perpendicular to a camera principal axis (See Figure 7.1 and explanation). In these cases, the methods are used to locate targets only in this plane and to improve their locations relatively to the GTs.

But in this thesis, it is important to know each targets positions in the WCS - more specifically in the floor. Therefore, by evaluating these tracked positions with the radiation dose rate distribution database, it is possible to compute the radiation dose each worker receives while walking through Argonauta's Room.

The transformation between the image plane and the WCS needs the estimation of some transformation matrices. Besides this, the acquired images in general present distortions that may be corrected - a process know as rectification - to result in more precise positioning in the WCS[1].

The only method among the selected ones that estimates targets position in the floor from projective geometry is CPMD-RPT, but it reads the projective matrices

[1] In the present case of Argonauta's Room, the rectification might be a minor problem, since distortions are very little, even in the image borders, where they typically cause more distortion. Further, these distortions occur in regions where the dose rates present the lower values, and little deviations in the workers' positions and resulting received radiation doses supposedly caused by the distortions would probably be negligible. The main region of interest with higher dose rate values, near J9, is nearer the frame center, where distortions are really negligible. But, anyway, the typical procedure of image rectification was performed, since the intrinsic and extrinsic parameters had to be obtained, as will be explained in the following Sections.

available from the video databases it uses, and those videos are already corrected for distortions. For this thesis, all the tasks of calibrating cameras, rectifying images and estimating the projection matrices should be done, as long as we are dealing with a newly generated database. All these processing is performed offline, after installing the cameras. The transformation matrices, cameras intrinsic parameters, once obtained, can then be applied online for incoming frames during operation[2].

Figure 7.1 shows the mapping between the 3D WCS and the image plane one, by using the pinhole camera model [15]. A point **X** with coordinates (X, Y, Z) is projected onto the image plane as point **x** with coordinates (x, y). This transformation combines both 3D rotation and translation, indicated respectively as **R** and **t** matrices. Camera center is indicated by **C**, its principal axis is indicated by Z_{CAM}, and principal point **p** is the intersection between principal axis and the image plane.

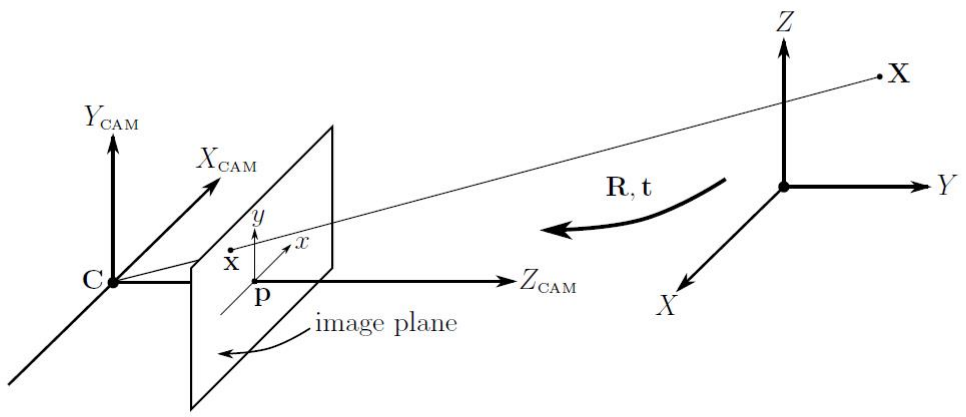

Figure 7.1: Pinhole camera model and the transformation.

The WCS is Euclidean, but the 3D scene projection into the image plane is not; rather it is known as projective space [15]. It is not a curved space: lines in the Euclidean space remain lines when projected to the projective one. But, if parallel lines in the Euclidean space "meet" at infinity, in the projective space they meet at finite vanishing points. Another important characteristic of projective space is that points projected into the image plane may correspond to an arbitrary number (infinite) of points in 3D when over a line passing through their 2D projection and the camera center (See Figure 7.1).

In projective geometry, it is common to use homogeneous coordinates in which both the 3D and the corresponding 2D projected points have their dimensions increased by one more variable; thus, a point (X, Y, Z) in 3D, in homogeneous coordi-

[2]In this thesis, all the video-based processing methodology was performed offline for previously recorded videos, split into frames. In the future, this system can be implemented online to process the incoming video directly, as long as there are software functions to acquire video streams and process them.

nates is now described by $(X, Y, Z, 1)$, while a 2D projected point (x, y) is described as $(x, y, 1)$.

Projections can be of four types:

1. 3D to 2D projection (or vice-versa): needs a projection matrix of dimension 3 x 4;

2. 2D to 2D projection, in the case of planar objects or regions in the the WCS, to the image plane (or vice-versa): needs a homography matrix of dimension 3 x 3 - a particular case of the above mentioned one;

3. 2D to 2D projection, in the case of correspondence between a pair of inter-related image planes: needs the fundamental matrix;

4. Transformation similar to the above mentioned one, but relating three image projections: needs the tri-focal tensor.

This thesis scope includes only the two first cases: projection and homography matrices. The transformation given by the projection matrix is given by:

$$w\mathbf{x} = \mathbf{PX} \tag{7.1}$$

where:

- $\mathbf{X}$: homogeneous coordinates of a 3D point in WCS;
- $\mathbf{x}$: homogeneous coordinates of its 2D projection in the image plane;
- $\mathbf{P}$: projection matrix;
- w: arbitrary scale factor.

The factor w is due to the fact already mentioned, that a 2D projected point may correspond to a line in the 3D WCS. After transformation from the 3D WCS to the image plane, for example, an arbitrary w value results; one just needs to normalize the other coordinates by this.

The projection matrix, in turn, can be decomposed as:

$$\mathbf{P} = \mathbf{K}[\mathbf{Rt}] \tag{7.2}$$

where:

- $\mathbf{R}$: 3D rotation matrix for 3D WCS to 2D transformation;
- $\mathbf{t}$: 3D translation vector for 3D WCS to 2D transformation;

- $\mathbf{K}$: camera intrinsic parameters matrix.

The matrix $[\mathbf{Rt}]$ comprises the camera extrinsic parameters, and corresponds to the spatial transformation; while the $\mathbf{K}$ matrix corresponds to the camera internal characteristics, and is given by:

$$\mathbf{K} = \begin{bmatrix} \alpha_x & s & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{7.3}$$

where:

- (x_0, y_0): coordinates of the principal point $\mathbf{p}$;
- (α_x, α_y): focal distance;
- s: skewness - inclination between camera plane coordinates axes -, null if perpendicular, which is the case for modern cameras.

This projection matrix computation considers the image is already rectified (undistorted).

The homography matrix is a particular case occurring when considering a planar object in the WCS - the floor, or any other one. Since the Z coordinate is thus null, the projection simplifies to:

$$w \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \mathbf{K} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 & \mathbf{t} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} = \mathbf{K} \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \mathbf{H} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \tag{7.4}$$

where:

- $\mathbf{H}$: homography matrix;
- Rotation matrix is split into its columns: $\mathbf{R} = \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3 \end{bmatrix}$.

The solution of either one of these matrices - projection or homography - can be obtained by the Direct Linear Transformation (DLT) method[15], or by the Random Sample Consensus (RANSAC) [15, 202] - the latter, robust to outliers.

The projection matrix has twelve elements, but only eleven degrees of freedom, while the homography matrix has nine elements, but only eight degrees of freedom[15]. Therefore, as each point in the 3D WCS (or in the projected image plane) has two coordinates, the projection needs five points and a half to be determined, while the homography needs four points. Knowing the coordinates of these points in the 3D WCS, one needs to mark or detect the corresponding ones in the image plane, to solve the matrices. But it is possible to supply more then the minimum required number of points for each matrix computation, in which cases the systems will be overdetermined and numerically solved.

7.1.1 Cameras Calibration

Cameras calibration may be done by either using a 3D well-defined object, or alternatively, by a number of views of a planar object in different poses. The latter approach was adopted in this thesis. A common approach is to use a planar object similar to a chessboard, but rectangular, not square. Grids dimensions in the WCS must be known. By acquiring a number of images of such an object at different positions and poses, it is possible to estimate both the intrinsic camera parameters and the distortion ones. Figure 7.2 shows the pattern used. Each square has dimension of 120 mm; it was printed in real full dimensions and pasted to a rigid planar object of approximately the same pattern dimensions.

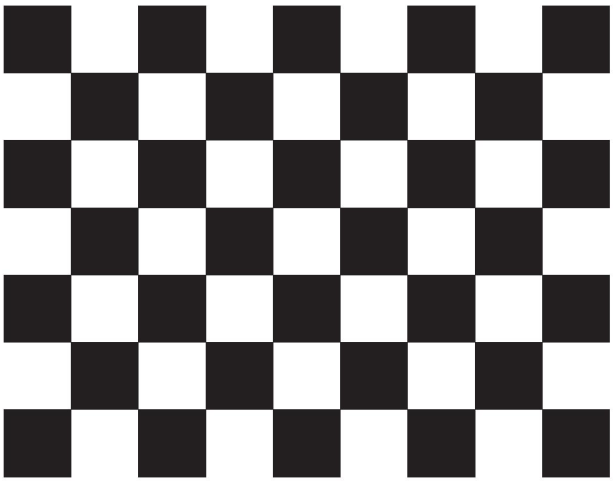

Figure 7.2: The calibration pattern used.

Figure 7.3 shows some views of the pattern image acquisition. It was used to cover different distances/positions and poses, including the boarders, where some distortion is more likely to occur, as explained in the following. A total of 30 image samples were used for the right camera, while 32 for the left one.

The distortion parameters are of two types, according to the pinhole camera model:

- Radial distortion;

Figure 7.3: Some calibration pattern sample acquisitions.

- Tangential distortion.

The first one increases from the center to the boarders of the image, and is given by:

$$x_{dist} = x(1 + k_1r^2 + k_2r^4 + k_3r^6) \tag{7.5}$$

$$y_{dist} = y(1 + k_1r^2 + k_2r^4 + k_3r^6) \tag{7.6}$$

where:

- x_{dist}, y_{dist}: distorted coordinates;
- x, y: undistorted coordinates;
- k_1, k_2, k_3: radial distortion coefficients;
- $r^2 = x^2 + y^2$.

In general, only the two firsts coefficients are needed, unless in the cases of fish-eye lens. The tangential distortion is given by:

$$x_{dist} = x + [2p_1 y + p_2(r^2 + 2x^2)] \tag{7.7}$$

$$y_{dist} = y + [p_1(r^2 + 2y^2) + 2p_2 x] \tag{7.8}$$

where:

- p_1, p_2: tangential distortion coefficients;
- Other parameters, as defined for the radial distortion case.

Tangential distortion occurs when the lens and the plane where image is projected (such as the charge-coupled device (CCD)) are not parallel. It is not common to occur in modern cameras.

These distortions can be used to rectify the images, as can be seen in Figures 7.4 and 7.5.

After calibration, the intrinsic parameters matrices resulted, for the right camera:

$$\mathbf{K}_r = \begin{bmatrix} 1462.9 & 0 & 643 \\ 0 & 1462.2 & 527.7 \\ 0 & 0 & 1 \end{bmatrix} \tag{7.9}$$

and, for the left camera:

$$\mathbf{K}_l = \begin{bmatrix} 1467.3 & 0 & 630.9896 \\ 0 & 1464.5 & 543.5115 \\ 0 & 0 & 1 \end{bmatrix} \tag{7.10}$$

7.1.2 Obtaining the Extrinsic Parameters

As explained in Section 7.1, first the projection matrix can be obtained, and then the homography results by eliminating the third column. But in this thesis, some factors led to obtaining these matrices separately:

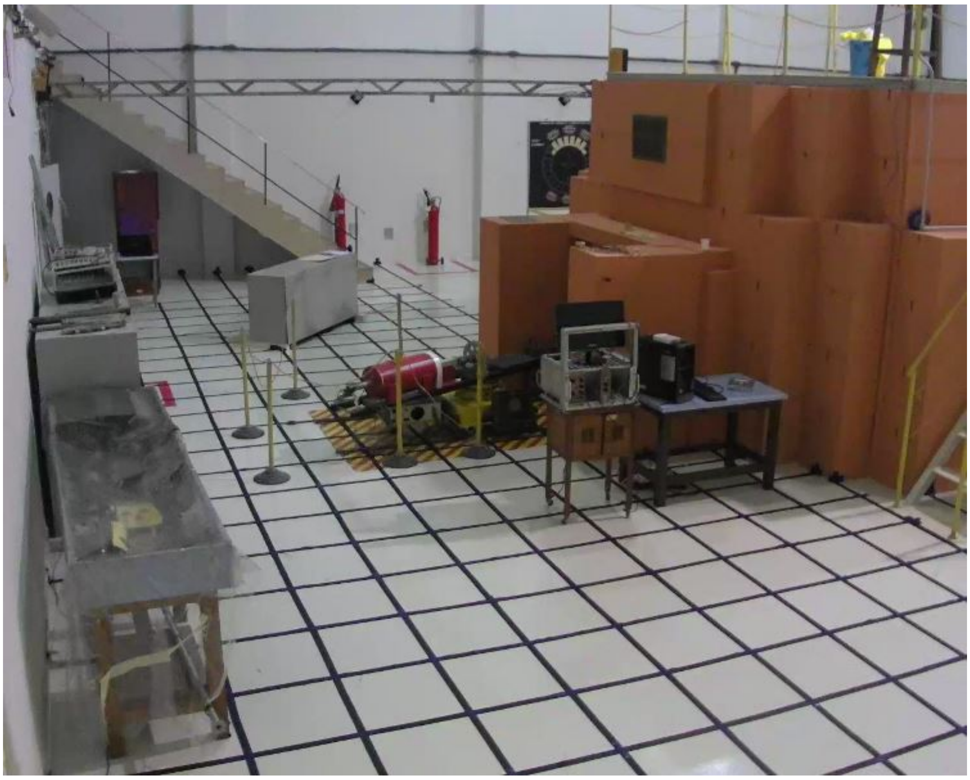

Figure 7.4: Distorted image.

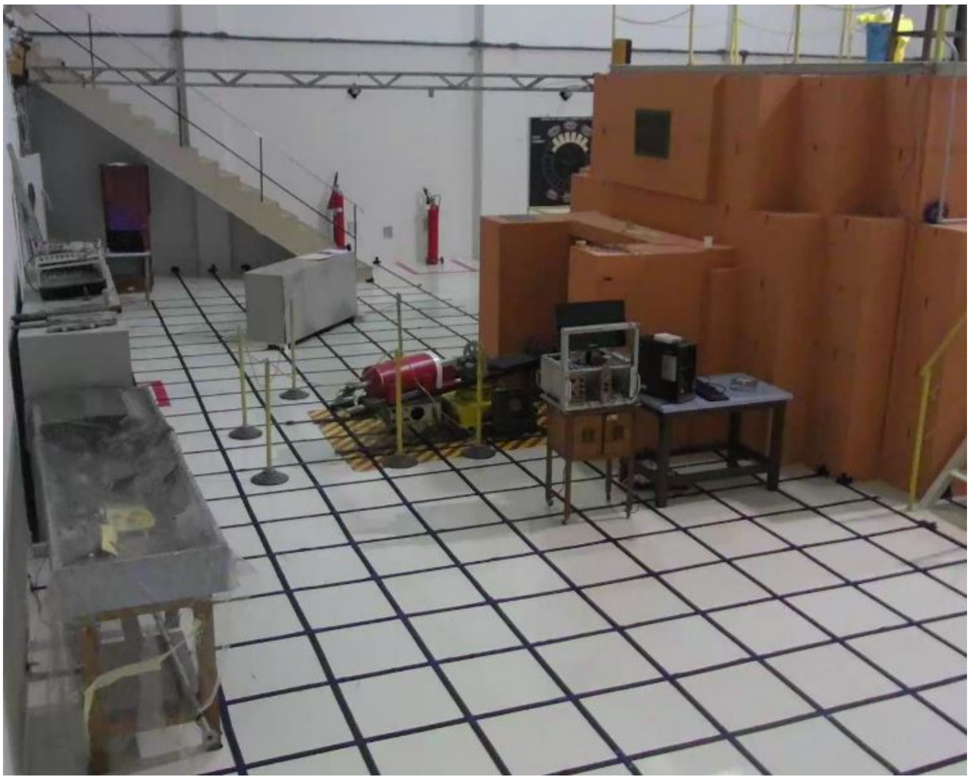

Figure 7.5: Rectified image.

- The only reason for using the projection matrix was for the CPMD-RPT method, which locates targets heads - that are out of the floor and thus needs 3D to 2D transformation;

- The most important task in this thesis is to locate targets in the floor - which needs only homography;

- There were more points in the floor, due to the tapes used for marking it, than available points in 3D, what would cause 2D to 2D transformations to the floor more precise.

Therefore, the homographies - one for each camera - were obtained separately from the projection ones. Further comparison eliminating the projections third columns showed good approximations.

The points considered for the homographies estimations were those corresponding to the tapes crossings/intersections marked in the floor - as show in the rectified image of Figure 7.5. Some corresponding points were marked in the image planes - corresponding to the right and left cameras views, for which the former is shown in Figure 7.5. Only the most visible points were considered, from visual inspection in the image planes, to discard those which could cause computations more imprecise.

For the projection matrix estimation, besides some points from the floor, other ones were collected from corners off the floor, mainly at Argonauta walls. These WCS coordinates were obtained from architectural plants and from measurements taken directly in Argonauta and its room.

For the homographies estimations using RANSAC, a total of 76 points were supplied for the right camera, of which only 29 were inliers. For the left camera, 56 points were supplied, of which only 24 were inliers.

The homographies resulted, for the right camera:

$$\mathbf{H}_r = \begin{bmatrix} -0.003995 & 0.1060 & 53.4123 \\ -0.005903 & 0.004412 & 365.0846 \\ -6.0676E-05 & 2.5397E-05 & 1 \end{bmatrix} \tag{7.11}$$

and, for the left camera:

$$\mathbf{H}_l = \begin{bmatrix} -2.8558 & 2.8238 & 3060.9623 \\ 0.51798 & 0.1009 & -12725.1893 \\ -0.002176 & -0.001114 & 1 \end{bmatrix} \tag{7.12}$$

The average reprojection error in the WCS was lower than 1 cm, which surpasses by far what was needed according to the measurements grid resolution.

The projections resulted:

For the right camera:

$$\mathbf{Rt}_r = \begin{bmatrix} 0.3602 & 0.9328 & 0.008546 & -6.0922E3 \\ 0.2713 & -0.09598 & -0.9577 & -1.6777E3 \\ -0.8926 & 0.3473 & -0.2876 & 1.5038E4 \end{bmatrix} \tag{7.13}$$

Using Equation (7.2), and after normalization:

$$\mathbf{P}_r = \begin{bmatrix} -0.003124 & 0.1056 & -0.01147 & 50.3607 \\ -0.004944 & 0.002855 & -0.1032 & 364.5711 \\ -5.9352E-05 & 2.3094E-05 & -1.9126E-05 & 1 \end{bmatrix} \tag{7.14}$$

For the left camera, the final result was:

$$\mathbf{P}_l = \begin{bmatrix} 4.7954 & -5.0572 & -1.3663 & -3.4219E3 \\ -0.96081 & -0.4038 & -6.8662 & 2.3302E4 \\ 0.003581 & 0.1512 & -0.002159 & 1 \end{bmatrix} \tag{7.15}$$

7.2 Obtaining the Ground Truths

This generated video database required obtaining the GTs, after image rectifications. This task was carried with the aid of a software supplied by some colleagues, developed by them for their application [203][3]. It runs on Linux, and has a friendly interface that enables marking objects - in this case, persons' bodies - with cursors limits, resulting in rectangular bounding boxes, with good accuracy. Only some frames need to be marked on frame intervals; linear interpolation is performed in the remaining frames between each pair of marked ones. Proper interval choices result in good GTs. Some care, though, must be taken, as explained in the following:

- A target moving in a straight - or quasi-straight line -, enables choosing greater intervals between frames, as the linear interpolation supplies good results in these cases;
- The same observation applies for a target that remains standing still - or almost standing still - for some time;
- A target that that performs a curved trajectory, or varies its velocity, requires finer intervals marking;
- The same above observation for a target that changes abruptly its trajectory.

[3]These authors gently supplied their code for our staff to generate our GTs, for which we acknowledge them.

Despite the intervals initially chosen, it is always possible to refine more them, in a posterior analysis, if interpolation does not result good enough.

More than one bounding box can be marked at each frame, resulting in different labels; all of them are saved in a text file, comprising, for each frame, each one's label and corresponding bounding box coordinates.

Figures 7.6 and 7.7 show the interactive GT marking process. The former shows the cursor placed in one ROI corner, while the latter shows the completed marking. In this case, the objective was to mark a ROI corresponding to the feet position.

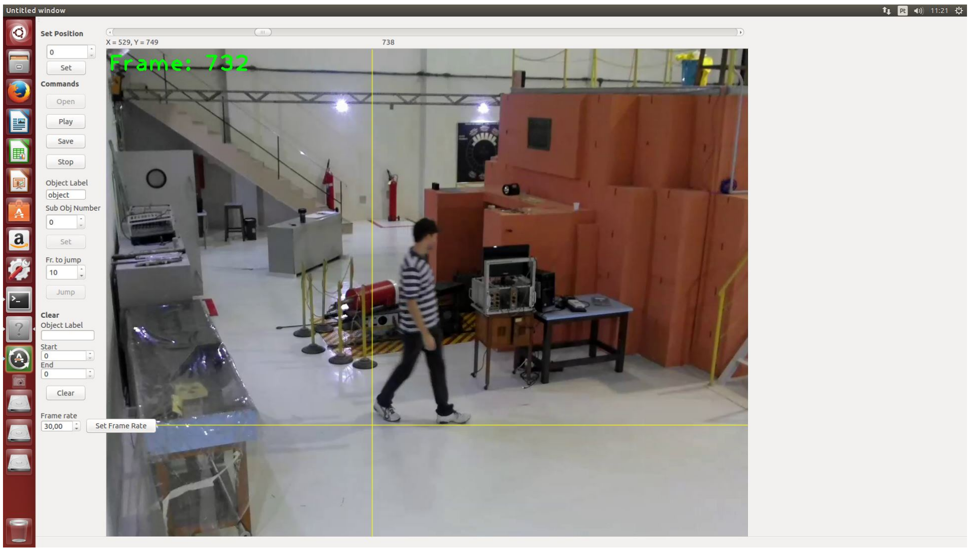

Figure 7.6: Interactive software for marking GTs.

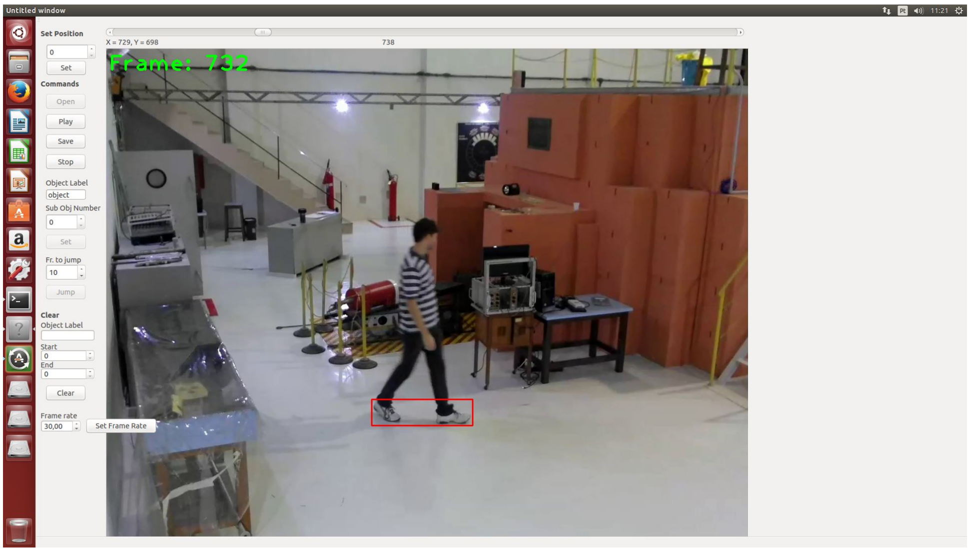

Figure 7.7: Feet GT ROI final marking.

For this thesis, besides the ROIs corresponding to a target whole body, also ROIs corresponding to his feet were also planned to be marked. The feet ROI may serve for better targets location, because it was noticed that the whole body's ROI base do not necessarily represent well the targets location in the floor, as can be noticed by comparing Figures 7.7 and 7.8, where, in the latter, the target position is most likely to be away from the whole body's ROI base. Even if it was not sure whether the tracking system would estimate or not this position, it would be good to have it recorded for future developments. These positions are the more precise, the gold ones, for estimating each one's position in the WCS. In the cases it was not possible to mark feet GTs, due to feet occlusions by objects in the environment, the whole bodies GT base center was used instead.

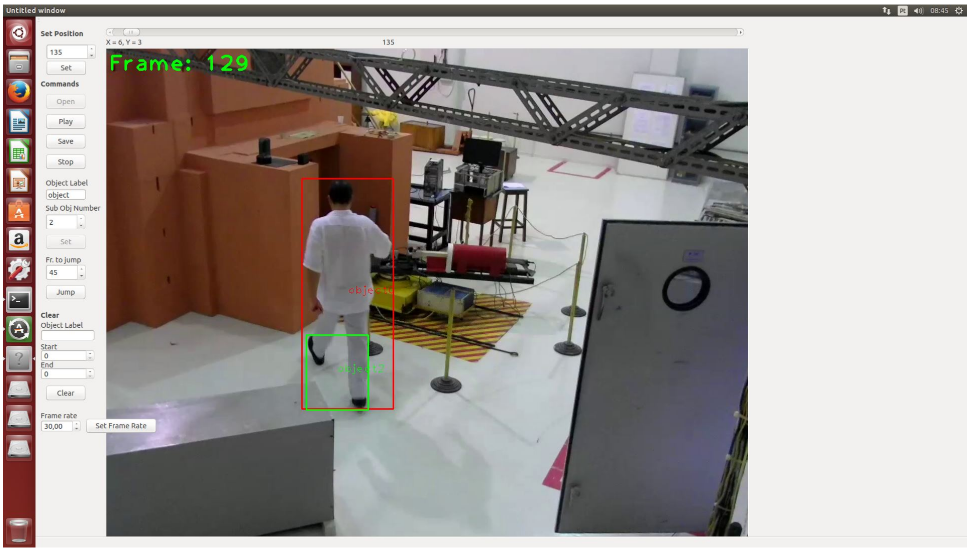

Figure 7.8: Example of a feet location most likely to be different from a whole body's ROI base.

7.3 Remarks

This chapter showed the cameras calibration process, and also how the transformation matrices (both homographies and projection ones) were obtained, for both cameras; all these processing, based on projective geometry concepts.

The cameras calibration process comprised a very important step for this thesis, since it enabled to rectify the frames/videos. Homographies, as well, are also very important, since the targets must be tracked also in the WCS - in the floor. Projection matrices were important only for the CPMD-RPT method execution, since it estimates heads positions, which are off the floor plane.

The whole bodies GTs were important in a first moment, to evaluate the selected computer vision-based methods at least for tracking in the image plane; the ones with worst performance could be readily discarded.

Considering the objective of computing received doses, most interesting GTs are the ones corresponding to feet, since targets should be better located in the WCS, although neither method predicted this position. Thus, the feet GTs center were kept where possible; otherwise, the bodies GT base center were used.

Chapter 8

Methodology

This chapter shows, in Section 8.1, the simulations results for the pre-selected methods. As explained in Sections 4.3 and 5.9, those methods were combined in some ways, as some of them do not perform alone all the tasks of detection and tracking in video. Then, Section 8.2 details the proposed solutions.

8.1 Methods Evaluation

The methods used and/or combinations adopted were:

1. GMM only, for detecting targets frame by frame;
2. Camshift initialized by GMM;
3. FragTrack initialized by GMM;
4. TLD initialized by GMM;
5. CPMD-RPT run alone.

These initial tests were performed first for a video with two persons - that one of spectrometry experiment, as listed in Section 6.3 -, since such a situation was expected to cause typical problems, as occlusions and switching among targets, and thus the more interesting situations to evaluate with the methods. The right camera view was chosen as first option since, as already mentioned, it resulted in the best views.

This first test considered, in a first moment, detection and tracking in the image plane, and was aimed at choosing among the methods with best performances for further analysis: if a method performed well in the image plane, of course it would also do for the WCS, since it would be just to make a geometric transformation to the floor.

The GMM and Camshift methods are already implemented in the computer vision free software package Open Source Computer Vision (OpenCV) [83, 204] - and the latter covers GMM version 2, with more functionalities, as the enhanced GMM implementations as described in Subsection 5.4.3, specially shadow discrimination. OpenCV is freely available for download from [205]. Thus, this favoured ready implementation of this combination.

FragTrack is not implemented in OpenCV, but was written in C language and is available for download from [206]; thus, it was also suitable to call OpenCV GMM implementation to initialize it.

TLD is implemented in MatLab [207], and is available for download from [208]. Thus, the already run GMM resulting ROIs were used to initialize it offline. An issue related to TLD is that it did not run on the operational system and software versions then in use in our Laboratories: Windows 7 and MatLab version 2014. After some tries, a virtual machine was installed with an older 32-bits Windows version, and with an older version of MatLab. Of course, this turned execution much more slower, but these were the means in which it was able to run.

These methods were used, as much as possible, as obtained from authors, for comparison - similarly to what is reported by them - following also somehow what was commented by CPMD-RPT authors in their articles [163, 164], and as is usual among authors, for comparative analysis among methods. But some modifications had to be readily implemented, as trying to extend Camshift, FragTrack and TLD for multiple targets tracking - as they were intended to track only one target at a time -, by supplying them with more than one initial ROI, from which they could begin different tracking instances from frame to frame.

Another modification - similarly to as CPMD-RPT authors also reported then - was in relation to FragTrack, which had to have its searching region enlarged to cope with the HD frames. We had to extend it to a 170 x 170 pixels-region. This was done by try, gradually, until resulting in good tracking. Probably, such a great region was needed due to the the fact that, besides videos were obtained in HD, in some scenes the targets are close to the camera, resulting big relatively to the frame. This modification increased too much the processing time, but resulted in better matching of the regions, and thus in better tracking results[1]. Another modification in relation to FragTrack was that it was not possible, at the time tests began, to use multiple instances to be tracked simultaneously, thus, each instance was run separately. CPMD-RPT authors also report this approach for their comparative analysis, in which each target was individually tracked and the results combined for the final objective evaluation [164]. This seems not likely to cause problems

[1]This modifications turned its use almost practically infeasible, since the processing of some videos usually take a number of days of continuous processing.

relatively to tracking, since along the frames, all targets are present and moving together, sometimes very near from each others, and FragTrack could easily distract and switch among them even with only one instance running at a time - what in fact occurred sometimes.

The CPMD-RPT method is implemented in MatLab, and is available for download [209]. It has its own initialization strategy in that it combines FPDW-based detection, validated by background removal with ViBE (as explained in Section 5.8). FPDW tolerates some superposition among targets, differently from GMM, ViBE or $W4$. FPDW was available for download [210], and ViBE - although protected by different patents -, has a version for download also available from [211]. But we did not have to deal with this, since CPMD-RPT method has both FPDW and ViBE already embedded as third parties. Therefore, as the CPMD-RPT method is a complete/stand-alone implementation that does not depend on initialization based on extra-methods, it was used directly.

An important fact is that the first mentioned tracking methods (Camshift, FragTrack and TLD), could also have been initialized by the same strategy used by CPMD-RPT: FPDW validated by ViBE, or even by FPDW only. But we noticed, in our video database, the common situation in which people enter Argonauta's Room very near each other, and experience substantial occlusion, such that one of the targets is commonly lost soon after some frames, then appears again, and so on. As we were aiming at testing those methods by initializing them only once, at beginning, to evaluate them as similarly to as implemented by their authors, this could cause premature targets death, which would not be re-initialized - unless we introduced, at that stage, modifications to those methods to improve them with re-initialization capabilities. Therefore, we chose to initialize them with GMM, and had to wait for separate ROIs for this.

Figures 8.1 - frame number 870 -, to 8.3 - frame number 958 -, show examples of some of the evaluated frames[2]. The marked GTs are shown in red, while GMM detections are shown in yellow. In the frame shown in Figure 8.3, the separate ROIs that arose were used to initialize the tracking methods, which were run from then on.

As the CPMD-RPT method re-initializes targets many times, it was difficult (if not infeasible) to represent its results together with the other methods, due to color confusion. First results presented in the next figures are related to those other methods: GMM-only, Camshift, FagTrack and TLD. Figure 8.4 shows tracking results for some representative frames of this sequence, which followed after frame number 958 (shown, in turn, in Figure 8.3). The color legend is:

[2]Frame numbers are shown in the bottom left part of each figure.

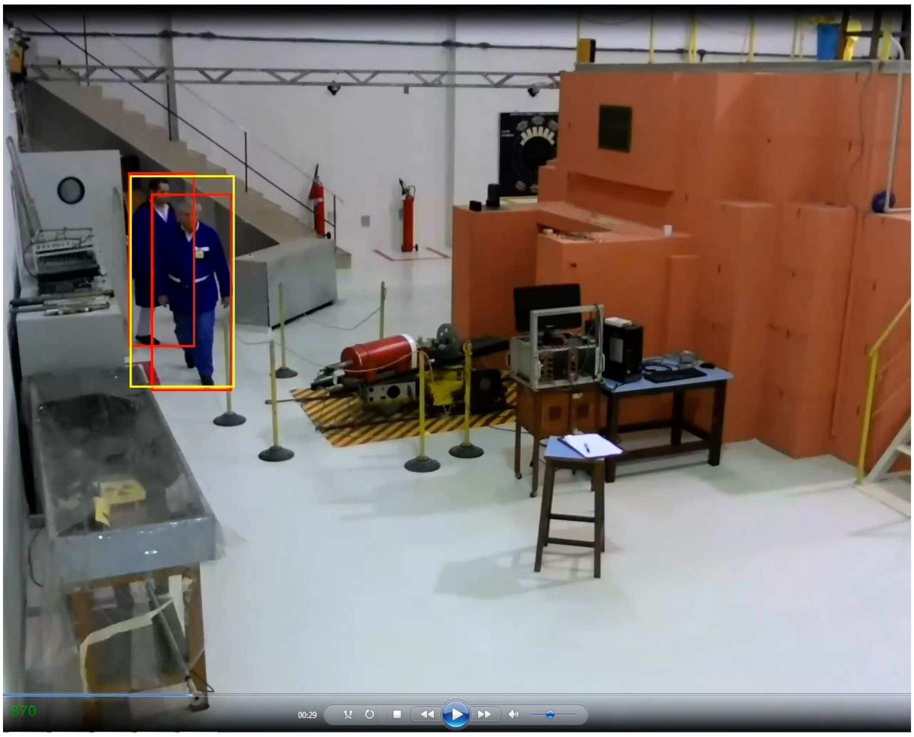

Figure 8.1: Frame 870. Both persons with substantial superposition; only one GMM detection.

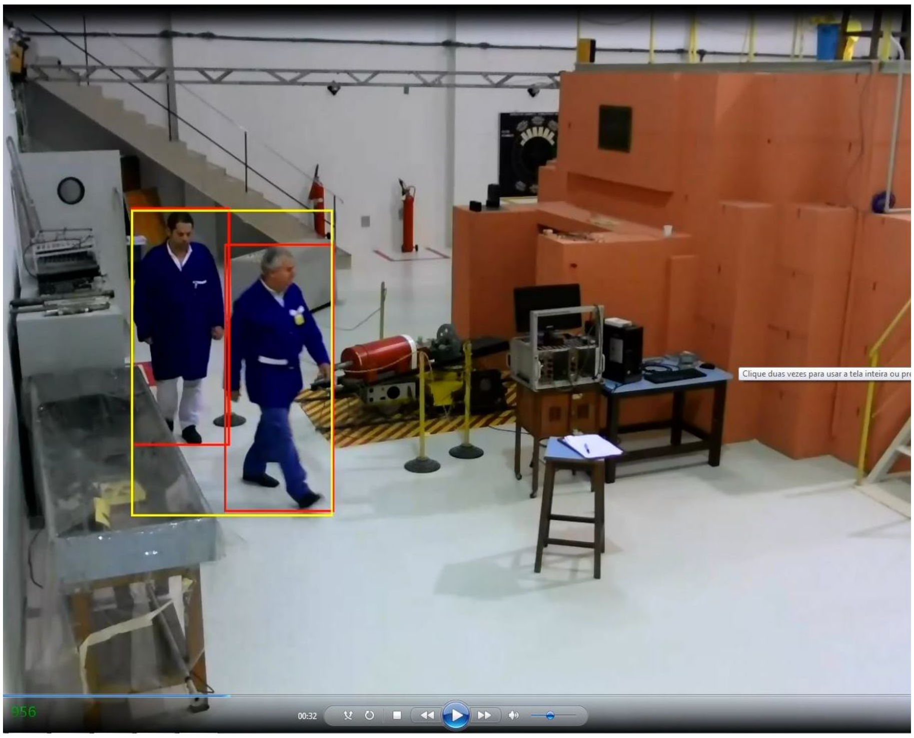

Figure 8.2: Frame 956. Both persons with almost no superposition; GMM still detects only one ROI.

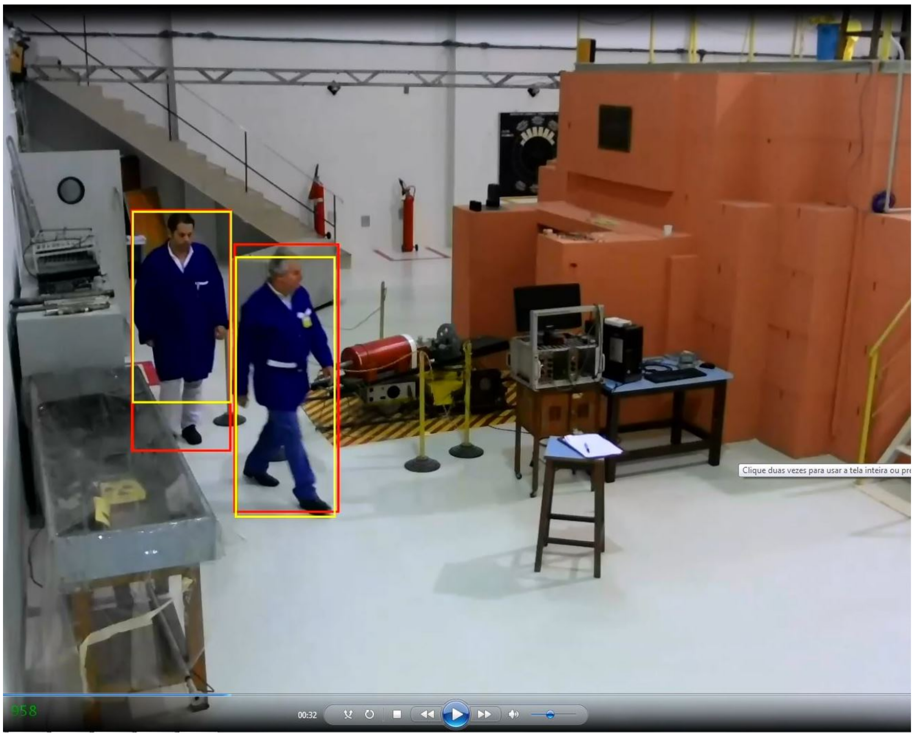

Figure 8.3: Frame 958. Both persons sufficiently apart to generate two separate GMM detections that initialize the first three evaluated tracking methods.

- Red: GTs;
- Yellow: GMM detection;
- Blue: Camshift tracking;
- White: TLD tracking;
- Green: FragTrack tracking.

Simulations showed - even by just a first visual inspection - that both Camshift and TLD had the worst results. For Camshift, this was not so a great surprise, because it is the simplest method among the others. It tracks the whole target ROI, and thus has low robustness to occlusion. But we did not expect it would have such a bad performance, drifting too soon. In this simulation, drifts with Camshift were sometimes so bad that tracked areas decreased or increased unrealistically, and/or got stuck over dark regions (as shown by Figure 8.4); drifts can be noticed since the second frame show in this figure (frame 995).

It was noticed, already in former tests, that Camshift behaved relatively well when targets moved against backgrounds with good contrast with targets colors, but loosing some targets parts for similar colors between background and at least part of the targets colors. For example, when moving against the light-color background

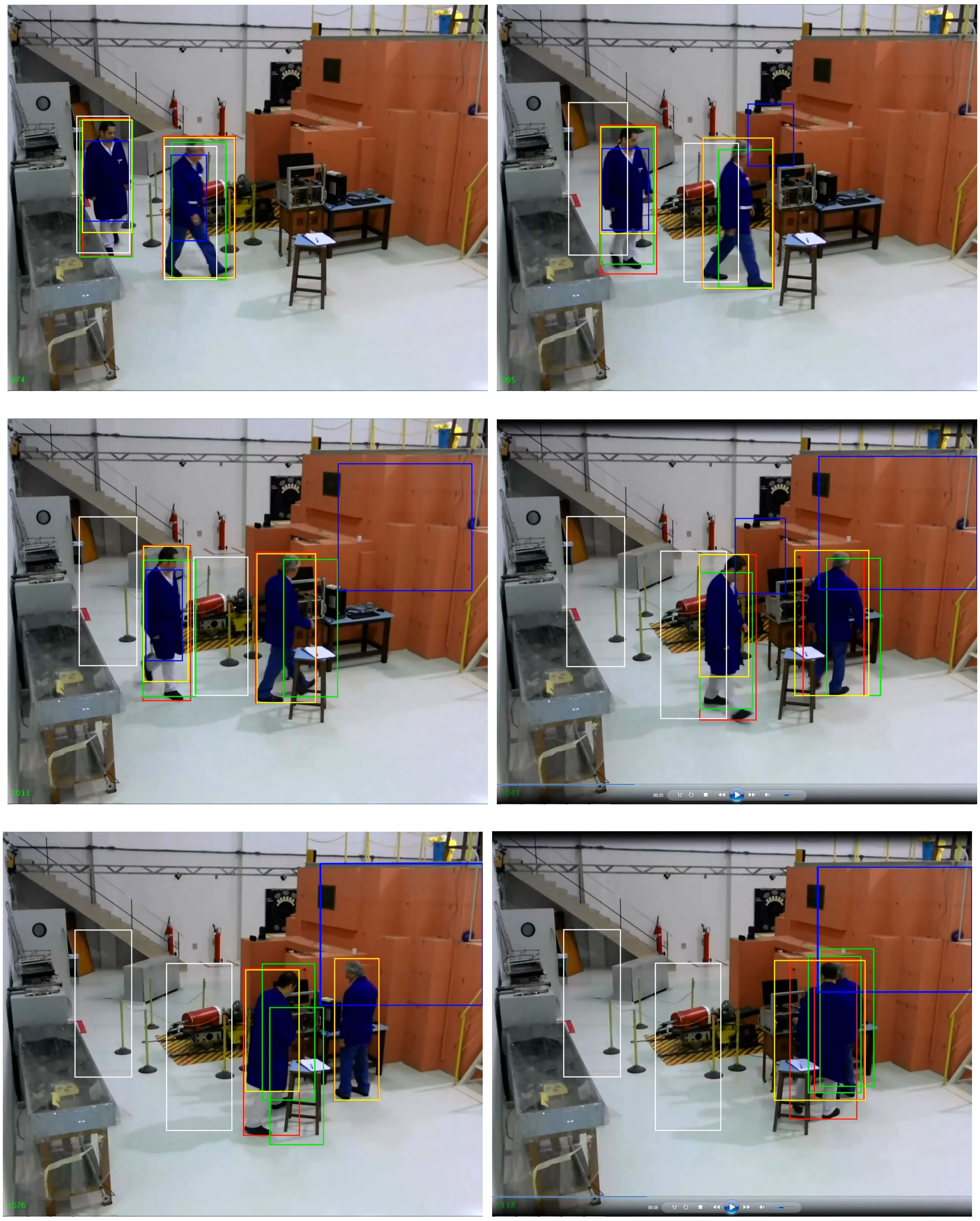

Figure 8.4: Tracking results in the image plane. From left to right, and from top to bottom: frames 974, 995, 1011, 1043, 1076, 1118. Red: GTs; yellow: GMM detections; blue: Camshift tracking; white: TLD tracking; Green: FragTrack tracking.

of the floor, the dark clothes were well tracked, whereas not the light color-ones. It also showed low robustness to clutter, when targets passed against the dark color of Argonauta walls, or other dark regions. These former tests were performed with videos recorded by other former cameras, with lower resolution, and showed the need for Camshift frequent re-initialization by GMM, by monitoring abrupt tracked ROIs locations or areas variations, and some area variations were due to color confusion between target and background. These former results were obtained before the

Qualifying Exam, and can be found in [212, 213].

A greater (bad) surprise, though, was related to TLD that, according to its authors, has some capability to adapt to targets appearances changes. But when, for our video database, people made curved trajectories, even slowly varying their poses, TLD drifted by getting stuck on some stopped positions, not following the targets anymore (as can also be noticed in Figure 8.4), beginning with a small drift just from the frame 995.

FragTrack and CPMD-RPT performed better, tracking targets for more time. But, as long as: (i) people stay stood still at some places, together, very near each other; (ii) sometimes with superposition - either low or severe -; (iii) with clothes of similar colors - by using upper bodies uniforms -; and (iv) against dark background clutter. They also drifted somewhat and also lost targets for some times.

To perform comparative analysis for those other methods (Camshift, FragTrack and TLD), associations were made between the detected and tracked ROIs with the GTs, in the initialization frame (number 958) for the whole bodies centroids distances in the image plane; the developed evaluation code paired tracked ROIs with the nearest GT ones, to enable and assess tracking along the video. Another measure was the ratio of intersection to union of each tracked ROI to its corresponding GT: the Jaccard distance [214]. This enables detection of unrealistically increasing tracked ROIs, even if they have good intersection with their corresponding GTs. Further, targets positions both in the image plane and in the WCS were evaluated. Most of these comparative analyses are commonly made by computer vision experts, and were performed at least initially to assess the methods performances and to make decisions about which methods would be discarded or not.

But, important to mention, the final and most important results supplied by the system proposed in this thesis, is comparative analyses in terms of received dose along paths, for each worker, instead of simply the trajectories in the WCS - although the estimated received doses depend directly on this latter result -, or in the image plane; and even less in terms of centroids distances and intersection to union ratios in the image plane. And this dose-based evaluation metric - particular to this thesis -, should be done by computing dose for trajectories tracked, by each method, in the WCS, comparing them with the corresponding GTs computed doses. In that, re-identification plays an important role for dose accounting, besides trajectories accuracy, to associate correctly the received doses for the right workers.

As the first evaluated tracking methods enable to associate each tracked target with its corresponding GT, it was easier to verify errors as: (i) drifts; (ii) targets losses; or (iii) switchings among them. But for the GMM-only and the CPMD-RPT, analyses were performed separately. GMM simply detects targets, but without knowing who is who; it served most to tasks as initializing tracking methods, for example. And CPMD-RPT changes targets IDs many times; when re-initializing lost or badly tracked (drifted) targets, it resulted in an increasing number of labels along the video, even when only two persons were really participating in it - for instance, with this first evaluated video with only two persons, after some running term, the ID number eight arose soon. Therefore, the analyses of CPMD-RPT had to be performed separately, and strongly supported by visual inspection.

At this point, both Camshift and TLD had already been discarded, due to their poor performances for this video database - and TLD, also considering its slow processing[3].

Although FragTrack was one of the best performing methods, it was noticeable that it drifts sometimes. When both persons remained standing still, or performing few movements, as near the computer desk, FragTrack resulting ROIs drift in some frames, both approximating to the same person's position, increasing the distance from the drifted ROI relatively to the corresponding GT. Figure 8.5 shows some representative frames of CPMD-RPT tracking, taken specially at the beginning of the video, in which severe occlusions occur - a characteristic and difficulty of this video. Thus, targets are lost, drift results as well as IDs change[4]. But after passing trough that part of the video, things get better, until the two persons become standing still (or almost) near the computers desk; or when one of them goes to the region in front of J9. As already mentioned, CPMD-RPT re-initializes a lost or badly tracked target with a new ID, not attempting to re-identify them.

This method seems that had been developed dedicated towards solving occlusion situations (even severe ones), and performs very well in that, by using its patch-based tracking strategy, and occlusion is a very important issue for targets tracking. This method tries to discriminate among targets by the color-based parameters of its patches during tracking - and it works very well for many cases of general videos of people walking in public places. But it always gives new IDs, not trying to re-identify them with previously terminated tracks, or between a person that goes out and returns to the scene. Thus, summarizing these comments: re-identification is not part of it - at least as supplied by its authors up to the version used.

Figure 8.6 shows also some challenging situations, where the two persons are very near each other, and problems as drift, targets loss and IDs switching occur. CPMD-RPT relies on: (i) targets movement prediction most of the times; (ii) on color-based tracking - even using patches -; and (iii) also on detections. Thus, when people are sufficiently apart from each others - or simply crossing among

[3]Of course, if it performed well for this work, it would be considered, even with slow processing. This could be solved later by any means, as porting it to more recent operating system and software versions, or even to C.

[4]IDs are not associated with colors; but are rather shown by their numbers, approximately at the center of each tracked ROI.

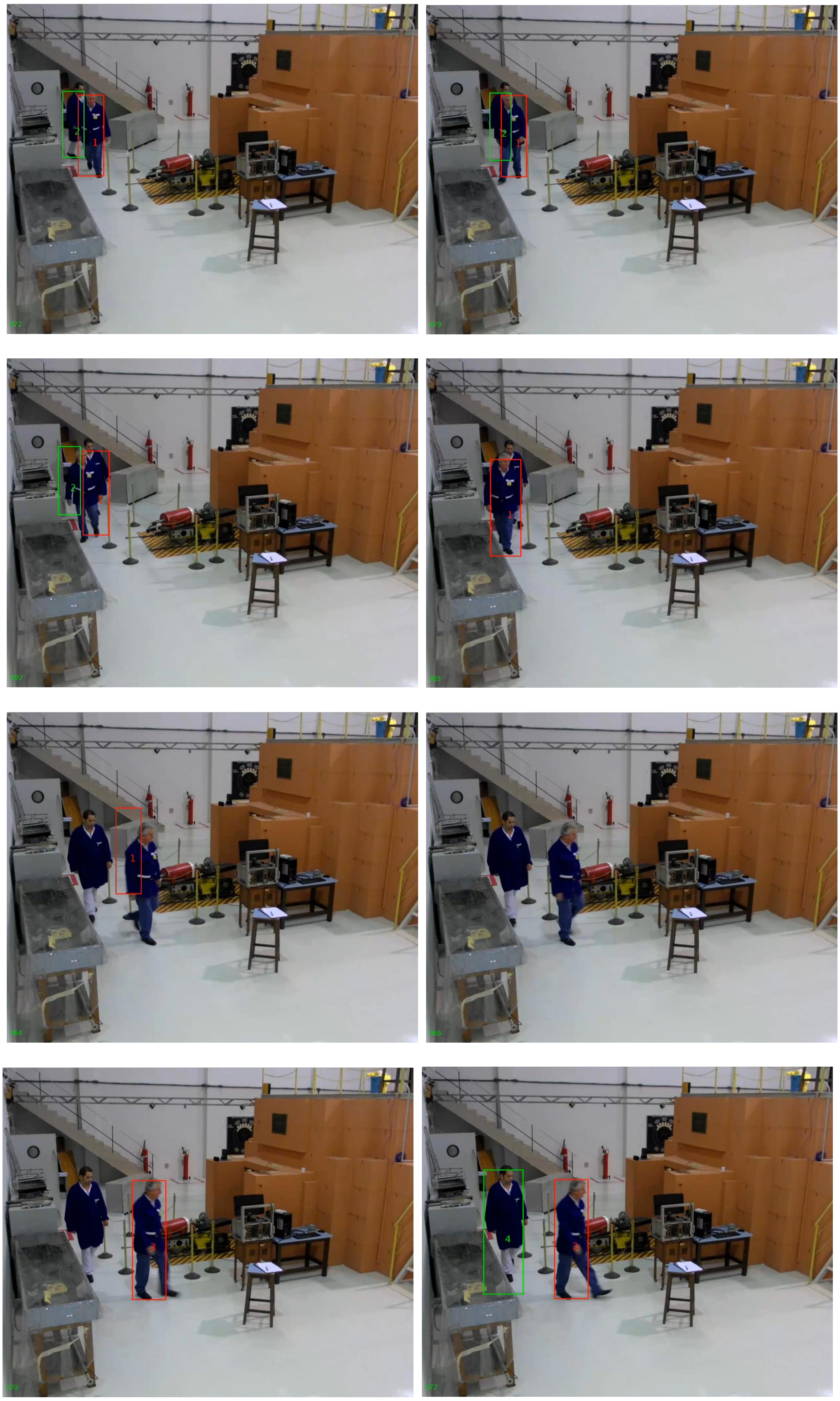

Figure 8.5: Tracking results by CPMD-RPT. From left to right, and from top to bottom: frames 872, 879, 892, 905, 964, 966, 970, 972.

themselves -, CPMD-RPT authors had already proven their method is robust against occlusion and distraction among different persons, with benchmark videos. But here, people almost standing still and using clothes of similar color, poses somewhat harder difficulties.

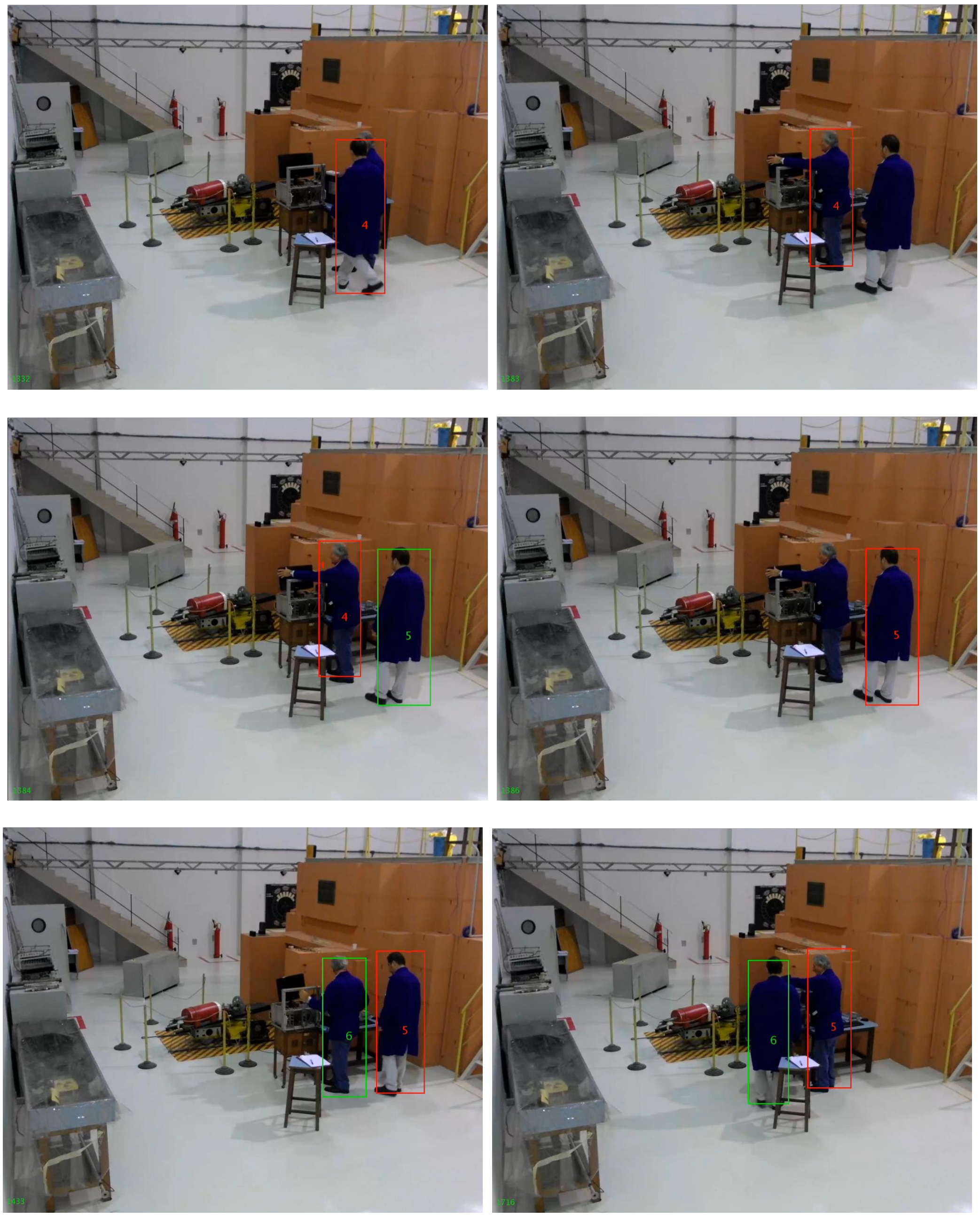

Figure 8.6: Tracking results by CPMD-RPT. From left to right, and from top to bottom: frames 1332, 1383, 1384, 1386, 1433, 1716.

Sometimes, it would be possible to correct targets IDs, by considering their WCS positions: when IDs switch for two standing still targets (for example), their WCS positions at subsequent frames could be used decide to undo the switching.

The left camera view could also aid in some situations, as occlusions occurring relatively to one of them, may not occur in the other view.

In summary, problems occurred mainly for:

- Severe long term occlusions;
- People stood still for long term very near to each others;
- People crouching in J9, even more when being outside the line if sight of a camera, occluded by some other person.

It has to be stressed out that these severe problems would be expected to distract and mislead any method. No method seems to be able to: (i) track well a severely occluded target, even more for long term - including those above mentioned cases where targets remain just behind another one that, unfortunately, remained exactly, or almost exactly, in the same line of sight of the camera -; or (ii) targets of similar colors standing still very near each other for long term; still (iii) it is also not likely that methods deal well, in principle, with targets varying so much their appearances as when crouching.

As even the evaluated methods with the better performances generated tracking or re-identification mistakes, these results analyses should be made by assessing both the tracked positions together with visual inspection, for the following reasons:

1. Each moment a specific method lead to an error, it should be annotated the corresponding situation for which that error occurred (be it drift, targets loss or targets switching errors);
2. For each moment an error occurs for a given method, it should be verified if another method behaved well in that same situation or not;
3. For each error that occurs from a given camera view, it should be verified if the other view at that same moment led also to errors.

First item would give information about the weaknesses of each method to given situations, as: (i) occlusion (partial, total, short term or long term); (ii) distraction due to color (among users, or between a target and clutter objects or same-color background); (iii) people crouching at some place.

Second item would show if another method could compensate for a former method failure, or whether some situations could weaken all methods simultaneously.

Third item would show if some errors were due only to an issue related to a camera bad view, as total occlusion, and whether it could or not be compensated by the other camera view.

8.2 Proposed Methodology

Given the considerations made in the last section, the proposed methodology can involve one or more of the following strategies:

1. To combine results from the best performing methods, as it was supposed that each one would probably fail in different situations, due to their different operating characteristics and principles;

2. To combine results for the two cameras views, since it was supposed that, when somebody becomes occluded in one view, that person could possibly not be occluded in the other one;

3. To propose strategies for targets re-identification, based on different approaches;

4. To try using heuristics to consider targets birth and death only at specific regions, as at the Argonauta entrance door or at the right lateral boarder, to avoid false alarms and targets loss.

It is possible that both cameras views present problems at the same scenes intervals, and for all methods used. But it was expected that, if it occurred, it would be for few intervals, in such a way that the combination strategies could work well most of the time.

The first item was discarded. First, due the high time consuming characteristics of FragTrack. Also, due to the fact that CPMD-RPT, besides losing targets IDs many times, loses also the trajectories for some not negligible terms in some parts of the videos, and it also considers targets movement for its tracking approach, what is not the case in many parts of the generated video database. Therefore, a new methodology was proposed.

8.2.1 Re-identification by Previous Positions

This is the simplest and most obvious re-identification strategy. This thesis does not consider any trajectories prediction. Tracking based on movement prediction tends to work well for people who keep walking most of the time, as in scenes recorded in common public situations; but may not work so well to track people, discriminating among them, for those who remain stood still for long term, as occurs with some relative frequency in this video database; as trajectories prediction may not supply useful information in these situations.

Therefore, a first choice was to use simply human targets detections frame by frame, without any explicit tracking; combined with association of detections in subsequent frames by the nearest positions detected before, in the WCS. But this can be done only if targets remain relatively apart in the WCS, to avoid positioning mistakes, and thus targets switching among them.

When targets get very near each other, all methods tend to fail. In the situations where targets are very near from each other, their positions are considered to be, in the proposed solution:

- If some targets are lost such that only one target remains detected, the same position may be assumed for the lost one;
- If targets are detected, but it is not possible to re-identify them due to high proximity or superposition - neither from previous positions nor from extracted parameters -, their positions still can be averaged and assumed the same for all, or at least for some groups not identified.

When targets become apart from each others again, association based on their trajectories might not work at all, as already commented in the last section. Considering also that people may be using uniforms, the methods evaluated - and all of them make use of color, even in patches - color-based re-identification for targets bodies could not work well too, because of high parameters similarity.

Therefore, two strategies were adopted for re-identification in these cases:

- One based on parameters extracted from the faces/heads ROIs; and
- Another one based on parameters extracted from the bodies ROIs.

8.2.2 Re-identification by faces/heads parameters

Taking into consideration the difficulties already stated to re-identify targets in this video database, the persons faces and/or heads seemed to be a good choice to distinguish among them. Our first attempt was to compare their faces/heads ROIs between subsequent frames as much as possible, due to the fact that no one, in general, changes his faces/heads pose so abruptly in subsequent frames - in a frame rate of 30 fps. Thus, it seems most likely that similarity - for the parameter chosen to be evaluated - should be higher between faces/heads of the same person between two subsequent frames, than among different persons. This could aid in solving ambiguities and undesired trackings switchings.

But, detecting, tracking and/or associating faces/heads regions are not so easy tasks, for freely-varying poses. The most common methods used to detect and/or track faces - as the Viola-Jones [118, 119], for example -, have few robustness to pose changes out of plane. They are rather trained towards detecting faces in some pre-defined poses, as in frontal ones; and fail easily when someone changes poses in 3D. Training such systems with different 3D poses samples may not work well too. It is more common that trainings are performed for different situations, as: (i) a pre-trained system for frontal faces; (ii) other for lateral faces views; or (iii) other ones trained to detect limbs; among other examples. Other methods for face recognition or tracking based on specific faces regions or points, require that faces images are sufficiently in close view, which is not the case in this video database; and may also fail for lateral or back faces/heads views.

First tries in this new proposal, considered simple geometrical relations between a person's body and his face/head region, as described next:

1. First step is obtaining a subregion comprised by the upper one sixth of person's body height;
2. Then, within this region, extracting the central third part, where the face/head is most likely to be located.

Preliminary tests were performed by extracting manually persons faces/heads ROIs, for a number of frames. Then, an approach for color-based parameters extraction from these ROIs we chosen, and showed encouraging results. The parameters extraction from the faces/heads ROIs were the color-based parameters used by CPMD-RPT: by using the CIELab color space, and the Battacharyya distance between histograms. But other color representations could have been used alternatively, as the HSV of Camshift, and with other metric distances. Further, other parameters may be used, instead of color, as HOG, for example. Some of them may even be extracted to compose a voting system. But, in fact, HOG parameters were evaluated and showed not to be so discriminative as the color-based chosen method. Also, SURF parameters - among other ones -, were also evaluated and led to problems, since - although the frames are in HD - faces at distance result somewhat blurred, with badly defined contours, when considering only faces/heads ROIs.

But, although the encouraging results obtained with manually extracted faces/-heads ROIs, in a real operating system these ROIs must be obtained automatically. FPDW was chosen, since it performs well, in general, to detect a person's body ROI with good accuracy. The proposed approach is shown in Figure 8.7, already considering real bodies detected from video, instead of fictitious typical bodies illustrations.

This re-identification approach showed good performance when persons were sufficiently apart from each others, even when they performed curved trajectories: the models adapt from frame to frame, and it was expected that it would perform the comparisons between the current frame with the one immediately before.

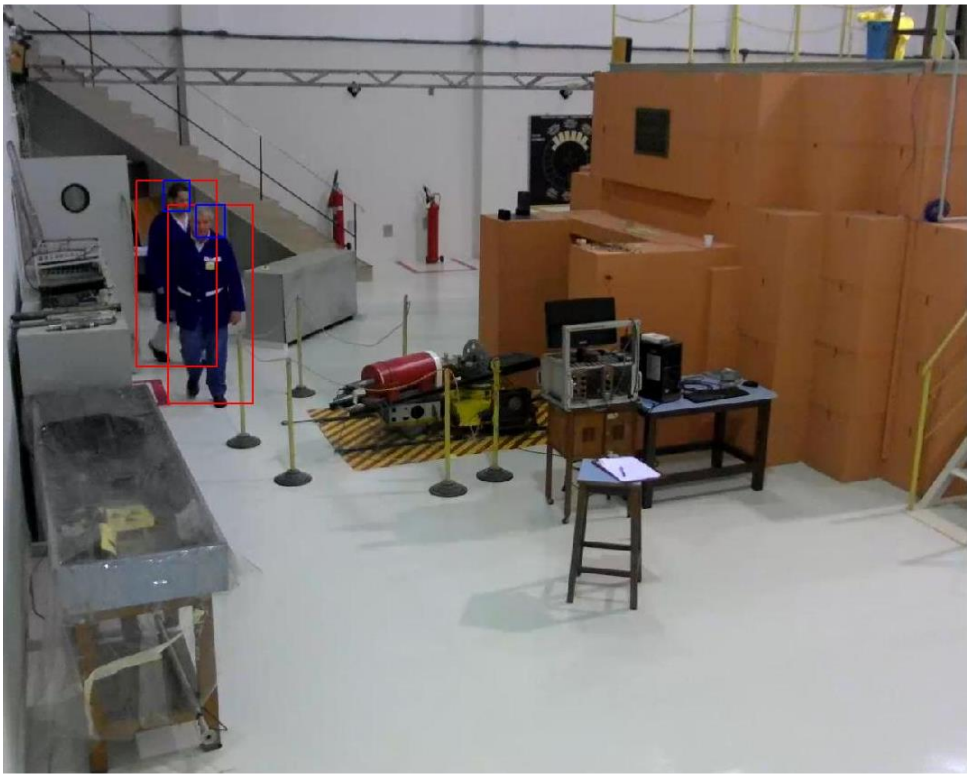

Figure 8.7: The geometrical relation to detect faces/heads given whole bodies ROIs; the latter shown in red, and the former in blue.

But later, some problems arose. First, relatively to the simple geometrical relation:

1. Sometimes a person moves his face/head in a way it gets out of that central third part within the sixth upper region; this may include too much background pixels information, and adapt the model in a wrong way.

2. Sometimes, the ROI obtained by FPDW diverges somewhat from the person's body real dimensions, maybe resulting bigger or smaller than it, and thus the face/head ROI may result also badly located, and background pixels information also included in the model.

3. There are frames where somebody becomes very occluded (almost totally if not totally), in which cases it is not possible to detect and extract any characteristics from faces/heads regions; sometimes because one of the targets is simply not detected.

Figure 8.8 shows examples of the two firsts listed issues. Figure 8.9 shows an example of the third listed issue.

To cope with the first two issues mentioned just before - that leads faces/heads getting out of the third horizontal upper region -, a strategy was proposed by com-

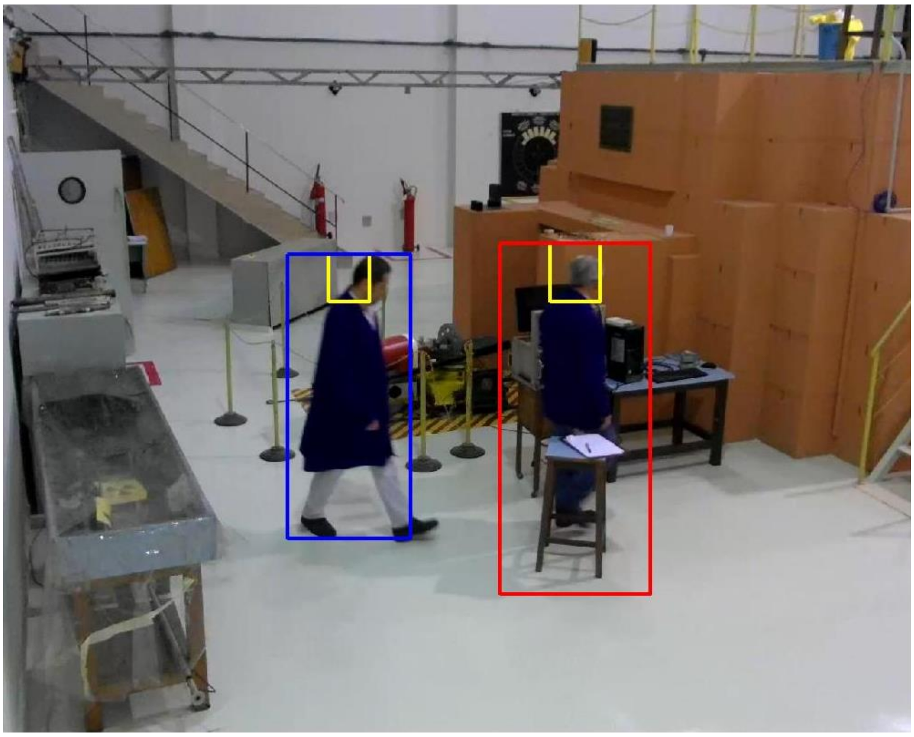

Figure 8.8: For the leftmost person: badly located head due to person's movement by inclining his head; for the rightmost person: badly located head due to a body ROI bigger than his body.

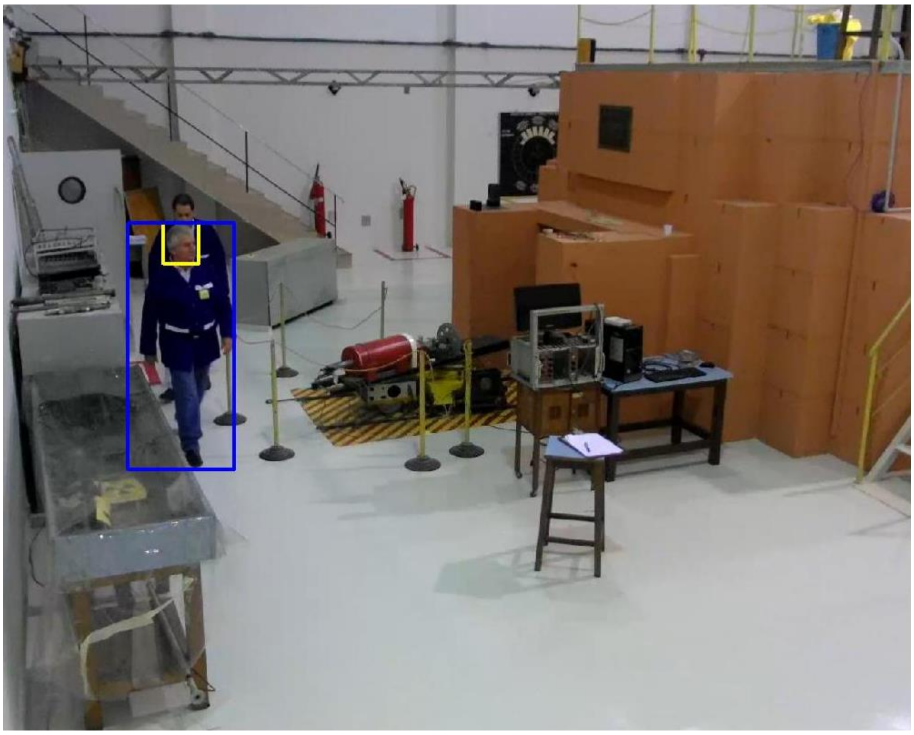

Figure 8.9: One head not detected due to a target lost by FPDW.

bining detections by FPDW with foregrounds obtained by a background removal method (such as ViBE, or another one), as explained in the following:

1. At a first step, a subregion comprised by the upper one sixth in height of the body ROI is obtaining, as before;

2. Then, the foreground within this sub-region is considered for defining a rectangular region to better locate the face/head and extract the chosen parameters.

In the second step, the face/head foreground supply the minimum and maximum x coordinates in the image plane, as well as the minimum y coordinate[5]; the maximum y coordinate in the image plane is given by the upper sixth geometrical limit.

Since ViBE is a time consuming method - background removal usually takes around 20 s per frame, to process -, we adopted alternatively the $W4$ method, with shadow discrimination [86–88]. This latter can be used in indoors environments, such that in which videos were acquired, and processes very fast, giving results immediately.

Figure 8.10 shows this processing. The dislocation of the face/head ROI is clearly noticed for the rightmost person.

Despite solving that former problem, another one arose: when two or more persons get near each others, sometimes their faces/heads foregrounds may be dislocated in such a way that become fused among themselves, enlarging unrealistically the regions in which parameters should be extracted, which would not correspond neither to one nor to the other faces/heads, or may include other regions corresponding to the other person's body, not only his head one.

Therefore, all the faces/heads parameters extraction is performed only when these regions are sufficiently apart from each other - and this must be evaluated in image plane coordinates, not the in WCS ones. Imagine two persons sufficiently apart from each other in the WCS but, unfortunately, almost aligned to the camera view in use. Although apart in the WCS, their faces/heads foregrounds might still experience fusion, due to high proximity in the image plane. Therefore, those parameters are extracted and saved frame by frame, only if persons are sufficiently apart in the image plane coordinates. Further, this evaluation is performed only for the upper sixth sub-regions: people may have partial - even severe - occlusions on their bodies, but with faces/heads still sufficiently free to be processed; or also the counterpart problem may occur: bodies may be not experiencing occlusion - or at least severe occlusion -, but faces/heads regions may still be.

[5]In the image plane, the y coordinate grows from the top to the bottom of the frame.

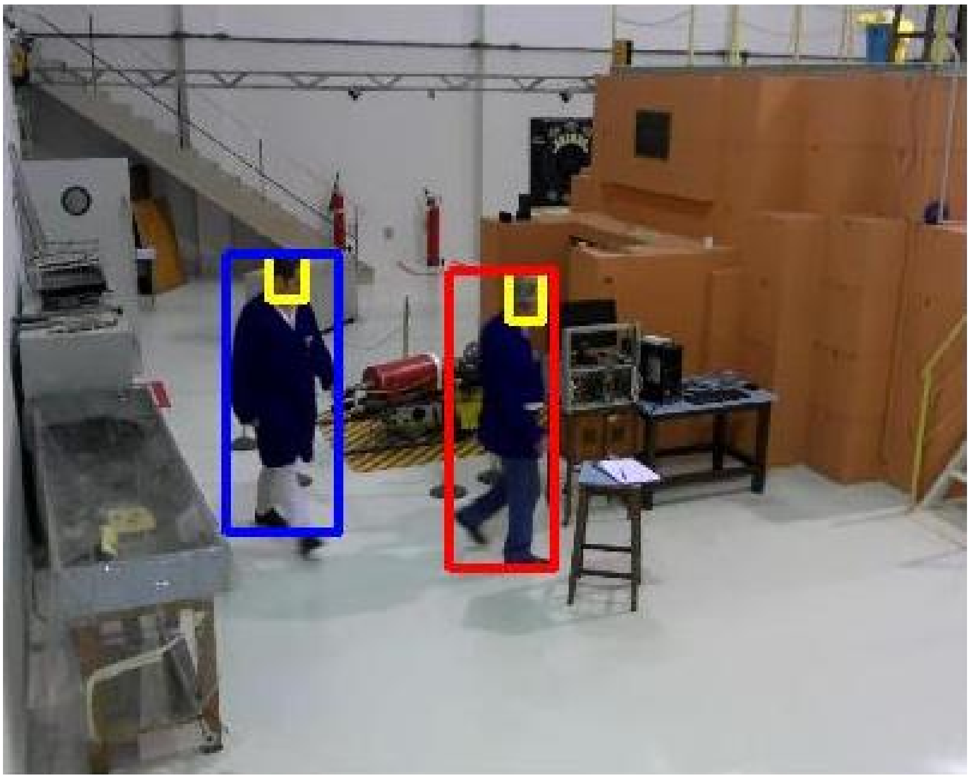

Figure 8.10: Best located head by combining FPDW detection with $W4$ background removal.

Then, the sixth upper body region of each person is evaluated against the whole body ROI of another near one. If this sixth upper part is sufficiently free - above a threshold -, face/head ROI is obtained by combining the geometrical relation with the background removal, and color-based parameters extracted; otherwise, not: it is said the person is in a prohibitive situation, and no ROI nor parameters are obtained, so as to not extract other persons' information as it were from the one under evaluation. Figure 8.11 shows such a situation.

Summarizing what was explained until now relatively to faces/heads analyses and re-identification in general:

1. When people are sufficiently apart from each other in the WCS: re-identification is performed by associating the nearest corresponding positions in the WCS from frame to frame;

2. When people are sufficiently apart from each others in the image plane - and taking into account all those considerations of mixed solutions for geometrical and not occluded faces/heads foregrounds - these ROIs parameters are extracted and saved from frame to frame, even when not used, but distances instead;

Figure 8.11: The foremost person is in prohibitive condition, while the other behind, not.

3. When people get apart after being very near to each others in the WCS, re-identification is performed by faces/heads parameters.

Thus, re-identification based on faces/heads parameters is not intended to be used alone; it does not aim at being an alternative method to the one described is Subsection 8.2.1, but rather combined with it. This combined re-identification strategy is shown the algorithm 4:

Combining FPDW targets detection with foreground extraction brings one more advantage, besides better faces/heads locations. As FPDW sometimes returns ROIs that do not correspond exactly to a person's location, the minimum and maximum horizontal and vertical bodies' foreground limits are used to modify the ROI. This, in turn, besides helping even more in the faces/heads locations - for the cases where the FPDW ROI is well above one's head -, also results in a better location of persons in the WCS - for those cases where the FPDW ROI is well below a person's body.

8.2.3 Re-identification by Bodies' Parameters

After evaluating different parameters for the faces/heads ROIs, without much success than the color-based one, a tentative was made for the whole bodies ROIs,

Algorithm 4 Re-identification strategy based on faces/heads color-based parameters and WCS positions.

for $EachFrame$ **do**
    if $Targets\ are\ appart\ in\ WCS$ **then**
        $Re - identifies\ by\ WCS\ past\ positions$
    else
        $Assumes\ same\ or\ averaged\ position\ and\ waits\ unil\ they\ get\ apart$
    end if
    if $Targets\ are\ appart\ in\ image\ plane$ **then**
        $Extracts\ and\ saves\ faces/heads\ parameters\ for\ visible\ ones$
    else
        $Do\ not\ extract\ nor\ save\ parameters$
    end if
    $Priority\ for\ position - based\ re - idenfification\ even\ if\ existing\ faces/heads\ parameters$
    $Swithes\ for\ faces/heads\ parameters\ only\ when\ necessary$
end for

instead. And SURF features extracted from bodies ROIs resulted in a great discrimination capability, and - a bit surprisingly -, even when targets were experiencing severe occlusions. It is well-known that SIFT method - to which SURF has similarities - extracts highly discriminative keypoints, even under severe conditions as occlusions or background clutter. The SIFT author presents, in his articles, examples of matching parameters extracted from objects ROIs to scenes where the same object is "camouflaged" in very cluttered scenes, under in-plane rotation, and with partial occlusions. Due to some conceptual similarities between SIFT and SURF, the latter may be used alternatively to SIFT - and, in fact, it was done. First tests results using GTs ROIs were encouraging, showing the robustness of this method, as shown in the following, from Figures 8.12 to 8.14. For each case, the leftmost image is the GT ROI of one person in a given frame, matched with the subsequent frame shown in the right. It is noticeable that SURF parameters match mostly in the same person's region in the subsequent frame, even in case of proximity or partial occlusion.

The same strategy used for the faces/heads parameters extraction - the combination with background removal - was also implemented here, for similar reasons:

- First, to locate better each person's body ROI, avoiding to extract too much parameters from the background;
- Second, the persons' WCS locations would also be better estimated.

SURF-based re-identification is also used combined with past coordinates in the WCS, when targets are sufficiently apart from each others - similarly to what is done

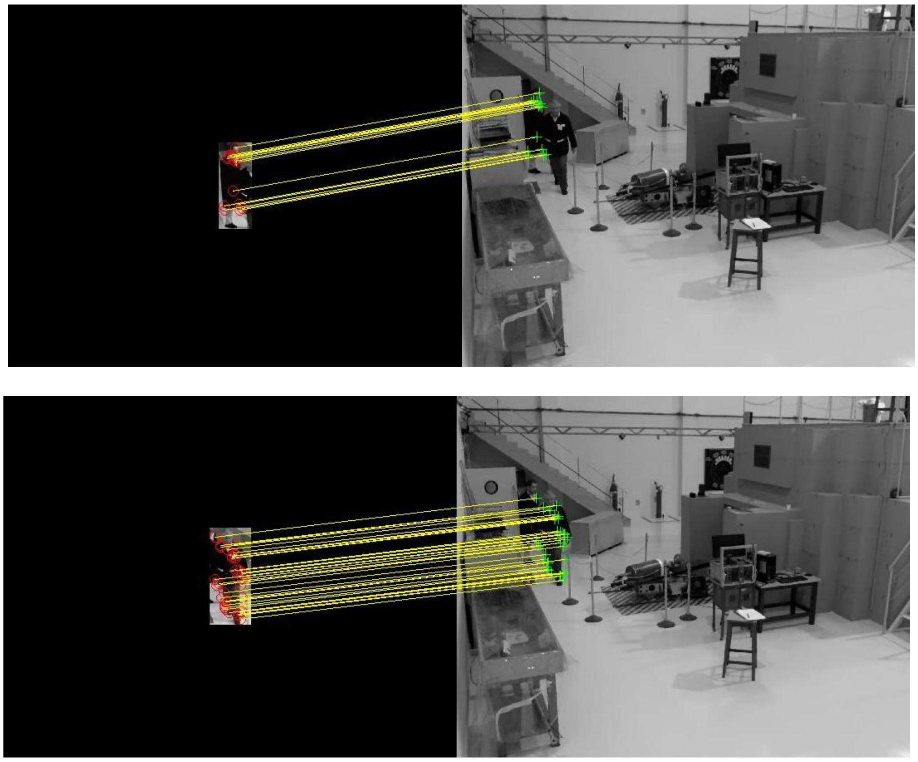

Figure 8.12: Preliminary matching results for SURF parameters, for severely occluded GTs.

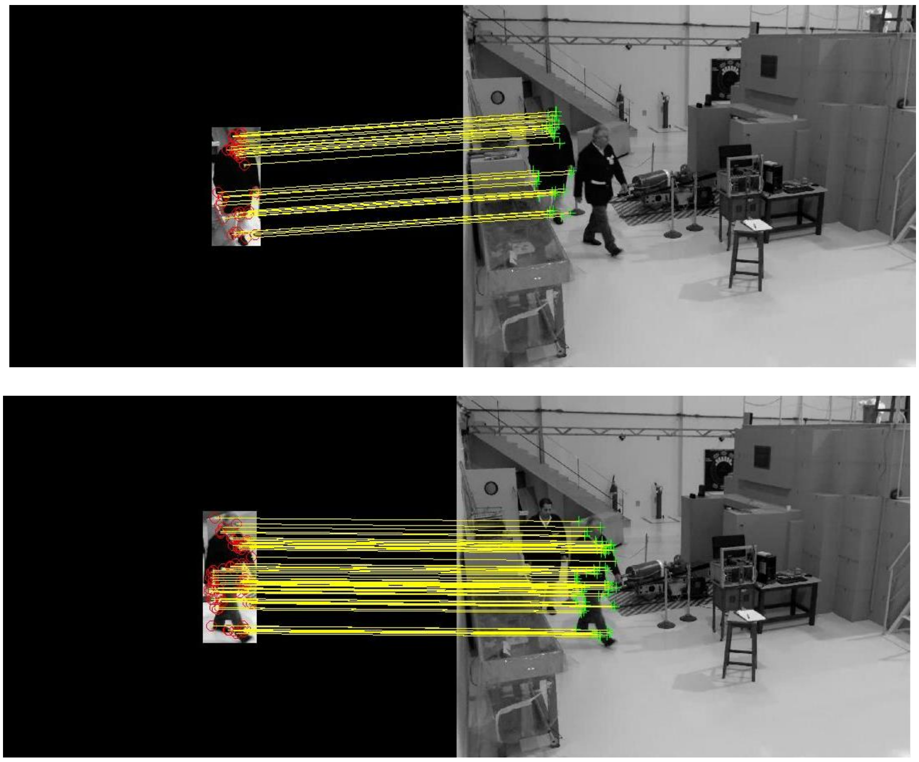

Figure 8.13: Preliminary matching results for SURF parameters, for near GTs.

with the faces/heads parameters. The strategy using SURF parameters is described in the algorithm 5.

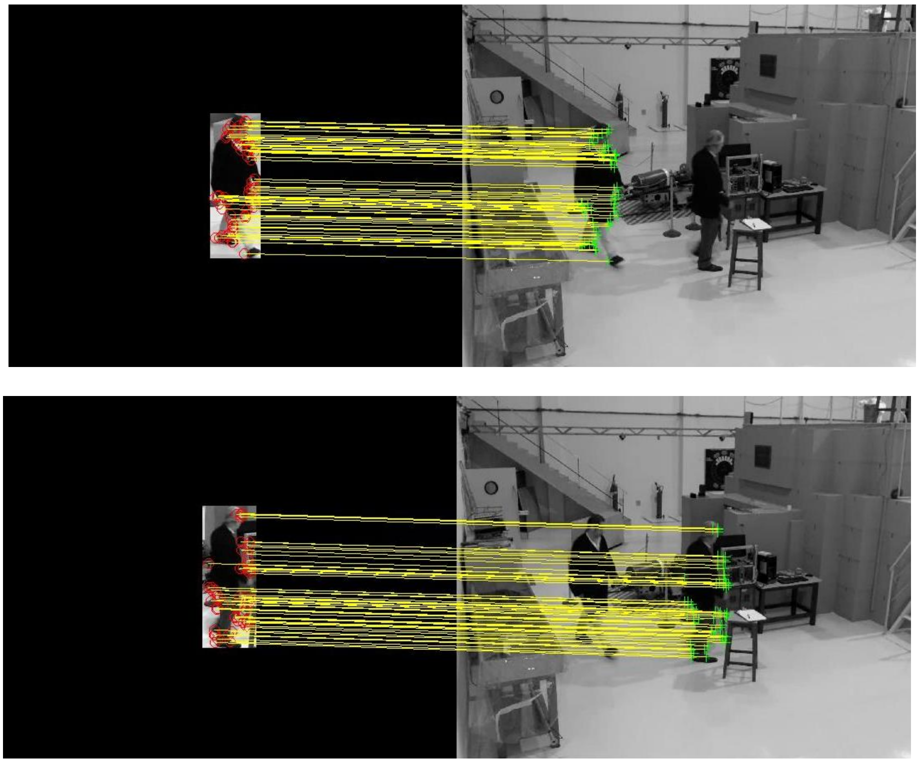

Figure 8.14: Preliminary matching results for SURF parameters, for separate GTs.

Algorithm 5 Re-identification strategy based on bodies' SURF parameters.

for *EachFrame* **do**
    Saves FPDW ROIs from former frame
    Extracts SURF parameters within those FPDW obtained ROIs
    Extracts SURF parameters for the whole current frame
    Matches each previous ROIs SURF parameters with current frame ones
    Detects new ROIs in current frame by FPDW
    Verify to which current ROIx the formerly obtained SURF parameters best fits and associate them
end for

The comparison to decide for this re-identification verifies the percentage of points matched between frames, what resulted more robust than evaluation only the number of matched points. Sometimes, a target is matched between frames with fewer points than another one, what seemed to be important was the percentage of matched points.

8.2.4 Multi-layer Strategy for Matching Parameters

Further, since faces/heads parameters can not be extracted nor saved in certain situations, this re-identification strategy may not ever be performed from the current frame to the one immediately before - for the simple and complicated reason that the faces/heads ROIs may not be visible in that to do any comparison. It also occurs with SURF analysis: sometimes a target is lost, and there is no parameter in the past frame to do any comparison with the current one.

In these cases, a further strategy was adopted for comparing parameters, by searching backwards, in a proposed multi-layer approach, following priority levels, as shown in the following algorithm 6[6].

It is expected that comparisons are made to previous frames most of the time, which is more favourable for re-identification, due to almost none pose changes. But, of course, the greater the interval for comparison - the more frames and levels the algorithm jumps to -, the greater the probability of matching mistakes. But, as long as sometimes there is no information in the immediately past frame, there is no other way, parameters must be searched in past frames.

The Multi-level strategy is detailed for the case of faces/heads color-based parameters, in algorithm 7.

This averaging process for faces/heads color-based parameters adopted in algorithm 7 avoids wrong identifications, for the following reason. Imagine an ID is matched for few frames in the current level, but with somewhat high distances - what suggests a wrong identification. Suppose also another ID is matched for the majority of frames in this level, but with much lower distances - what suggests correct identification. By just summing the distances along the frames, it may happen that the lowest matched distances surpasses the sum of the wrong ones, leading to the wrong choice of the latter. Therefore, the averaged values indicate correctly the correct ID.

The Multi-level strategy is detailed for the case of SURF parameters, in algorithm 8:

For this case, averages could also be done, but as long as the logic is inverse of that of the faces/heads parameters - here the highest percentage means best matches - no average was performed, just a sum. This also results, in the end, in considering simultaneously the percentage and the number of matched SURF points. A threshold is considered for matching, of 50 percent: below this, the system does not consider a matched actually occurred. By varying this threshold, greater importance is given to the percentage or to the number of matched points:

1. If threshold is lowered, more matches will be considered in the sum but, for this same reason, matches with lower number of points will also be considered as right ones;

2. The contrary, in counterpart.

[6] At each level, the frames already analysed in the former ones are not analysed anymore, to obviously not lead to redundant comparisons. Also, the limit of 600 frames that can be alternatively posed to the highest algorithm level, was to avoid an ever increasing time processing for each frame, for long term videos.

Algorithm 6 Multi-layer strategy algorithm for matching parameters: faces/heads color-based parameters or bodies SURF ones

```
if Faces/heads or bodies ROIs are detected in the previous frame then
    (This is the highest priority level for comparisons)
    Do comparisons
    if Matched then
        Re − identifies targets and returns
    end if
end if
if No Faces/heads or bodies ROIs detected or not matched in the previous
  frame then
    (This is the second priority level for comparisons)
    Searches for the corresponding ROIs backwards up to complete
  ramaining thirty frames
    Do comparisons
    if Matched then
        Re − identifies targets and returns
    end if
end if
if No Faces/heads or bodies ROIs detected or not matched in the previous
  thirty frames then
    (This is the third priority level for comparisons)
    Searches for the corresponding ROIs backwards to complete ramaining
  150 frames
    Do comparisons
    if Matched then
        Re − identifies targets and returns
    end if
end if
if No Faces/heads or bodies ROIs detected or not matched in the previous
  150 frames then
    (This is the fourth and last priority for comparisons)
    Searches for the corresponding ROIs backwards ramaining frames up
  to beginning or, alternatively, to a pre − defined limit, as up to ramaining
  650 frames
    Do comparison
    if Matched then
        Re − identifies targets and returns
    end if
end if
```

Tests showed initially good results[7]. But soon some problems arose, not, in principle, supposed to be directed related with this algorithm, nor even less with the parameters chosen, themselves - since previously manually marked ROIs showed those parameters robustness. The problems were more related to misleading detec-

[7]For instance, considering all the problems that may occur, parameters matching in automatic form can only be performed by an algorithm such as the one shown in 6.

Algorithm 7 Details of Multi-level strategy based on faces/heads color-based parameters.

for *EachLevel* **do**
    Color − based parameters of faces/heads ROIs in incoming frame are compared with ones obtained in past frames
    Battacharyya distances are summed along level, separately for each ID
    For each ID, results are averaged by number of matched frames
    Winner ID in current level is that with lowest average distance
end for

Algorithm 8 Details of Multi-level strategy based on SURF parameters.

for *EachLevel* **do**
    SURF parameters in past bodies′ ROIs are matched with incoming frames
    Targets in incoming frame are detected by FPDW
    if *Percentage of matched SURF parameters between each past frame with new detected ROI is equal or greater than* 50 *percent* **then**
        Sum percentages along current level, separately for each ID
    end if
    Winner ID in current level is that with highest percentage sum
end for

tions by FPDW, and these, in turn, are due also partly to the difficult conditions found in this video database, which, then, usually cause:

- Some very severe occlusions with bodies' ROIs highly superimposed, disabling to do any targets associations;
- Some targets losses;
- Some false detections - this latter problem, due, in principle, to FPDW failures itself[8].

If targets detections worked always well in all frames, those parameters-based re-identification methods would do very good targets association, specially SURF - they behaved well with the GTs ROIs. But all the re-identification depends strongly on good automatic persons' bodies detections, and good detections are not expected to be achieved always, as they are obtained automatically by a method that, although considered a very good one, still fails in some frames.

As already mentioned, when more than one target are detected very close to each other, no association can be made: not from past positions in the WCS, nor from parameters matching. Instead, false temporary IDs are given to them. When they

[8] False detections may be due to FPDW itself - of course, worsened by this video database. CPMD-RPT method also finds FPDW false detections, and it manages to eliminate them by validating these detections with extracted foregrounds.

get apart again, the system tries to re-identify them from one or another parameter, depending on the method chosen.

This problem is worse when there are also false detections near the true ones. In such situations, foreground can not be used to discard the false detections, since there exist significant foreground regions also within the false ones, due to the fact that all detections - the true and the false ones - are very near to each others. Although this is a problem, the false detections usually disappear in few frames, and can be discarded at least in a post-processing.

There is a very special situation when two true targets do not have prohibitive superposition between them - and thus, they should be identified -, but a false detection arises between both, causing prohibitive occlusions between itself and each one of the true targets. This originates two groups of non-identified persons - that is, with false temporary IDs.

A much more complicated problem arises with target loss. Suppose there are two true targets moving in the scene. Depending on the distances between them in the WCS - and also in the image plane - they may be identified or not, by position in the WCS, or by one of the chosen parameters. But when suddenly one detection is lost, no group with false ID can be formed, in principle, with the only one remaining target. Then, the developed system tries to re-identify the remaining target by the parameters in use. There are cases where the remaining target's ROI belongs to a person, and the system makes a mistake in associating it with the other person's parameters, for which detection was lost. If this situation remains for a number of frames, the parameter model - be it based on the faces/heads color parameters or on the bodies SURF ones - begins to learn parameters in a wrong form, associating the wrong parameters to the other person, and then targets are switched and probably not re-identified correctly from then on.

We tried to manage eliminating targets losses and false detections by considering problem-specific conditions, as explained by [75]; more specifically, in its Chapter 8: "Context modelling", in which some specific regions can be defined where people are supposed to enter or leave the scene - not necessarily only on its boarders. We tried to enable people entering or leaving Argonauta's Room only near its entrance door, or appearing or disappearing only on the right boarder, or near J9. But further problems arose then.

When a false detection occurs away from the permitted regions, it would be eliminated, what is good. But suppose one true target is lost, away from these pre-defined regions. But since the lost target disappeared away from a permitted region, when it appears again, this context modelling condition would disable to consider it as a true target. We tried to implement it: when a target is lost, and the system tries to re-identify it - sometimes in a wrong way, as explained three paragraphs before -, we tried to set a group, so as the system would not try to re-identify the remaining target. But when the previously lost target appeared again, it would be considered a false detection, and the group would be kept, and no one would be re-identified anymore, nor when separating among themselves. Of course, this is a very bad situation.

Therefore, when a target is lost, we could not find, up to now, any online means of avoiding the system trying to re-identify the remaining target, sometimes in a wrong way; and this may switch among targets IDs.

In the way the algorithm for multi-layer strategy for matching parameters was initially proposed, this problem could be worsened, because it tries to match ROIs (faces/heads or bodies ROIs) preferably to the frame immediately before the current one; or with past ones, giving higher priorities to most recent frames, if not found in it, and so on. If a match is found, the search gets out of the algorithm and a new incoming frame is processed. If a target is lost, and the system tries to match the remaining one with the wrong person - as commented before -, and a match is considered as right, the targets re-identitification has a chance of failure, because the error occurred in the last frame, or in a recent one.

Therefore, new versions of the "Multi-layer strategies for matching parameters" were proposed, to achieve more robustness and avoid failures in re-identitification. The second version tried to consider, in some form, the contributions of all the levels: (i) the immediate previous frame; (ii) the second level comprised by remaining thirty frames; (iii) the third level one with remaining 150 frames; and (iv) the fourth and last one with all remaining - or limited remaining 650 - past frames. The objective was to find matches in *all* these levels - not only for the first match found - and decide for a targets ID by a voting system that can use different weights for each level. It was expected that this voting strategy considered somehow each target history, not only the first match found: if a target's ID changes in a given frame, this change is not promptly accepted, but compared to his past ID history. It is expected that sudden ID changes could be reversed to the correct ones, by considering the past ID history, because the other levels could vote for the correct ID, winning over the most recent change.

But there were still problems in some parts of the video - as when the two persons are almost stood still near the computer desk, and then one of them become totally occluded by the other for long term. FPDW detection of the occluded person is lost, and the system tries to identify the remaining one, and if, in a frame, ID is switched for some reason - as a wrong voting result -, the system becomes also learning parameters in a wrong way, and associating them with the wrong person.

Therefore, a third version of the algorithm for "Multi-layer strategy for matching parameters" was proposed. Here, comparisons are made with all the frames within each level; frames in which matches are stronger, receive higher weights to vote for a winning ID in its level. Then, the winners of each level are further weighted, as in the second algorithm version, to vote for targets IDs. The comparisons made within each level are weighted differently, depending on the type of parameters chosen. For the bodies' SURF parameters, the percent number of points matched for each candidate comprises the weight. Here, the higher the percentage, the higher the weight given. For faces/heads color-based parameters, the Battacharryya distances for all candidates were averaged, to avoid outliers, and each candidate distance was compared to this average to compose the weights. In this case, the lower the distance, the higher the weight.

This seemed to be the more robust form of comparison, to consider each target's history. But the choice of weights for each level alters significantly the system result: it seems these parameters must be finely adjusted by repeating tests.

Anyway, if some target has its ID wrongly switched, a post-processing analysis can be performed where, also based on the WCS, decisions can be done to reverse wrongly switched IDs, ant to give IDs to those targets for which a false temporary ID had been assigned. This strategy was implemented, and is explained in Subsection 8.2.1.

Another alternative is to use the other camera, which can aid in solving for this problem, since the lost target in one view is not likely expected to be lost in the other view simultaneously. And a decision could be made in terms of WCS positions compared between the cameras results: if, in one view, a target has its ID switched, but not in the other one, the system could decide to keep to former ID, if that second view is more confident for that target's WCS region. It would be the case if, in a view, targets become occluded and one is lost; but not in the second camera view: of course this second camera is more likely to give the right results, and this could be used to make a final decision.

8.2.5 Further Analysis to Re-identify Targets

The use of either one of the proposed parameter-based targets re-identification, combined with the re-identification by subsequent WCS positions for sufficiently apart targets, and with the multi-layer strategy for matching parameters, achieved good robustness against wrong targets identification. But, even so, problems may still occur, since no method is perfect. But this possible wrong identification have to be avoided as much as possible, to assign correctly the received radiation dose for each worker.

Therefore, a further analysis was proposed, in that the estimated WCS positions of each person are used to assess the reliability of the given IDs along a recorded scene. Suppose, for example, two targets are walking - or standing still, it does not matter the specific situation in which they can be found. When they are sufficiently apart from each other, priority in re-identification is given for the WCS position-based strategy described in Subection 8.2.1. When they approach each other, re-identification is switched to paramater-based ones (as described in Subsections 8.2.2 and 8.2.3). But these latter are done only for those persons in no-prohibitive condition.

There are situations in which targets are below the minimum distance to be re-identified by subsequent WCS positions, and the system tries to re-identify them by one of the parameters chosen. Although the lastly developed multi-layer strategy for matching parameters, from many cases, turns the system more robust by considering the past ID history of each target, the voting system may still make some mistakes, due to a misleading levels weights combination that gives also a misleading decision for the current ID of a target.

Thus, the proposed further analysis can recover the true targets IDs by assessing their WCS positions in subsequent frames. If the IDs switch, as the WCS coordinates do not normally change much between subsequent frames, this can be inferred as being a false switch, which can be reversed. Targets would not jump simultaneously almost exactly to the other ones' WCS positions in 1/30 s; this is not a realistic situation.

There are other situations in which this further analysis proposed can recover targets IDs:

- Sometimes one of the targets may be identified, and the other not; and then receives a temporary false ID.
- Other times none of them can be identified, and a group is formed and given also a false ID.
- In other cases, they might not had been under no prohibitive conditions at all, but suddenly a false detection arises near one of them, or both:

  - In the first case, the one in no prohibitive condition is re-identified, and the other forms a group and is given a false identify until the false detection disappears;

  - In the latter, two groups are formed, until the false detection disappears.
- Sometimes, a person may just disappear behind another one, and the developed system tries to re-identify him, maybe wrongly.

In the first case listed just before, the proposed further analysis can recover that target ID that was given a temporary false ID, by comparing distance between its current WCS position and the former one, when it had a true ID. Of course this must be done immediately when a false ID is given, because the ID in the immediate past frame must be a true one.

In the second case listed, the false ID results in only one WCS position for both persons. The proposed further analysis can associate this WCS position to both former true targets IDs.

In the third case, two solutions may be given:

- The person which remained with his true ID, remains unaltered; the other who formed a group and was assigned a false ID, can have his true ID recovered, since the other one is known: thus, the one with false ID is surely the other person.
- If two groups are formed, there is no way to know which WCS position corresponds to a person; therefore, these two false-IDs WCS positions are averaged and assigned to both targets.

This latter is an approximation, of course. But one must notice that both persons are very near to each others, and such an average would be a good approximation to compute the received dose for both during this temporary situation, in the absence of other solution.

In the last case listed before (when a target disappears behind another one), a group can be forced, and and WCS coordinates assigned to both targets.

All this strategy is shown in the following algorithm 9.

This algorithm is currently performed offline in a post-processing stage. But it can be included in the existing algorithms that implement the methods combination to perform these corrections online, avoiding the models to learn from wrong parameters.

8.3 Remarks

The proposed methodology is summarized in Figure 8.15.

By closing this Methodology Chapter, some considerations should be made.

- First off all, it should be highlighted that the proposed methodology belongs to the tracking-by-detection class of methods, in which tracking results by detecting the targets frame by frame. Data association is performed between successive (or even not successive) video frames, by different approaches. Therefore, tracking results by combining detection with data-association for targets identification.

Algorithm 9 Further analysis to re-identify targets

```
if In targets below minimum distance − parameters re − identification condition then
    if IDs switched then
        Assesses WCS coordinates of both targets betweeen subsequent frames − cartesian norm
        if Possible to associate WCS coordinates then
            Rejects switch
        end if
    end if
    if One target assigned false ID then
        Assesses WCS coordinates of this target betweeen subsequent frames
        if Possible to associate WCS coordinates then
            Recovers target ID
        end if
    end if
    if Two targets assigned false IDs then
        if One group − one WCS coordinates then
            Assigns WCS coordinates to both targets IDs
        end if
        if Two groups − two WCS coordinates then
            Averages WCS coordinates
            Assigns WCS coordinates to both targets IDs
        end if
    end if
    if One target disappears behind another then
        (System tries re − identification of remaining one by paraqmeters − one WCS coordinates)
        Assign WCS coordinates to both targets IDs
    end if
end if
```

- From the evolution occurring in the computer vision community, it seems the main interest has been directed towards solving more and more the problems due to occlusions; and the developed methods have been each time more well succeeded in this task, by adopting patch-based tracking, for example. But, for this current application, re-identifications is a key issue, to account correctly for each one's received radiation dose.

- The selected methods, based on the bibliographical research, used either stand alone or in combination, solved completely all the problems of the specific application in hand, with this particular video database - even considering that some of them belong to the state-of-the-art methods in the computer vision field, for tracking.

- CPMD-RPT, for instance, gave relatively good tracking results, but does not

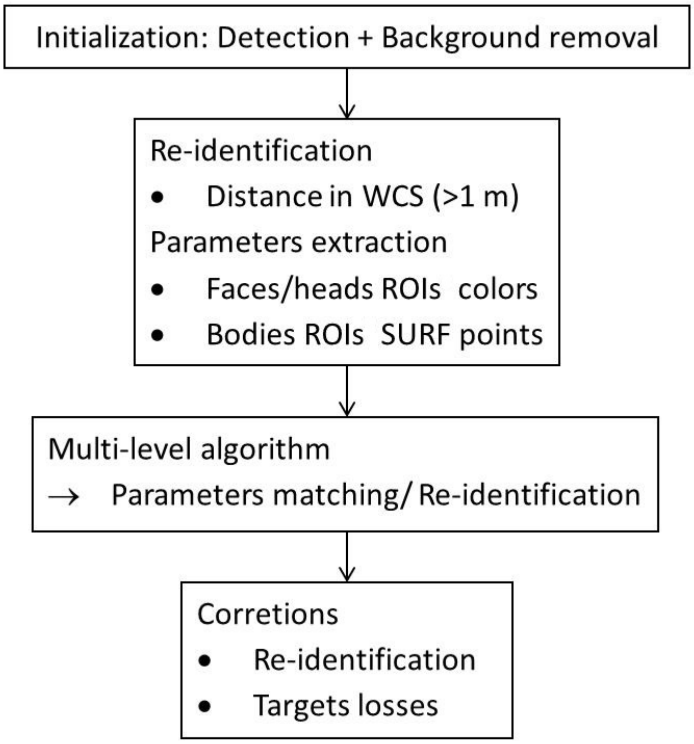

Figure 8.15: Block diagram of the proposed methodology.

aim at re-identifying people. Maybe it can be used in a complimentary form to the proposed ones - as no good method should be discarded -, but after passing through algorithm 9 to try associating lost targets with new birth ones; but we noticed that sometimes a target does not arise again in a number of frames; it should be further assessed if association based on WCS coordinates could give good results in these cases. The proposed methods give detections for almost all, if not all, frames; there is no considerable lack of frames without detections based on FPDW.

- The video database generated for this thesis is very particular, with many specific situations that are not so common in other general video databases available. Thus, some problems that occurred here were due to: (i) color-based distraction; (ii) people standing still very near to each others for some long term - avoiding the use of movement to distinguish among persons -; (iii) people crouching.

Further considerations:

Although first tests with re-identification based on faces/heads color parameters were encouraging, during the implementation of this approach, some problems arose, as: (i) sometimes, bad faces/heads localization due to inclination out of that central third part, or to badly obtained ROIs by FPDW; (ii) sometimes, faces/heads regions fused when using also foregrounds, for somewhat near people; (iii) lack of information for severe occlusion or faces/heads regions, sometimes due to their projection onto other person's body parts; and so on. These problems were approached one by one.

But, since this proposed re-identification method has caused all those issues, simultaneously to its implementation and problems solving, a further complimentary literature review showed some other approaches somewhat similar to the proposed one for head detection/location.

For example, in [215, 216], authors also make use of some geometrical measures, but considering an one seventh upper part of whole person's body. Within this region, they search for heads by combining different parameters, as separate colors spaces for hair and skin - and by considering different hair and skin colors. They represent skin and hair colors in HSV and XYZ spaces, respectively. They also combine SURF parameters with colors. They try to fit an ellipse ROI to the matched regions.

Their purpose is to detect people under severe bodies occlusions, by considering the hypothesis that at least faces/heads should be not occluded in these cases. As seen by their results, they are able to detect faces/heads, but the obtained ROIs do not coincide always well with those desired regions, sometimes loosing some parts of faces/heads, or including other parts of the bodies or the background[9].

There is also a work [217], in which authors make some criticism about that formerly cited one [215, 216], due to supposedly issues related to color-based head detection. This current authors propose a method in which they remove background, separate persons even with some superposition - although small ones, as shown in their paper. But here we make a comment that detecting persons with even more severe superpositions than they show, can be obtained by FPDW, and this causes problems with fusing faces/heads, as we already explained; extracting heads locations from persons with low superposition is easier. They divide each ones' body into variable parts - arms and legs - and invariable ones - torso and head. Once discarding the variable ones, only torso and heads rest, and they inspect these two regions intersection, inferring that it corresponds to the neck; and thus accurately detect heads locations. Although they comment on using people in frontal positions to the camera, this paper shows interesting figures in which heads are detected from whole bodies ROIs, not always centralized, bust also dislocated; that is, similarly to what we achieved by combining the upper one sixth whole bodies with foreground segmentation. Their objective is to count people from their heads.

Another work mention a strategy to locate heads [218]. This work first considers a simple head location based on whole bodies ROI, supposing also a high central

[9] This is noticed more clearly in the thesis [216], which shows more experiment results.

position. Then the head location is refined by considering the blob minimum and maximum horizontal positions; a median computation is made to minimize the influence of variations due to the inclusion of some neck or torso pixels. An upper part sub-region of 15 percent of the whole bodies ROIs height was considered - what corresponds approximately to the one sixth we considered in our approach. This work objective was to detect people's orientation from their heads.

There is still a work [219] to detect heads positions, but by using a quite different approach from ours and the other ones cited before. From (i) foregrounds; it (ii) computes first its centroid; then, (iii) extracts contour; (iv) measures all contour distances to the centroid; (v) detects maximum distances; and (vi) selects the one with highest height coordinates, taking measures to avoid confusion with persons' hands when arms are directed upwards. It is really quite different from those others; and from the figures shown, it does not seem to detect heads with accuracies good enough for our purpose of extracting parameters for re-identification.

Besides trying to re-identify people by their faces/heads regions, we also attempted to do it by means of parameters extracted from their whole bodies; for this purpose we have used SURF parameters, since they are highly distinctive, even for similar targets - as was verified later during the preliminary tests. This approach showed more robust than the former one based on faces/heads color parameters, since faces/heads regions are blocked by objects in some frames, turning the re-identification then a difficult task - and it happens for the left camera view, when an iron bar, part of the Argonauta's Room structural construction, occludes people's faces/heads regions.

After all, with the combination of all the strategies proposed in this thesis, it was possible to re-identify targets - or to correct wrongly identified ones -, in many situations. The algorithm 9 was capable to recover targets IDs, even not using the left camera, for those severe occlusions cases. When using that left view, results will tend to improve even more.

One consequence of algorithm 9 was that: it was noticed, in some cases, that two targets had IDs wrongly switched, by one of the parameters based approach; when, suddenly, the parameter-based method corrected by itself the IDs, algorithm 9 did not enable the switch; and here, it was a correct switching. But, even though, these problems occur just when targets are very near to each others; and thus, the consequences for dose computation are not so severe. And, considering that the use of algorithm 9 solved more problems than those it created, its use was an advantage overall.

To finally close this chapter, it seems that data association is still a challenging task in computer vision tracking applications. Targets death, drift or switching are issues that may occur with some more or less frequency, depending on the database and on the application in hand. The data association may be based on different approaches, such as trajectories of moving targets; or targets' colors, appearances or other specific characteristics which may be used to distinguish among them. There are approaches based on global optimization to associate interrupted targets trajectories, but this considers that targets are moving along the video sequence, such that this association may be done. When dealing with targets that: (i) experience severe occlusions; or (ii) that have similar colors; or (iii) that remain very close to each others or even standing still for long term, the challenges are higher. The last challenge mentioned, means there may be no trajectories at all to associate a new target birth with a former near target death.

Thus, there is still an open field for more research on re-identification, to strengthen this aspect of video-based surveillance systems, which might improve the specific objective of this thesis: accounting for received radiation dose by each person.

To the extent of the knowledge of the author, the proposed system to re-identify human targets by faces/heads parameters has some particularities from other similar works aiming at detecting and/or tracking faces and/or heads, which makes the proposed system different from them.

Also, to the extent of the knowledge of the author, the strategies of comparing parameters extracted from targets as in the different versions of the algorithm "Multi-layer strategy algorithm for matching parameters" 6, by using levels for comparing them with other past frames, they way developed here, is a new one. This is an important contribution for targets tracking in general, to cope with the problems of missing detections or tracks in some frames.

And, ratifying what was stated in Section 1.5, the approach of using computer vision-based methods for targets detection, tracking and re-identification for radiological protection is, to the extent of the knowledge of the author, an original contribution. Even more because of the form different methods and proposed strategies were used to result in a system capable of performing such task, to compliment traditional radioprotection methods, is unique.

Chapter 9

Results

This chapter shows, in Section 9.1, the simulations results for the proposed methodology. Section 9.2 gives remarks about these results.

9.1 Results and Discussion

Figures 9.1 to 9.3 show some interesting situations. The frames are shown along with the estimated targets WCS coordinates. Below this latter, information shows if targets were re-identified by distance in the WCS or by the parameter-based method. Also, the winner identifiers for each level - level one to the fourth, from left to right -, and the final voted ID, are shown. The latter depends on the weights given by each level. A more detailed explanation of this voting system is given in Figure 9.1. These results were obtained before the application of the algorithm "Further analysis to re-identify targets" 9.

Figure 9.1 shows a situation where re-identification was recovered after a group situation due to a false detection. Here it is possible no notice that correction by algorithm 9 is possible by re-assigning the right ID of the target that fell on the group and was assigned a temporary false detection; the WCS displacement from frame to frame enables easily to associate correctly this target with its previous ID.

Explaining better the voting system mechanism:

In the upper frame shown in Figure 9.1, for example - frame number 960 -, the foremost target was given ID "1", as indicated in the first numerical line below the WCS map, by the resulting number to the right of the vertical bar. In the left part of this bar, four numbers are shown, the sequence: "1 1 2 0". As already explained in the first paragraph of this Section, each one of these numbers shows the winner ID at each level, from the first with only one frame - the leftmost number -, to the last level, with the remaining of the 650 frames - the rightmost number. This means that, in this example, the first and second levels voted for ID "1", the third voted for ID "2", and the last, for ID "0" - this latter value means that this frame is in

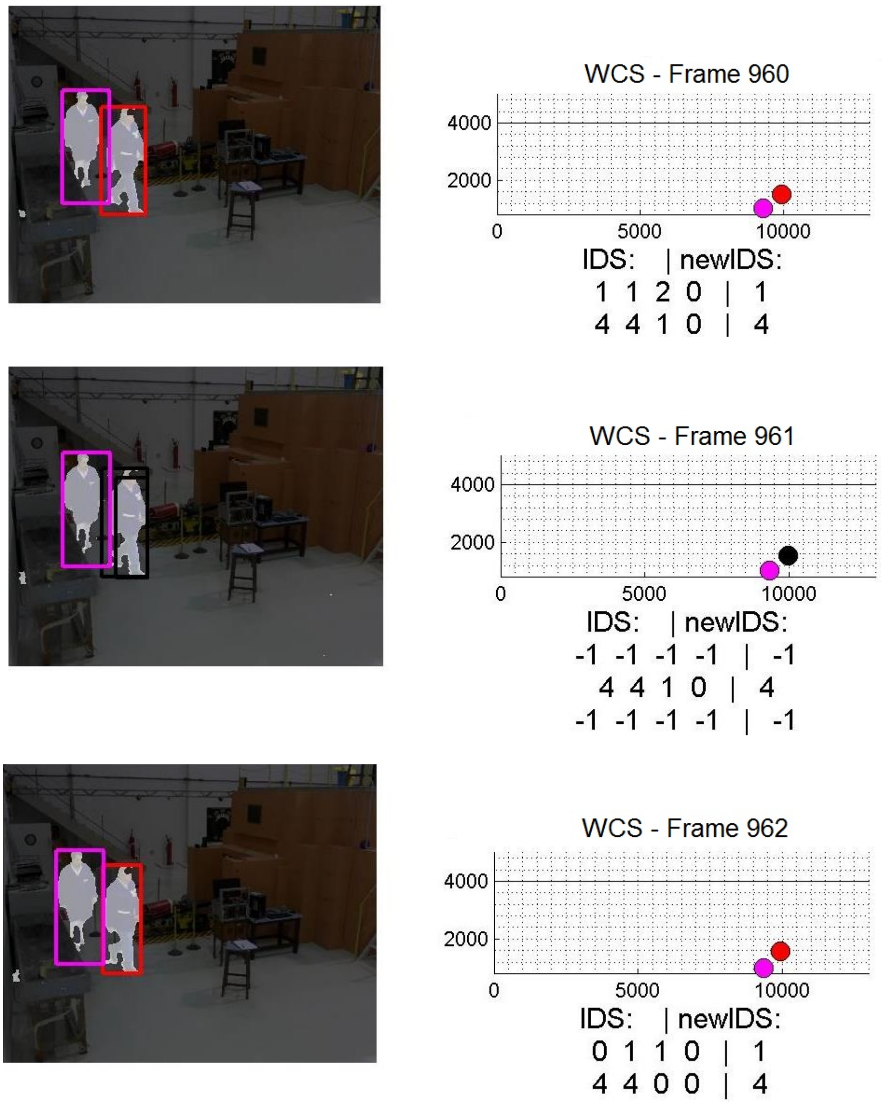

Figure 9.1: Tracking results by SURF; re-identification recovering after a group situation.

the beginning of the video sequence, and thus there are less or equal to 150 frames processed, and no frame at the last level.

If all levels are given equal weights, all of them vote equally to give the final winner ID for this target. But, during experiments, it was verified that sometimes the voting scheme resulted in a wrong ID, even if it did not make sense for a human observer. For example, sometimes the far past target's history - represented by levels three and fourth levels, for example - gave an ID, supposedly the right one. But, depending on the winner IDs at the two first levels, the final ID was wrongly changed. Therefore, the weights of each level were adjusted by try, to result in a good sense final result. But this adjustment is somewhat sensitive. Further tests can improve it in the future.

And last, but not least, for the remaining cases where an ID is wrongly changed, despite the weights adjustments, the final system stage - performed by algorithm 9

- can retrieve the correct results by associating the WCS targets' positions.

Similar explanation is given to the second target in the upper frame shown in Figure 9.1. The levels voted, from the first to the last, the IDs: "4 4 1 0". This target ID was clearly changed, despite its farther history given by the third level, due to the influence of severe occlusions that occurred before, and resulted in changing this target ID. When this change passed to the second level, the two first levels voted for this new ID and won over the third one. Here, again, it should be mentioned the sensitive character of the weights adjustment of this voting system, when there are severe occlusions possible with lost targets, or false alarms. And once more, algorithm 9 can retrieve the correct results.

Taking now the mid frame shown in Figure 9.1 as another example - frame number 961 -, there was a false alarm almost coinciding with the foremost target position in the image plane. This put both detections - the true and the false one - in prohibitive condition, as long as the severe occlusion avoided extracting and identifying the parameters with confidence. In this case, temporary false IDs were given to both - indicated by "-1". The final WCS was obtained by averaging both detections projections to the floor. The other target was not affected by any occlusions, then in no prohibitive situations, and thus was normally identified.

In the bottom frame shown in Figure 9.1 - frame number 962 -, as soon as the former false detection disappeared, the foremost target got out the prohibitive condition, and its parameters could be extracted and compared to the ones previously extracted at frame number 960, and matched correctly. This example also shows a situation where parameters matching is made between non-consecutive frames.

Figure 9.2 shows a situation where re-identification was recovered after some problems: (i) first, a false detection was assigned the identification of one true existing person, who was given a new ID; (ii) then the false ID persisted a bit; and (iii) finally, it was recovered. From the figures, it is possible to notice that these failures can be corrected by associating the nearest respective WCS coordinates in a post-processing stage. Here it is also possible no notice that correction by the algorithm 9 is possible by also assessing the WCS coordinates from frame to frame.

Figure 9.3 shows a situation where a target was lost for some time[1]. Here, the system can not form a group and tries to identify the remaining person; and does this by matching the parameters of the other person. Even though, when they got apart again, the correct IDs were recovered. Here, algorithm 9 can do the following: in the first frame (frame 1111), when a target is lost behind the other, the system - without algorithm 9 - would try to re-identify it based on parameters - what is noticed in this frame figure. Here, algorithm 9 forces a temporary false ID and assigns the only one WCS coordinates to both persons. This avoids the

[1]This persists for a number of frames, which were omitted here since they are very similar.

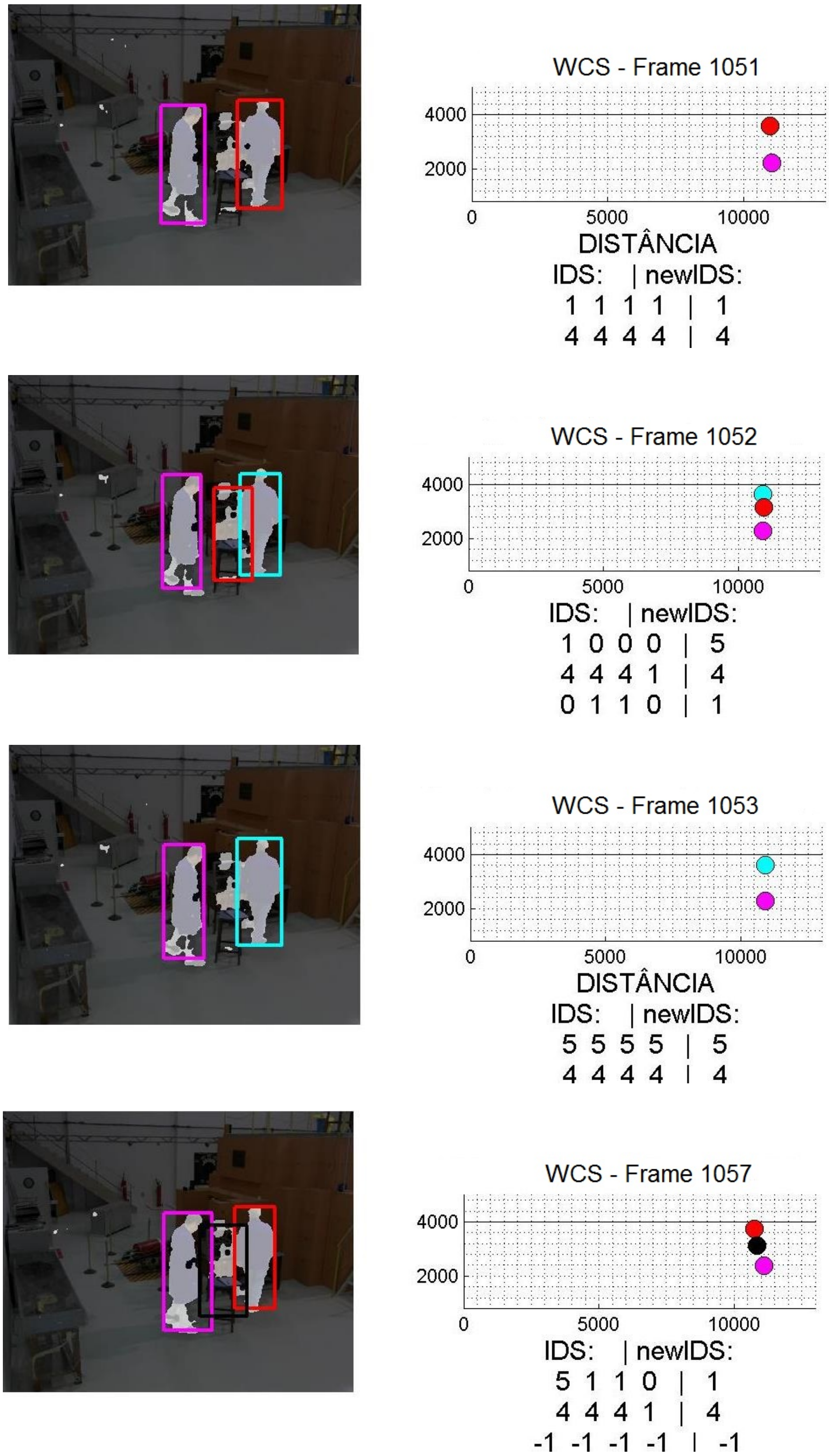

Figure 9.2: Tracking results by SURF; re-identification after a group situation.

possible learning of wrong parameters, corresponding to a wrong target. When the previously lost target re-appears, it is re-identified by parameters again.

An example of a resulting received dose estimation for workers are shown in the following. This was performed for the video with two persons representing a

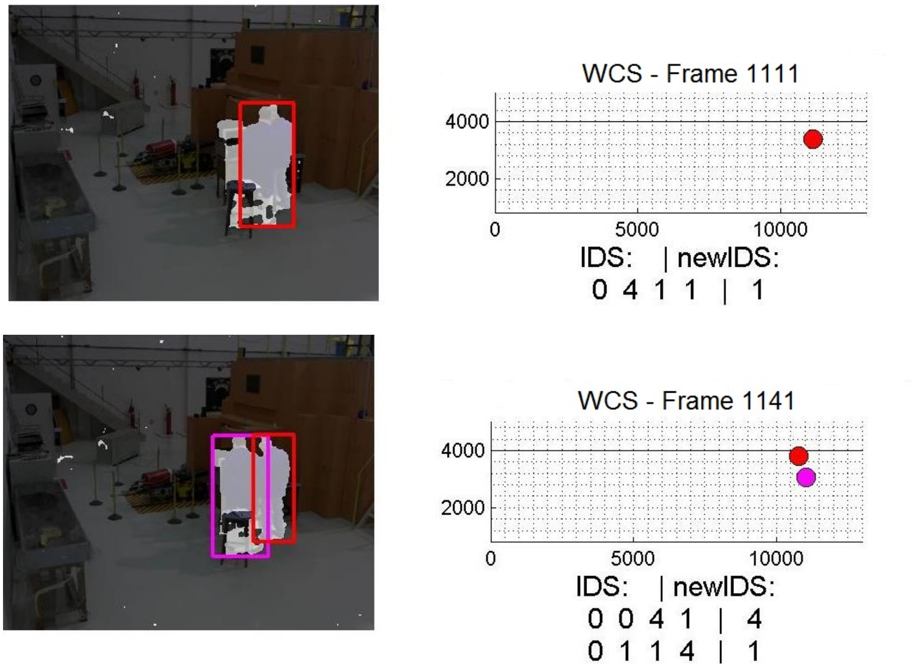

Figure 9.3: Tracking results by SURF; re-identification recovering after target loss situation.

spectrometry experiment. This is one of the most complicated ones in the generated database, due to:

- Severe occlusions;
- Targets losses;
- Targets remaining standing still - or almost standing still - near each others;
- Targets crossing among themselves during this last situations (after being stood still near each other);
- Targets using clothes of similar colors.

These results were the final ones obtained after application of algorithm 9.

In the following, some examples are shown for the most relevant results aimed by this research: the final received radiation dose estimates for personnel, based on computer vision methods for video-based surveillance. Some Error cones are shown in the figures, due to the intrinsic measurement errors resulting from the experimental setup, where a Geiger-Müller sensor was used: plus or minus 20 percent.

Figure 9.4 shows the results for a video with one person performing the tasks related to the spectrometry experiment. It shows a comparative analysis between received doses computed from the WCS ground truth and the tracked WCS positions by the faces/heads color-based approach.

Some points corresponding to particular situations are indicated:

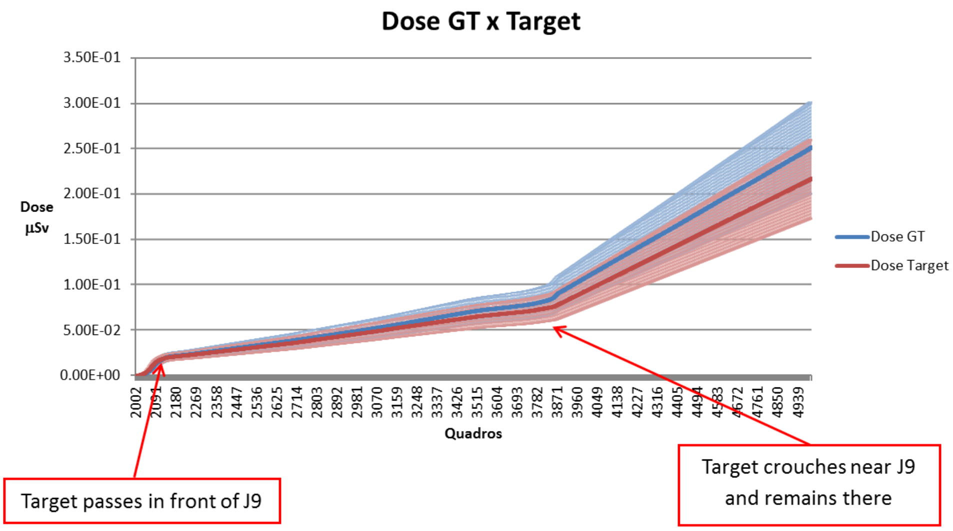

Figure 9.4: Received accumulated radiation dose by Target. One person, faces/heads color-based approach.

1. When, after entering the Argonauta's Room, the target passes in front of J9, the received dose increased in a higher rate, as expected;

2. When the target goes to the region near J9, in the spectrometer, and crouches remaining there, the received dose also increases in a higher rate.

Figure 9.5 shows the results for another video with one person performing tasks related to the spectrometry experiment, by the SURF approach[2]. In this case, the target crouched near the J9 region, but behind the spectrometer, from the point of view of Camera 1.

In this case, a deviation resulted between the estimated and the ground truth-based computed received doses. Also, it is possible to notice that both curves agree in a great part of the graphics, deviation only from the point indicated in the figure on. Figure 9.6, though, shows the cause of this deviation: the target remained crouched with his feet and part of his legs occluded by the spectrograph. The ROI detected by FPDW is shown, where it is possible to notice this failure in estimating the feet position in the image plane. When projecting it to the WCS, a deviation

[2]It has to be emphasized that, in this latter result, it was observed that, in parts of the video, some false detections arise with bad aspect ratios. Therefore, an additional criterion was used then, to eliminate those false detections by filtering the ones out of realistic intervals - similarly as in CPMD-RPT. Thus, as the experiments were carried out, some new situations suggested that the last module of the developed system, the "Further analysis to re-identify targets" (see Chapter 8, Subsection 8.2.5), can be incremented with new criteria, in the future. Another criterion that was envisioned latter, was that false detections can be filtered also by their heights in the WCS, by using the projection matrix to estimate the heads positions - also, similarly as in CPMD-RPT -, combined with the homography for estimating feet positions. But this latter one will be part of future developments.

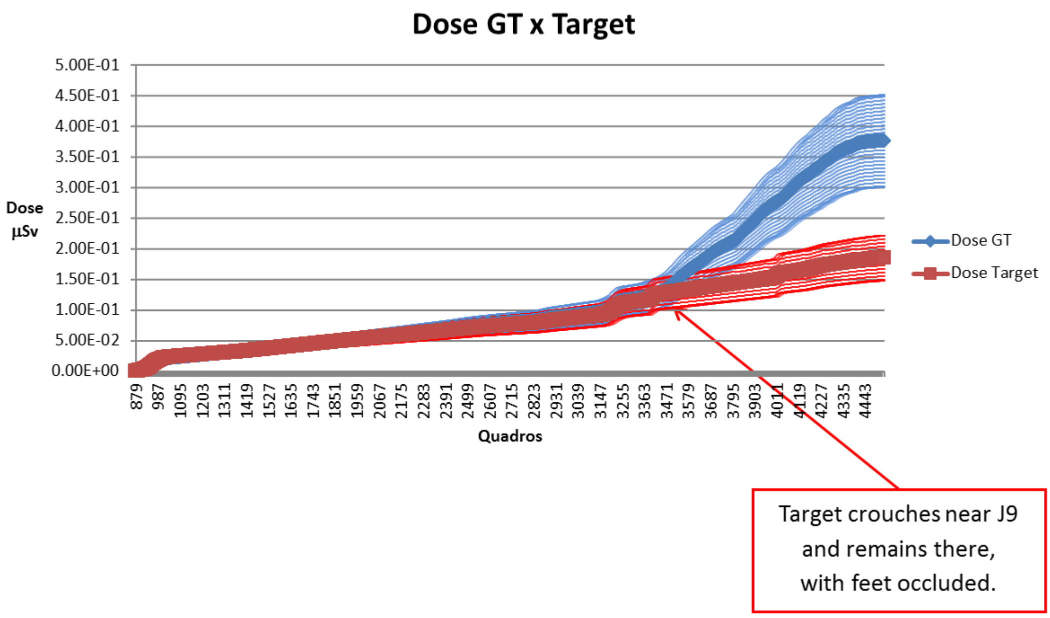

Figure 9.5: Received accumulated radiation dose by Target, during occlusion. One person, SURF-based approach.

of around one meter arose, to the left. Considering that this occurred near J9, where radiation dose rate varies with higher rate (see Figure 6.3), this resulted in a significant difference in the received dose estimation, relatively to the ground truth-based one.

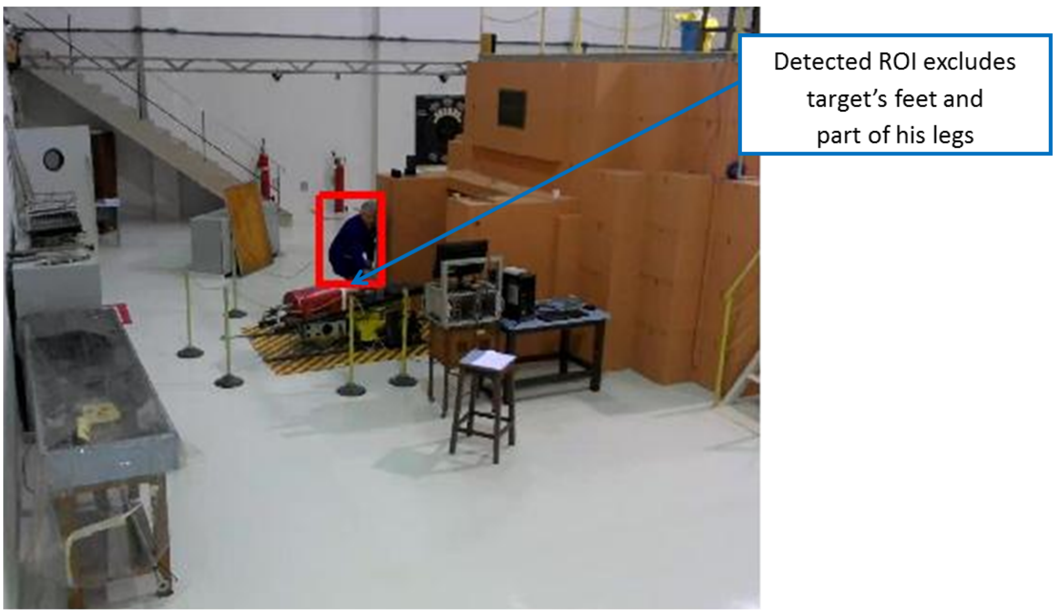

Figure 9.6: The bad feet position estimate in the image plane due to occlusion.

But, in counterpart, Figure 9.7 shows the corresponding scene recorded from the point of view Camera 2, where it is possible to notice that the target is free from any occlusions, and his feet can be well located by the developed system. Therefore, the occlusion problems can be solved by switching between the two cameras.

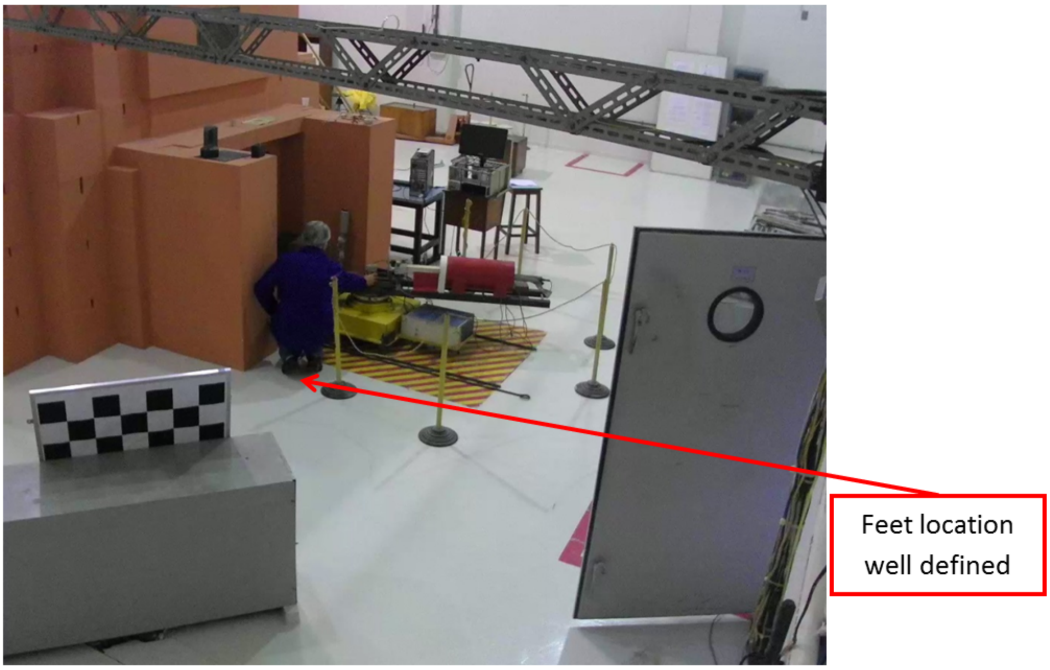

Figure 9.7: The same corresponding scene free from occlusion, taken by the other camera.

Figures 9.8 and 9.9 show comparatively the received accumulate doses for a video with two persons performing tasks related to the spectrometry experiment, by the SURF-based approach. Target 1 is the foremost person (see Figure 8.2), while Target 2 is the one just behind.

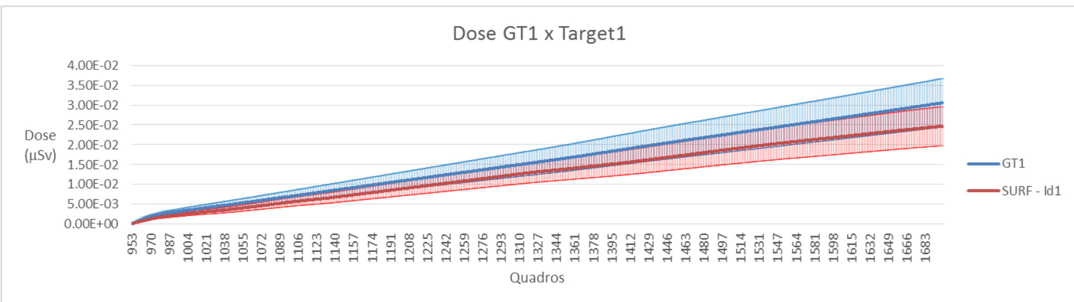

Figure 9.8: Received accumulated radiation dose by Target 1. Two persons, SURF-based approach.

As these computations was performed beginning from frame 952, it is possible to notice a higher increasing rate of received dose by Target 2, when it passes just in front of J9, as indicated in the figure.

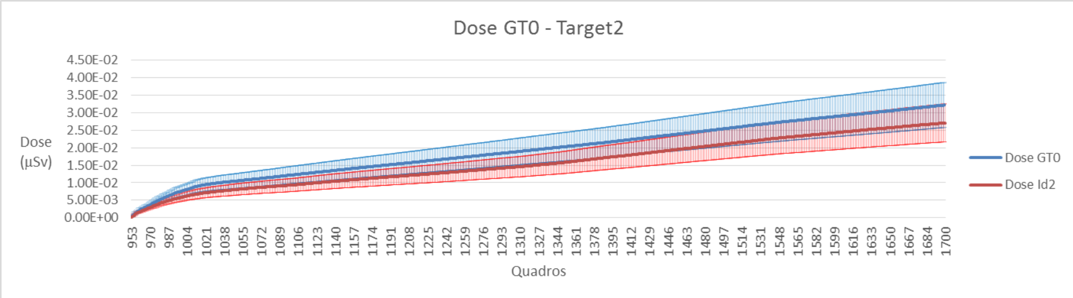

Figure 9.9: Received accumulated radiation dose by Target 2. Two persons, SURF-based approach.

It has to be mentioned that the two simulations with one person were performed by using background removal with $W4$, while the one with two persons was performed by using GMM. Both methods can be used alternatively, as they present similar results. But it was noticed a little difference in the results, for the two-person example, relatively to the capability of eliminating false alarms, when they arise between two true targets. Sometimes, the foreground supplied by $W4$ was a bit worse in such a way this little differences avoided the elimination of some false alarms, leading to switching the true identities in the beginning of that video. Of course, with the last processing stage, the “Further analysis to re-identify targets”, all these changes can be corrected, and at thus, become irrelevant. But, during the system development testing stage, a try was made by using alternatively GMM, to verify the difference in the results. In the future, on background removal should be chosen to perform all simulations, for uniformity in the processing tasks.

9.2 Remarks

By closing this Results Chapter, some considerations should be made. First and main comment is that the objective of this thesis was achieved: it is possible to estimate received radiation dose for nuclear plants personnel, based on computer vision methods for video-based surveillance. Figures 9.4, 9.5, 9.8 and 9.9 shown in Section 9.1 demonstrate this, by the agreement between the ground truth-based and estimated dose curves. The deviation shown in Figure 9.5 was explained and justified, and it was shown that the other camera view supply corresponding images without occlusion. This last cited example showed a case of occlusion caused by an obstacle of the environment, but it could be due to occlusion by another person.

In the following, some problems more related to practical issues are commented:

- There were problems relatively to the few number and the locations where the cameras were authorized to be installed, combined to typical movements of people in the environment: there were too much occlusions due to people alignment to both cameras, besides obstacles in the left camera view.

- As the system showed good performance for some parts of the scenes where people were not aligned relatively to the cameras, it seems that, with more appropriate installation of cameras, both in number and in their locations, many problems that occurred can be solved, as already explained in the last paragraph.

Relatively to the two items commented just before, some further considerations can be made about possible solutions, in the future:

1. Installing more appropriate cameras for internal surveillance, with better aperture angles, and for which we could have more control on their recording parameters, as using cameras own manager to deal with multiple cameras; and also trying to use open video formats.

2. The left camera must be better located, to avoid obstacles of the environment in its view: the criterion for installation must be strictly technical, from the computer vision perspective, not from any others, - as was the case, since cameras had to be installed near the existing conduits.

3. At least one more camera must be installed in front of J9 - or approximately in front of it - to enable a transversal view that could aid in solving some problems due to severe occlusions that occurred with people aligned simultaneously to both the right and the left cameras. This (or these) new camera could also solve better the locations of people when in front of J9, the hottest region during Argonauta operation; this (or these) camera should be installed in such a way to supply a good coverage of this region, by taking into account its aperture angles and local for installation.

4. For an even better solution, at least one camera could be installed in the roof - or in another high position, relatively to the Argonauta's Room. This (or these) cameras could track with greater accuracy the trajectories of each worker, what is very important for dose accounting. Of course this (or these) upper cameras could not solve for re-identification, since it (or they) would perform a rather point-like tracking; but its (or their) best trajectories estimation in the WCS could be combined to the re-identification supplied by the lower cameras that can capture better the faces/heads, and whole bodies, to obtain the interest-parameters for re-identification.

All the above considerations are related to surveillance of the regions of interest treated up to now: Areas “1” to “3”. Other regions within Argonauta’s Room could also be covered.

Chapter 10

Conclusions

10.1 Conclusions

As commented in Section 9.2, the objective of this thesis was achieved: it is possible to estimate received radiation dose for nuclear plants personnel, based on computer vision methods for video-based surveillance. It must be emphasized that this result is original: there were no similar work found in the literature. This opens a new field of research, since it was demonstrated that computer vision-based methods can be used for safety purpose during routine tasks in nuclear plants, or other plants (as radioactive ones) where people are subject to radiation dose exposition. This can fill the current gap of existing monitoring approaches, as already explained in Sections 1.5 and 3.3, and can be used in conjunction with these existing approaches to achieve redundancy, or to supply estimations when the existing methods may fail.

Also, as commented in Section 9.2, there is still an open field for research on targets re-identification in video-based surveillance systems. It is expected that the proposed system operates well most of the time, or during those common situations.

Relatively to received radiation dose accounting, we should remember also that the existing methods for measuring doses, both ambient doses and the ones received by each worker of the nuclear field, have still their own limitations and intrinsic failures[1].

In critical plants such as nuclear or radioactive ones - just to cite examples related to this thesis theme[2] -, redundancy is an important matter. Also, not all critical work will be subject to decisions taken by computers, without any human intervention. Two extremes might be envisioned, but are not good choices:

- That in which all work is performed manually by operators - what is likely to cause incidents, due to their lack of attention even for a short period of time,

[1]Refer to Chapter 3, especially to Section 3.3.
[2]But it should be applied to other types of critical plants or tasks.

and due to the burden caused by hard and tedious work;

- Another in which all tasks are given to computers - what is likely to cause accidents.

Computer-based systems should be rather used as support systems to aid operators in burden and tedious tasks, but still letting them some freedom to act. Information supplied by computerized systems must rather give support to final human decisions. Relatively to this work, if any failure occurs in dose accounting, an alarm can be generated when computed doses seem to be wrong in some sense - as, for example, showing bumps or abrupt reductions in some moments, for one or other worker, what could indicate that persons might have been switched. In those situations, verifications should be performed to assess what really occurred, by comparing results obtained by different approaches. Of course, the system is expected to operate well most of the time, and nobody could take into account workers' trajectories and dose computations manually, if not by using any automatic means... Human intervention should occur in some few situations only.

Anyway, for such a system to be adopted by the nuclear field as equivalent to the existing methods relatively to safety and confidence, it should be evaluated for a somewhat long time. But, for use in parallel to the existing methods, it could be readily implemented, if supplying relatively good results.

To close this Subsection, it must be emphasized that the results achieved demonstrated the feasibility of estimating received radiation dose by each person within a plant subject to a radiation field, by using only the developed computer vision-based system. The graphics show this. Thus, the objective of this thesis was achieved successfully.

10.2 Future Work

First of all, the use of the left camera could improve even more the solution of the problems due to severe occlusion relatively to the right camera view. It could be done soon.

Other improvements are related to the last module of the developed system, the "Further analysis to re-identify targets", can be incremented with new criteria, in the future (as already explained in Section 9.1). Also, the most those corrections can be made online, they will be.

Further work can be carried out in terms of including other computer vision-based methods to this system. Some of them were indicated during the Qualifying Exam for evaluation, including some with good capabilities to deal at least with partial occlusion, by adopting patch-based tracking, as FragTrack and CPMD-RPT.

The latter one is capable of dealing even with short term total occlusions, and has also a pipeline for detection, validation, and tracking management for re-initializing lost targets. At that time - two years ago -, no-one would know if they could solve completely the specific task of this thesis, or not, nor how much they would solve it; or if they would need improvements.

The specific situations in this video database related to persons' colors and typical movement (or lack of movement), and high proximity among them, plus severe occlusions, posed issues for detection, tracking and re-identification - specially the latter task. Besides this, things were worsened by the strict conditions in which we performed the experimental tasks: (i) with only two borrowed cameras, installed under very strict *sine qua non* condition for use; (ii) restrictions to the form with which we could mark the floor; and (iii) also the pressure of time to give back the borrowed cameras so as not to weaken the security system then in operation.

As commented in Section 9.2, more and better installed cameras could solve part of the problems. Also, including new computer vision-based methods or new developments over the existing ones, could aid in solving them better. The former would help the tasks from an engineering point of view; while the latter could supply more computer vision capabilities to cope with the specific situations of the task in hand - specially re-identification.

Online implementation can be achieved by possibly reducing somewhat the frame rate with a compromise of not lowering the methods tracking and data-association efficiencies. Other improvement from the engineering point of view, would be porting the methods codes to C, for example, as a number of them were recently developed, or are still under development, and as such are implemented in MatLab. And finally, be it alternatively - or even in a complementary form to porting it or not to faster execution languages -, parallel computing can be used to implement this system online, specially by using Graphical Processing Units (GPU), an approach that have demonstrated, in the last years, its great potential to perform costly computing tasks, even better than using former approaches as using computer clusters. And there are diverse possibilities of using GPUs, even with Matlab.

Relatively to computer vision-based improvements, some other methods also referred to in the literature as good ones, could be also tested; some of them were mentioned along Chapter 4. One very promising method is DP+NMS, which tries to solve re-identification problems on its own form, by using Markov-chain Monte Carlo (MCMC)-based approach for data association. The DP+NMS method is non-causal, meaning it could not be used online - it makes use of future information to achieve its final results. This means it must run first, and after all, compose the final result offline.

Anyway, as the proposal, up to now, has been combining results from different methods and cameras, to compensate for failures of one another, probably the system would not need to operate online in the sense of giving the final received doses - even because the results may need further verifications, as the task is quite serious.

And even if the system operates offline, the videos will be recorded online along with the real dose rate distributions - which may be obtained by more sophisticated means, either by computing approach or by considering fluctuations collected by monitors (as already explained in Chapter 6, more specifically concluded in Section 6.3). Thus, all the doses computed by visual-based tracking would take into account for doses variations, including peaks that could affect more workers' health.

What can be performed in real-time, or online, is that possible surge dose rate peaks can generate alarms for the CSPR and/or Argonauta heads, who should alert personnel; but the dose computations may be done by processing at the end of the day, or during the following night. Therefore, everyone would know their received doses soon, and decisions could be made whether anyone could continue doing his or her work or might reduce work for some period of time, or at least compare to results obtained by other means to take a decision. We must remember that even this solution of processing soon after the operations, is better than the current ones, in which people only know their received doses one month later, or by a low accuracy "pen" dosimeter.

In terms of computer vision-based methods improvements, I suggest to consider DP+NMS readily for tests, in the near future. The better the methods available, the better the results. And we would have one more method to cope with re-identification, besides the ones proposed in this thesis. This could compose a better voting system for re-identification. Another suggestion would be to include CPMD-RPT combined with algorithm 9.

The reference [75] addresses some interesting approaches, including some for dealing with re-identification, that could also be more explored in the future.

Bibliography

[1] VALENTIN, J., (ED.). "Publication 103 - 2007 Recommendations of the International Commission on Radiological Protection", *Annals of the ICRP*, v. 37, n. 2-4, pp. 1 – 332, April 2007.

[2] "Nuclear Engineering Institute (English version)". `http://www.ien.gov.br/ingles/index.php`.

[3] "National Commission on Nuclear Energy (In Portuguese)". `http://www.cnen.gov.br`.

[4] AUGUSTO, S. C., MÓL, A. C. A., JORGE, C. A. F., et al. "Use of Virtual Reality to Estimate Radiation Dose Rates in Nuclear Plants". In: *2007 International Nuclear Atlantic Conference (INAC 2007)*, v. DVD-ROM, 2007.

[5] AUGUSTO, S. C. *Use of Virtual Environments in Estimative of Radiation Dose in Nuclear Facilities (in Portuguese)*. Master Thesis, Civil Engineering - Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil, 2008. Available at `http://www.coc.ufrj.br/index.php/component/docman/doc_download/1446-silas-cordeiro-augusto-mestrado?Itemid=`.

[6] MÓL, A. C. A., JORGE, C. A. F., COUTO, P. M., et al. "Virtual Environments Simulation for Dose Assessment in Nuclear Plants", *Progress in Nuclear Energy*, v. 51, n. 2, pp. 382 – 387, March 2009.

[7] MÓL, A. C. A., AGHINA, M. A. C., JORGE, C. A. F., et al. "Virtual Simulations of Nuclear Plants for Dose Assessment with On-line Measurements Collected by Networked Radiation Monitors". In: *The 11th IEEE International Conference on Computational Science and Engineering (CSE) - Workshops*, v. CSEW, pp. 367–372, 2008.

[8] MÓL, A. C. A., AGHINA, M. A. C., JORGE, C. A. F., et al. "Nuclear Plant's Virtual Simulation for On-line Radioactive Environment Monitoring and Dose Assessment for Personnel", *Annals of Nuclear Energy*, v. 36, n. 11-12, pp. 1747 – 1752, Nov.-Dec. 2009.

[9] FREITAS, V. G. G. *Real Time Determination of Dose Radiation Through Artificial Intelligence and Virtual Reality (in Portuguese)*. Master Thesis, National Commission of Nuclear Energy, Nuclear Engineering Institute, Rio de Janeiro, Brazil, 2009. Available at http://www.ien.gov.br/posien/teses/dissertacao_mestrado_ien_2009_06.pdf.

[10] FREITAS, V. G. G., MÓL, A. C. A., PEREIRA, C. M. N. A. "Real Time Determination of Dose Radiation Through Artificial Intelligence and Virtual Reality". In: *2009 International Nuclear Atlantic Conference (INAC 2009)*, v. DVD-ROM, 2009.

[11] MÓL, A. C. A., PEREIRA, C. M. N. A., FREITAS, V. G. G., et al. "Radiation Dose Rate Map Interpolation in Nuclear Plants Using Neural Networks and Virtual Reality Techniques", *Annals of Nuclear Energy*, v. 38, n. 2-3, pp. 705 – 712, Feb.-March 2011.

[12] JORGE, C. A. F., MÓL, A. C. A., COUTO, P. M. "Nuclear Plants and Emergency Virtual Simulations Based on a Low-cost Engine Reuse". In: Tsvetkov, P. V. (Ed.), *Nuclear Power*, Sciyo/InTech, pp. 367–388, Rijeka, Croatia, 2010. Available at http://cdn.intechopen.com/pdfs-wm/11579.pdf.

[13] AUGUSTO, S. C., MÓL, A. C. A., COUTO, P. M., et al. "Using Virtual Reality in the Training of Security Staff and Evaluation of Physical Protection Barriers in Nuclear Facilities". In: *2009 International Nuclear Atlantic Conference (INAC 2009)*, v. DVD-ROM, 2009.

[14] JORGE, C. A. F., MÓL, A. C. A., COUTO, P. M., et al. "Virtual Simulation for Training Personnel in Emergency and Security Preparedness and Counteraction". In: *XIII Simpósio de Aplicações Operacionais em Áreas de Defesa (SIGE)*, v. DVD-ROM, 2011.

[15] HARTLEY, R., ZISSERMAN, A. *Multiple View Geometry in Computer Vision*. 2nd ed., Cambridge, UK, Cambridge University Press, 2003.

[16] "Argonauta Nuclear Research Reactor (In Portuguese)". http://www.ien.gov.br/oinstituto/instalacoes/serea/argonauta/index.htm.

[17] OLIVEIRA, E. S. L., CARDOZO, K. K. M., SILVA, J. C. P., et al. "Análise do Levantamento Radiométrico Durante Operação do Reator Argonauta". In: *IX Latin American IRPA Regional Congress on Radiation Protection and Safety (IRPA 2013)*, v. DVD-ROM, 2013.

Available at http://www.sbpr.org.br/irpa13/AnaisdoIRPA2013/Culturadelaseguridadypercepcionderiesgo/3796.pdf.

[18] "Eletrobrás Eletronuclear (In Portuguese)". http://www.eletronuclear.gov.br/.

[19] "Centro de Desenvolvimento da Energia Nuclear (In Portuguese)". http://www.cdtn.br.

[20] "Instituto de Pesquisas Energéticas e Nucleares (In Portuguese)". http://www.ipen.br.

[21] "Universidade Federal do Rio de Janeiro (In Portuguese)". http://www.ufrj.br.

[22] "Instituto Militar de Engenharia (In Portuguese)". http://www.ime.eb.br.

[23] FABRY, T. *Interactive Visual Intervention Planning in Particle Accelerator Environments with Ionizing Radiation*. PHD Thesis, University of Grenoble, France, 2014. Available at http://cds.cern.ch/record/1977459/files/CERN-THESIS-2014-202.pdf?version=1.

[24] STABIN, M. G. *Radiation Protection and Dosimetry: An Introduction to Health Physics*. Springer, 2008.

[25] TAHUATA, L., ALMEIDA, I. S., PRINZIO, R., et al. "Radioproteção e Dosimetria: Fundamentos, 10th Rev." Handbook, 2014. Available at http://www.ird.gov.br/index.php?option=com_docman&task=doc_download&gid=32&Itemid=49.

[26] "International Commission on Radiological Protection". [Online], . Available at http://www.icrp.org.

[27] "International Commission on Radiation Units and Measurements". [Online], . Available at http://www.icru.org.

[28] "International Atomic Energy Agency". [Online]. Available at http://www.iaea.org.

[29] "Diretrizes Básicas de Proteção Radiológica". Radiological Protection Guideline, 2014. Available at http://www.cnen.gov.br/seguranca/normas/pdf/nrm301.pdf.

[30] “Posição Regulatória 3.01/004:2011 - Restrição de Dose, Níveis de Referência Opcupacionais e Classificação de Áreas”. Radiological Protection Guideline, 2011. Available at http://www.cnen.gov.br/seguranca/normas/pdf/pr301_04.pdf.

[31] “Occupational Radiation Protection - Safety Guide RS-G-1.1”. In: *IAEA Safety Standards Series*, International Atomic Energy Agancy, 1999. Available at http://www-pub.iaea.org/MTCD/Publications/PDF/Pub1081_web.pdf.

[32] “Instituto de Radioproteção e Dosimetria (In Portuguese)”. http://www.ird.br.

[33] “Polimaster”. http://www.polimaster.com/.

[34] “X-Ray and Gamma Radiation Personal Dosimeters PM1621”. Operation Manual. Available at http://www.polimaster.com/files/downloads/PM1621_manual_eng.pdf.

[35] OLIVEIRA, M. V., AGHINA, M. A. C., SOARES, M. “Sistema de Monitoração de Radiação 7027”. In: *The 4th Brazilian National Meeting on Nuclear Applications (ENAN) and 11th Brazilian National Meeting on Reactor Physics and Thermal Hydraulics (ENFIR)*, 1997.

[36] OLIVEIRA, M. V. “Manual de Serviço do Monitor de Radiação MRA 7027 - MS-MRA-7027”. Service Manual, 2001.

[37] OLIVEIRA, M. V. “Manual de Serviço da Sonda Geiger-Müller SGM 7027”. Service Manual, 2001.

[38] LEWIS, M., JACOBSON, J. “Introduction”, *Communications of the Association for Computing Machinery – Special Issue on 'Game Engines in Scientific Research'*, v. 45, n. 1, pp. 27 – 41, 2002.

[39] ROSEMBLOOM, A. “Introduction”, *Communications of the Association for Computing Machinery – Special Issue on 'A Game Experience in Every Application'*, v. 46, n. 7, pp. 28 – 31, 2003.

[40] ZYDA, M. “Introduction”, *Communications of the Association for Computing Machinery – Special Issue on 'Creating a Science of Games'*, v. 50, n. 7, pp. 28 – 36, 2007.

[41] TRENHOLME, D., SMITH, S. P. “Computer Game Engines for Developing First-person Virtual Environments”, *Virtual Reality*, v. 12, n. 3, pp. 181 – 187, 2008.

[42] RINDAHL, G., MARK, N. K. F. "Norway: VRdose™ and Emerging 3D Software Solutions to Support Decommissioning Activities". In: *IAEA-TECDOC-1602 - Innovative and Adaptive Technologies in Deccommissioning of Nuclear Facilities*, International Atomic Energy Agency, pp. 147–158, 2008. Available at http://www-pub.iaea.org/MTCD/publications/PDF/te_1602_web.pdf.

[43] OH, W. Z., CHOI, W. K., LEE, K. W., et al. "Republic of Korea: Technology Development on the Decontamination and Decommissioning of the Nuclear Research Facilities in Korea". In: *IAEA-TECDOC-1602 - Innovative and Adaptive Technologies in Deccommissioning of Nuclear Facilities*, International Atomic Energy Agency, pp. 159–172, 2008. Available at http://www-pub.iaea.org/MTCD/publications/PDF/te_1602_web.pdf.

[44] "Institute for Energy Technology (English version)". http://www.ife.no/en.

[45] "Halden Virtual Reality Centre". http://www.ife.no/en/ife/laboratories/hvrc1/halden-virtual-reality-centre.

[46] "HVRC VRdose". http://www.ife.no/en/ife/departments/software-engineering/products/vrdose/hvrc-vrdose.

[47] SEBOK, A., NYSTAD, E., DRØIVOLDSMO, A. "Improving Safety and Human Performance in Maintenance and Outage Planning Through Virtual Reality-based Training Systems". In: *IEEE 7th Conference on Human Factors and Power Plants (HFPP)*, v. 8, pp. 14 – 22, 2002.

[48] NYSTAD, E., DRØIVOLDSMO, A., SEBOK, A. "Use of Radiation Maps in a Virtual Training Environment for NPP Field Operators". Halden Work Report, 2002.

[49] SZÖKE, I., LOUKA, M. N., MARK, N. K., et al. "New Software Tools for Dynamic Radiological Characterisation and Monitoring in Nuclear Site". In: *OECD NEA Workshop on Radiological Characterisation for Decommissioning*, 2012. Available at http://www.oecd-nea.org/rwm/wpdd/rcd-workshop/E-2___PAPER_New_software_tool_for_dynamic_radiological-Szoke.pdf.

[50] SZÖKE, I., JOHNSEN, T. "Human-centred Radiological Software Techniques Supporting Improved Nuclear Safety", *Nuclear Safety and Simulation*, v. 4, n. 3, pp. 219 – 225, 2013.

[51] VABØ, R., PIOTROWSKI, L., RINDAHL, G. "3D Representation of Radioisotopic Dose Rates within Nuclear Plants for Improved Radioprotection and Plant Safety", *Nuclear Safety and Simulation*, v. 1, n. 2, 2010.

[52] LOUKA, M. N. "Augmented and Virtual Reality Research in Halden 1998-2008". In: Skjerve, A. B., Bye, A. (Eds.), *Simulator-based Human Factors Studies Across 25 Years - The History of the Halden Man-Machine Laboratory*, Springer, pp. 287–302, 2011.

[53] "Japan Atomic Energy Agency (English version)". `http://www.jaea.go.jp/english/index.html`.

[54] IGUCHI, Y., KANEHIRA, Y., TACHIBANA, M., et al. "Development of Decommissioning Engineering Support System (DEXUS) of the Fugen Nuclear Power Station", *Journal of Nuclear Science and Technology*, v. 41, n. 3, pp. 367 – 375, 2004.

[55] RINDAHL, G., MEYER, G., LOUKA, M., et al. "Virtual Reality for Radiation Visualisation and Knowledge Management". Presentation, 2005. Available at `http://www2.hrp.no/vr/workshops/2005/rindahl.pdf`.

[56] ISHII, H., FUJINO, H., ZHIQIANG, B., et al. "Development of Marker-based Tracking Methods for Augmented Reality Applied to NPP Maintenance Work Support and its Experimental Evaluation". In: *5th International Topical Meeting on Nuclear Plant Instrumentation Controls, and Human Machine Interface Technology (NPIC and HMIT 2006)*, 2006.

[57] "Belgian Nuclear Research Centre (English version)". `https://www.sckcen.be/en`.

[58] VERMEERSCH, F., BOSSTRAETEN, C. V. "Development of the VISIPLAN ALARA Planning Tool". In: *International Conference on Topical Issues in Nuclear Radiation and Radioactive Waste Safety*, 1998.

[59] VERMEERSCH, F., BOSSTRAETEN, C. V. "Software VISIPLAN: a Powerful Tool for Optimisation". In: *2nd European Workshop on Occupational Exposure Management at NPPs*, 2000. Available at `http://jongeren.sckcen.be/~/media/Files/Visiplan/Literature/tarragona.pdf`.

[60] VERMEERSCH, F., NIJS, R. "The Use of Computer Simulations in Specific Job Training, Risk Communication and Safety". In: *3rd International Conference on Education and Training in Radiological Protection (ETRAP 2005)*, 2005.

[61] VERMEERSCH, F. “ALARA Pre-job Studies Using the VISIPLAN 3D ALARA Planning Tool”, *Radiation Protection Dosimetry*, v. 115, n. 1-4, pp. 294 – 297, 2005.

[62] HARDEMAN, F., DEBOODT, P., ANTOINE, P., et al. “20 Years of ALARA Management, Research and Development at the Belgian Nuclear Research Centre SCK•CEN”, *Annalen van de Belgische Vereniging voor Stralingsbescherming*, v. 37, n. 3, pp. 79 – 96, 2012.

[63] “Universitat Politècnica de València (English version)”. `https://www.upv.es/index-en.html`.

[64] “IBERDROLA Engineering and Construction (English version)”. `http://www.iberdrolaingenieria.com/ibding/home.do?__setlocale=en`.

[65] RÓDENAS, J., ZARZA, I., BURGOS, M. C., et al. “Developing a Virtual Reality Application for Training Nuclear Power Plant Operators: Setting Up a Database Containing Dose Rates in the Refuelling Plant”, *Radiation Protection Dosimetry*, v. 111, n. 2, pp. 173 – 180, 2004.

[66] RÓDENAS, J., ZARZA, I., FELIPE, A., et al. “Generation of Dose Databases for an Operator Training Application in Nuclear Facilities”. In: *3rd International Conference on Education and Training in Radiological Protection (ETRAP 2005)*, 2005.

[67] OHGA, Y., FUKUDA, M., SHIBATA, K., et al. “A System for Calculation and Visualization of Radiation Fields for Maintenance Support in Nuclear Power Plants”. In: *International Conference on Radiation Shielding and Topical Meeting on Radiation Protection and Shielding*, 2004.

[68] OHGA, Y., FUKUDA, M., SHIBATA, K., et al. “A System for the Calculation and Visualisation of Radiation Fields for Maintenance Support in Nuclear Power Plants”, *Radiation Protection Dosimetry*, v. 116, n. 1-4, pp. 592 – 596, 2005.

[69] PEREIRA, F., THEIS, C., MOREIRA, A., et al. “Evaluating Location Fingerprinting Methods for Underground GSM Networks deployed over Leaky Feeder”. In: *2011 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2011.

[70] PEREIRA, F., THEIS, C., MOREIRA, A., et al. “Multitechnology RF Fingerprinting with Leaky-feeder in Underground Tunnels”. In: *2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2012.

[71] PEREIRA, F., THEIS, C., MOREIRA, A., et al. "Performance and Limits of KNN-based Positioning Methods for GSM Networks Over Leaky Feeder in Underground Tunnels", *Journal of Location Based Services*, v. 6, n. 2, pp. 117 – 133, 2012.

[72] WOOD, R. T., EWING, P. D., KERCEL, S. W., et al. "Electromagnetic Compatibility in Nuclear Power Plants". Available at `http://web.ornl.gov/~webworks/cpr/pres/103297.pdf`.

[73] KERCEL, S. W., MOORE, M. R., BLAKEMAN, E. D., et al. "NUREG/CR-6436 ORNL/TM-13171 - Survey of Ambient Electromagnetic and Radio-Frequency Interference Levels in Nuclear Power Plants". Report, 1996. Available at `http://www.iaea.org/inis/collection/NCLCollectionStore/_Public/28/042/28042825.pdf`.

[74] SARTORI, C. A. F. "TNR5767 - Aspectos de Compatibilidade Eletromagnética em Sistemas Elétricos de Plantas Nucleares". University Lecture, 2005. `http://sistemas.ipen.br/Ensino/Consulta_Disciplinas_Centro.php?a=2008&ce=CCCH&ti=2`.

[75] MAGGIO, E., CAVALLARO, A. *Video Tracking Theory and Practice.* West Sussex, UK, John Wiley and Sons, Ltd., 2011.

[76] REGAZZONI, C. S., CAVALLARO, A., (EDS.), et al. *EURASIP Journal on Image and Video Processing, Hindawi Publishing Corporation - Issue 'Video Tracking in Complex Scenes for Surveillance Applications'*, v. 2008, n. 2-4, 2008.

[77] CHELLAPPA, R., SANKARANARAYANAN, A. C., (EDS.), et al. *Foundations and Trends in Signal Processing - Issue 'Statistical Methods and Models for Video-based Tracking, Modeling, and Recognition'*, v. 3, n. 1-2, pp. 1 – 151, 2009.

[78] REMAGNINO, P., JONES, G. A., (EDS.), et al. *Video-based Surveillance Systems: Computer Vision and Distributed Processing*, 2002.

[79] YILMAZ, A., JAVED, O., SHA, M. "Object Tracking: A Survey", *ACM Computer Surveys*, v. 38, n. 4, 2006.

[80] MOESLUND, T. B., GRANUM, E. "A Survey of Computer Vision-based Human Motion Capture", *Computer Vision and Image Understanding*, v. 81, pp. 231 – 268, 2001.

[81] MOESLUND, T. B., HILTON, A., KRÜGER, V. "A Survey of Advances in Vision-based Human Motion Capture and Analysis", *Computer Vision and Image Understanding*, v. 104, pp. 90 – 126, 2006.

[82] HU, W., TAN, T., WANG, L., et al. "A Survey on Visual Surveillance of Object Motion and Behaviors", *IEEE Transactions on Systems, Man and Cybernetics - Part C: Applications and Reviews*, v. 34, pp. 334 – 352, 2004.

[83] BRADSKI, G., KAEHLER, A. *Learning OpenCV*. Sebastopol, CA, O'Reilly Media, Inc., 2008.

[84] HARITAOGLU, I., HARWOOD, D., DAVIS, L. "W4: Who? When? Where? What? A Real Time System for Detecting and Tracking People". In: *Third IEEE International Conference on Automatic Face and Gesture Recognition*, 1998.

[85] HARITAOGLU, I., HARWOOD, D., DAVIS, L. "W4: Real-time Surveillance of People and their Activities", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 22, n. 8, pp. 809 – 830, Aug. 2000.

[86] JR., J. C. S. J., JUNG, C. R. "Background Subtraction and Shadow Detection in Grayscale Video Sequences". In: *18th Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI 2005)*, pp. 131–138, 2005.

[87] JR., J. C. S. J., JUNG, C. R., MUSSE, S. R. "A Background Subtraction Model Adapted to Illumination Changes". In: *IEEE International Conference on Image Processing*, pp. 1817–1820, 2006.

[88] JUNG, C. R. "Efficient Background Subtraction and Shadow Removal for Monochromatic Video Sequences", *IEEE Transactions on Multimedia*, v. 11, n. 3, pp. 571 – 577, April 2009.

[89] WREN, C. R., AZARBAYEJANI, A., DARRELL, T., et al. "Pfinder: Real-time Tracking of the Human Body", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 19, n. 7, pp. 780 – 785, July 1997.

[90] STAUFFER, C., GRIMSON, W. E. L. "Adaptive Background Mixture Models for Real-time Tracking". In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, v. 2, pp. 246–252, 1999.

[91] STAUFFER, C., GRIMSON, W. E. L. "Learning Patterns of Activity Using Real-time Tracking", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 22, n. 8, pp. 747 – 757, Aug. 2000.

[92] KAEWTRAKULLPONG, P., BOWDEN, R. "An Improved Adaptive Background Mixture Model for Real Time Tracking with Shadow Detection". In: *2nd European Workshop on Advanced Video Based Surveillance Systems 2001 (AVBS'01)*, 2001.

[93] ZIVKOVIC, Z. "Improved Adaptive Gaussian Mixture Model for Background Subtraction". In: *17th International Conference on Pattern Recognition (ICPR 2004)*, 2004.

[94] ZIVKOVIC, Z., HEIJDEN, F. V. D. "Recursive Unsupervised Learning of Finite Mixture Models", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 26, n. 5, pp. 651 – 656, May 2004.

[95] ZIVKOVIC, Z., HEIJDEN, F. V. D. "Efficient Adaptive Density Estimation per Image Pixel for the Task of Background Subtraction", *Pattern Recognition Letters*, v. 27, n. 7, pp. 773 – 780, 2006.

[96] ELGAMMAL, A. M., HARWOOD, D., DAVIS, L. S. "Non-parametric Model for Background Subtraction". In: *IEEE ICCV Frame Rate Workshop*, 1999.

[97] ELGAMMAL, A. M., HARWOOD, D., DAVIS, L. S. "Non-parametric Model for Background Subtraction". In: *6th European Conference on Computer Vision (ECCV 2000)*, pp. 751–767, 2000.

[98] ELGAMMAL, A. M., DURAISWAMI, R., HARWOOD, D., et al. "Background and Foreground Modeling Using Nonparametric Kernel Density Estimation for Visual Surveillance", *Proceedings of the IEEE*, v. 90, n. 7, pp. 1151 – 1163, July 2002.

[99] OLIVER, N. M., ROSARIO, B., PENTLAND, A. P. "A Bayesian Computer Vision System for Modeling Human Interactions", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 22, n. 8, pp. 831 – 843, Aug. 2000.

[100] PAPOULIS, A. *Probability, Random Variables and Stochastic Processes*. Third ed., McGraw-Hill, 1991.

[101] HAN, B. H., JAIN, R. "Real-time Subspace-based Background Modeling Using Multi-channel Data". In: *Lecture Notes in Computer Science: Advances in Visual Computing*, Springer, pp. 162–172, 2007.

[102] HALL, P., MARSHALL, D., MARTIN, R. "Merging and Splitting Eigenspace Models", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 22, n. 9, pp. 1042 – 1049, 2000.

[103] BARNICH, O., DROOGENBROECK, M. V. "ViBe: A Powerful Random Technique to Estimate the bBackground in Video Sequences". In: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2009)*, pp. 945–948, 2009.

[104] BARNICH, O., DROOGENBROECK, M. V. "ViBe: A Universal Background Subtraction Algorithm for Video Sequences", *IEEE Transactions on Image Processing*, v. 20, n. 6, pp. 1709 – 1724, June 2011.

[105] HYVÄRINEN, A., KARHUNEN, J., OJA, E. *Independent Component Analysis*. New York, John Wiley and Sons, 2001.

[106] TSAI, D.-M., LAI, S.-C. "Independent Component Analysis-based Background Subtraction for Indoor Surveillance", *IEEE Transactions on Image Processing*, v. 18, n. 1, pp. 158 – 167, Jan. 2009.

[107] CHEN, Z. *Object Extraction on Video Sequences Based on Spatiotemporal Independent Component Analysis*. Master Thesis, Ryerson University, Toronto, Ontario, Canada, 1996.

[108] YAMAZAKI, M., XU, G., CHEN, Y. W. "Detection of Moving Objects by Independent Component Analysis". In: *Lecture Notes in Computer Science: Computer Vision – ACCV 2006: 7th Asian Conference on Computer Vision, Proceedings – Part II*, v. 3852, Springer, pp. 467–478, 2006.

[109] SEKKATI, H., LAGANIÈRE, R., MITICHE, A., et al. "Robust Background Subtraction Using Geodesic Active Contours in ICA Subspace for Video Surveillance Applications". In: *Ninth Conference on Computer and Robot Vision (CVR 2012)*, pp. 190–197, 2012.

[110] ZHANG, X.-P., CHEN, Z. "An Automated Video Object Extraction System Based on Spatiotemporal Independent Component Analysis and Multiscale Representation", *EURASIP Journal on Applied Signal Processing*, v. 2006, pp. 1–22, 2006.

[111] CHEN, K. Y., CHIEN, C. C., ZHOU, Y. D., et al. "A Visual-based Dual-target Tracking System Using Independent Component Analysis", *Sensor Letters*, v. 10, n. 5-6, pp. 1216 – 1222, 2012.

[112] FAKHFAKH, N., KHOUDOUR, L., BRUYELLE, J. L., et al. "Intelligence Surveillance System Based on Stereo Vision for Level Crossings Safety Applications". In: El-Alfy, H. (Ed.), *Recent Developments in Video Surveillance*, InTech, 2012.

[113] JIMÉNEZ-HERNÁNDEZ, H. "Background Subtraction Approach Based on Independent Component Analysis", *Sensors*, v. 2010, pp. 6092 – 6114, 2010.

[114] BUCAK, S. S., GUNSEL, B., GÜRSOY, O. "Incremental Nonnegative Matrix Factorization for Background Modeling in Surveillance Video". In: *IEEE 15th Conference on Signal Processing and Communications Applications (SIU 2007)*, pp. 1–4, 2007.

[115] BUCAK, S. S., GUNSEL, B. "Video Content Representation by Incremental Non-negative Matrix Factorization". In: *IEEE International Conference on Image Processing (ICIP 2007)*, v. II, 2007.

[116] LUCAS, B. D., KANADE, T. "An Iterative Image Registration Technique with an Application to Stereo Vision". In: *7th International Joint Conference on Artificial Intelligence (IJCAI'81)*, v. 2, 1981.

[117] SHI, J., TOMASI, C. "Good Features to Track". In: *IEEE Computere Society Conference on Computer Vision and Pattern Recognition (CVPR'94)*, pp. 593–600, 1994.

[118] VIOLA, P., JONES, M. "Robust Real Time Object Detection". In: *IEEE International Conference on Computer Vision (ICCV) - Workshops*, 2001.

[119] JONES, M., VIOLA, P. "TR2003-96 - Fast Multi-view Face Detection". Technical Report, 2003.

[120] PÉREZ, P., HUE, C., VERMAAK, J., et al. "Color-based Probabilistic Tracking". In: *Seventh European Conference on Computer Vision (ECCV 2002)*, pp. 661–675, 2002.

[121] CHENG, Y. "Mean Shift, Mode Seeking, and Clustering", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 17, n. 8, pp. 790 – 799, 1995.

[122] COMANICIU, D., MEER, P. "Robust Analysis of Feature Spaces: Color Image Segmentation". In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 750–755, 1997.

[123] BRADSKI, G. R. "Real Time Face and Object Tracking as a Component of a Perceptual User Interface". In: *Fourth IEEE Workshop on Applications of Computer Vision (WACV'98)*, 1998.

[124] BRADSKI, G. R. "Computer Video Face Tracking of Use in a Perceptual User Interface", *Intel Technology Journal Q2*, pp. 705 – 740, 1998.

[125] COMANICIU, D., MEER, P. "Mean Shift Analysis and Applications". In: *Seventh IEEE International Conference on Computer Vision*, v. 2, pp. 1197–1203, 1999.

[126] COMANICIU, D., MEER, P. "Mean Shift: A Robust Approach toward Feature Space Analysis", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 24, n. 5, pp. 603 – 619, 2002.

[127] COMANICIU, D. "Nonparametric Information Fusion for Motion Estimation". In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, v. 1, pp. 69–76, 2003.

[128] COMANICIU, D., MEER, P. "Kernel-based Object Tracking", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 25, n. 5, pp. 564 – 577, 2003.

[129] FREEMAN, W. T., ROTH, M. "Orientation Histograms for Hand Gesture Recognition". In: *IEEE International Workshop on Automatic Face and Gesture Recognition*, pp. 296–301, 1995.

[130] DALAL, N., TRIGGS, B. "Histograms of Oriented Gradients for Human Detection". In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005)*, 2005.

[131] MORAVEC, H. "Rover Visual Obstacle Avoidance". In: *7th International Joint Conference on Artificial Intelligence (IJCAI'81)*, v. 2, pp. 785–790, 1981.

[132] HARRIS, C., STEPHENS, M. "A Combined Corner and Edge Detector". In: *Fourth Alvey Vision Conference (AVC 88)*, pp. 147–151, 1988.

[133] HARRIS, C. "Geometry from Visual Motion". In: *Active Vision*, MIT Press, pp. 263–284, Cambridge, MA, USA, 1993.

[134] SCHMID, C., MOHR, R. "Local Grayvalue Invariants for Image Retrieval", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 19, n. 5, pp. 530 – 534, 1997.

[135] LOWE, D. G. "Object Recognition from Local Scale-invariant Features". In: *Seventh IEEE International Conference on Computer Vision (ICCV'99)*, v. 2, pp. 1150–1157, 1999.

[136] LOWE, D. G. "Local Feature View Clustering for 3D Object Recognition". In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR'88)*, v. 1, pp. 682–688, 2001.

[137] WREN, C. R., AZARBAYEJANI, A., DARRELL, T., et al. "Distinctive Image Features from Scale-invariant Keypoints", *International Journal of Computer Vision*, v. 60, n. 2, pp. 91 – 110, 2004.

[138] CROWLEY, J. L., PARKER, A. C. "A Representation for Shape Based on Peaks and Ridges in the Difference of Lowpass Transform", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 6, n. 2, pp. 156 – 170, 1984.

[139] LINDBERG, T. P. "On the Construction of a Scale-space for Discrete Images - TRITA-NA-P8808". Internal Report, 1988.

[140] LINDBERG, T. P. "Scale-Space for Discrete Signals", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 12, n. 2, pp. 234 – 254, 1990.

[141] LINDBERG, T. P. "Scale-space Behaviour of Local Extrema and Blobs", *Journal of Mathematical Imaging and Vision*, v. 1, n. 1, pp. 65 – 99, 1992.

[142] MIKOLAJCZIK, K. *Detection of Local Features Invariant to Affine Transformations*. PHD Thesis, Institut National Polytechnique de Grenoble, France, 2002.

[143] LOWE, D. G. "Method and Apparatus for Identifying Scale Invariant Features in an Image and Use of Same for Locating an Object in an Image". 2004. US Patent No.: US 6,711,293 B1.

[144] "SIFT code, demo version". [Online], 2014. Available at `www.cs.ubc.ca/~lowe/keypoints/siftDemoV4.zip`.

[145] VEDALDI, A. "An Open Implementation of the SIFT Detector and Descriptor - Technical Report 070012". University Lecture, 2007. Available at `www.vlfeat.org/~vedaldi/assets/pubs/vedaldi07open.pdf`.

[146] BAY, H., TUYTELAARS, T., GOOL, L. V. "SURF: Speed Up Robust Features". In: *9th European Conference on Computer Vision (ECCV 2006)*, 2006.

[147] BAY, H., ESS, H., TUYTELAARS, T., et al. "SURF: Speed Up Robust Features", *Computer Vision and Image Understanding*, v. 110, pp. 346 – 359, 2008.

[148] KASS, M., WITKIN, A., TERZOPOPULOS, D. "Snakes: Active Contour Models", *International Journal of Computer Vision*, pp. 321 – 331, 1988.

[149] BLAKE, A., ISARD, M. *Active Contours: The Application of Techniques from Graphics, Vision, Control Theory and Statistics to Visual Tracking of Shapes in Motion.* Springer, 2000.

[150] KALMAN, R. E. "A New Approach to Linear Filtering and Prediction Problems", *Transactions of the ASME Journal of Basic Engineering*, v. 82 (Series D), pp. 35 – 45, 1960.

[151] WELCH, G., BISHOP, G. "An Introduction to the Kalman Filter". Technical Report, 1995.

[152] GELB, A., KASPER, J. J. F., NASH, J. R. A., et al. *Applied Optimal Estimation.* Cambridge, The MIT Press, 1974.

[153] RISTIC, B., ARULAMPALAM, S., GORDON, N. *Beyond the Kalman Filter: Particle Filters for Tracking Applications.* Artech House, 2004.

[154] VAN DER MERVE, R., DOUCET, A., FREITAS, N., et al. "The Unscented Particle Filter". In: *Advances in Neural Information Processing Systems 13 (NIPS 2000)*, pp. 584–590, 2000.

[155] ISARD, M., BLAKE, A. "Contour Tracking by Stochastic Propagation of Conditional Density". In: *Fourth European Conference on Computer Vision (ECCV'96)*, v. 1, pp. 343–356, 1996.

[156] ISARD, M., BLAKE, A. "CONDENSATION: Conditional Density Propagation for Visual Tracking", *International Journal of Computer Vision*, v. 29, n. 1, pp. 5 – 28, 1998.

[157] DOLLÁR, P., BELONGIE, S., PERONA, P. "The Fastest Pedestrian Detector in the West". In: *The 21st British Machine Vision Conference (BMVC 2010)*, v. 68, pp. 1–11, 2010.

[158] ADAM, A., RIVLIN, E., SHIMSHONI, I. "Robust Fragments-based Tracking Using the Integral Histogram". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2006)*, 2006.

[159] RUBNER, Y., TOMASI, C., GUIBAS, L. "The Earth Mover's Distance as a Metric for Image Retrieval", *International Journal of Computer Vision (IJCV)*, v. 40, n. 2, pp. 91 – 121,, 2000.

[160] DIHLL, L., JUNG, C. R., BINS, J. "Robust Adaptive Patch-based Object Tracking Using Weighted Vector Median Filters". In: *24th Conference on Graphics, Patterns and Images (SIBGRAPI 2011)*, pp. 149–155, 2011.

[161] BINS, J., DIHLL, L., JUNG, C. R. "Target Tracking Using Multiple Patches and Weighted Vector Median Filters", *Journal of Mathematical Imaging and Vision*, v. 45, pp. 293 – 307, 2013.

[162] JUNG, C. R., SAID, A. "Object Tracking Using Multiple Fragments". In: *16th IEEE International Conference on Image Processing (ICIP 2009)*, pp. 889–892, 2009.

[163] FÜHR, G., JUNG, C. R. "Robust Patch-based Pedestrian Tracking Using Monocular Calibrated Cameras". In: *25th Conference on Graphics, Patterns and Images (SIBGRAPI 2012)*, pp. 166–173, 2012.

[164] FÜHR, G., JUNG, C. R. "Combining Patch Matching and Detection for Robust Pedestrian Tracking in Monocular Calibrated Cameras", *Pattern Recognition Letters*, v. 39, pp. 11 – 20, 2014.

[165] GALL, J., ROSENHAHN, B., BROX, T., et al. "Optimization and Filtering for Human Motion Capture: A Multi-layer Framework", *International Journal of Computer Vision*, v. 87, n. 1, pp. 75 – 92, 2010.

[166] BHATTACHARYYA, A. "On a Measure of Divergence Between Two Statistical Populations Defined by their Probability Distributions", *Bulletim of the Calcutta Mathematical Society*, v. 35, pp. 99 – 109, 1943.

[167] VIOLA, J. J. L. "Double Exponential Smoothing: An Alternative to Kalman Filter-based Predictive Tracking". In: *Workshop on Virtual Environments*, pp. 678–689, 2003.

[168] ASTOLA, J., HAAVISTO, P., NEUVOS, Y. "Vector Median Filters", *Proceedings of the IEEE*, v. 78, pp. 199 – 206, 2003.

[169] KALAL, Z., MIKOLAJCZYK, K., MATAS, J. "Tracking-Learning-Detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 34, n. 7, pp. 199 – 206, 2012.

[170] KALAL, Z., MATAS, J., MIKOLAJCZYK, K. "P-N learning: Bootstrapping Binary Classifiers by Structural Constraints". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2010)*, pp. 49–56, 2010.

[171] KALAL, Z., MIKOLAJCZYK, K., MATAS, J. "Forward-backward Error: Automatic Detection of Tracking Failures". In: *20th International Conference on Pattern Recognition (ICPR 2010)*, pp. 23–26, 2010.

[172] BOUGUET, J. Y. "Pyramidal Implementation of the Lukas Kanade Feature Tracker: Description of the Algorithm". Technical Report, 1999.

[173] PIRSIAVASH, H., RAMANAN, D., KOWLKES, C. C. "Globally-optimal Greedy Algorithms for Tracking a Variable Number of Objects". In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, v. 14, pp. 8–14, 1998.

[174] BABENKO, B., YANG, M.-H., BELONGIE, S. "Visual Tracking with Online Multiple Instance Learning". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2009)*, 2009.

[175] BABENKO, B., YANG, M.-H., BELONGIE, S. "Robust Object Tracking with Online Multiple Instance Learning", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 33, n. 8, pp. 1619 – 1632, Aug. 2011.

[176] KWON, J., LEE, K. M. "Visual Tracking Decomposition". In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2010)*, 2010.

[177] GRABNER, H., BISCHOF, H. "On-line Boosting and Vision". In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2006.

[178] GRABNER, H., GRABNER, M., BISCHOF, H. "Real-time Tracking via On-line Boosting". In: *17th British Machine Vision Conference (BVMC 2006)*, 2006.

[179] GRABNER, H., LEISTNER, C., BISCHOF, H. "Semi-supervised On-line Boosting for Robust Tracking". In: *10th European Conference on Computer Vision (ECCV 2008)*, pp. 234–247, 2008.

[180] STALDER, S., GRABNER, H., GOOL, L. V. "Beyond Semi-supervised Tracking: Tracking Should be as Simple as Detection, but not Simpler than Recognition". In: *IEEE International Conference on Computer Vision (ICCV) - Workshops*, pp. 1409–1416, 2009.

[181] ROSS, D. A., GRABNER, H., GOOL, L. V. "Adaptive Probabilistic Visual Tracking via Incremental Subspace Update". In: *8th European Conference on Computer Vision (ECCV 2004)*, pp. 1409–1416, 2004.

[182] LIM, J., ROSS, D. A., LIN, R.-S., et al. "Incremental Learning for Robust Visual Tracking". In: Saul, L., Weiss, Y., Bottou, L. (Eds.), *Advances in Neural Information Processing Systems 17*, MIT Press, pp. 125–141, 2005.

[183] ROSS, D. A., LIM, J., LIN, R.-S., et al. "Incremental Learning for Robust Visual Tracking", *International Journal of Computer Vision*, v. 77, n. 1-3, pp. 125 – 141, May 2008.

[184] FUKUNAGA, K. *Introduction to Statistical Pattern Recognition.* Academic Press, 1990.

[185] PORIKLI, F. "Integral Histogram: A Fast Way to Extract Histograms in Cartesian Spaces". In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005)*, 2005.

[186] DEMPSTER, A., LAIRD, N., RUBIN, D. "Maximum Likelihood from Incomplete Data via the EM Algorithm", *Journal of the Royal Statistical Society*, v. 38 (Series B), pp. 1 – 38, 1977.

[187] DOLLÁR, P., WOJEK, C., SCHIELE, B., et al. "Pedestrian Detection: A benchmark". In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009)*, 2009.

[188] OLIVEIRA, M. V. "Monitor Portátil Modelo 7013 - Manual de Projeto". Design Manual, 1987.

[189] OLIVEIRA, M. V. "Monitor Portátil Modelo 7013 - Manual de Operação, Rev. 03". Operating Manual.

[190] TEAM, X.-. M. C. ""MCNP - A General N-Particle Transport Code, Version 5", Volume I: Overview and Theory". Technical Report - LA-UR-03-1987, 2003.

[191] CUNHA, V. L. L. *Simulação do Reator Argonauta para Estudo de Viabilidade de Uso em BNCT (in Portuguese)*. Master Thesis, Instituto Militar de Engenharia, Rio de Janeiro, Brazil, 2010.

[192] CUNHA, V. L. L., JUNIOR, W. F. R. S., SILVA, A. X. "Computational Simulation of Argonauta/IEN Nuclear Reactor Using MCNPX Code". In: *International Conference on Mathematics and Computational Methods Applied to Nuclear Science and Engineering (MC 2011)*, 2011.

[193] "AVTech". `http://www.avtech.com.tw/`.

[194] "Alpha Digi". `http://www.alphadigi.com.br/`.

[195] "Alpha Digi ADIP 357". User Manual. Available at `http://www.alphadigi.com.br/downloads/IP%20avtech/Manuais/ManualADIP357.pdf`.

[196] ELLMAUTHALER, A. *Multiscale Image Fusion*. PHD Thesis, Universidade Federal do Rio de Janeiro, Brazil, 2013. Available at `http://www.pee.ufrj.br/teses/textocompleto/2013121001.pdf`.

[197] ELLMAUTHALER, A., SILVA, E. A. B., PAGLIARI, C. L., et al. "A Robust Temporal Alignment Technique for Infrared and Visible-Light Video Sequences". In: *XXXI Simpósio Brasileiro de Telecomunicações (SBrT 2013)*, 2013.

[198] "FFMpeg - A Complete, Cross-platform Solution to Record, Convert and Stream Audio and Video". [Online], 2014. Available at `https://www.ffmpeg.org/`.

[199] CORTEZ, R. A., TANNER, H. G., LUMIA, R., et al. "Radiation Mapping Using Multiple Robots", *Transactions of the American Nuclear Society*, v. 99, pp. 157 – 159, 2008.

[200] CORTEZ, R. A., TANNER, H. G., LUMIA, R. "Distributed Robotic Radiation Mapping". In: Khatib, O., Kumar, V., Pappas, G. (Eds.), *Springer Tracts in Advanced Robotics*, v. 54, Springer, pp. 147–156, 2008.

[201] COSTA, R. G., MÓL, A. C. A., PEREIRA, C. M. N. A., et al. "Mobile Robotic System for Mapping Dose". In: *2011 International Nuclear Atlantic Conference (INAC 2011)*, v. DVD-ROM, 2011.

[202] FISCHLER, M. A., BOLLES, R. C. "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography", *Communications of the ACM*, v. 24, n. 6, pp. 381 – 395, 1981.

[203] SILVA, A. F., THOMAZ, L. A., DE CARVALHO, G. H. F., et al. "An Annotated Video Database for Abandoned-Object Detection in a Cluttered Environment". In: *2014 International Telecommunications Symposium (ITS 2014)*, pp. 1–5, 2014.

[204] LAGANIÈRE, R. *OpenCV 2 Computer Vision Application Programming Cookbook*. Birmingham, Mubai, Packt Publishing Ltd., 2011.

[205] "OpenCV - Open Source Computer Vision". [Online]. Available at http://opencv.org/.

[206] "Fragtrack - Robust Fragments-based Tracking using the Integral Histogram". [Online]. Available at http://www.cs.technion.ac.il/~amita/fragtrack/fragtrack.htm.

[207] "MatLab - The Language of Technical Computing". [Online]. http://www.mathworks.com/products/matlab/.

[208] "TLD - Track-Learning-Detection". [Online]. Available at http://personal.ee.surrey.ac.uk/Personal/Z.Kalal/tld.html.

[209] "Combining Patch Matching and Detection for Robust Pedestrian Tracking in Monocular Calibrated Cameras". [Online]. Available at http://inf.ufrgs.br/~gfuhr/?file=research/multi-people-tracking.

[210] "FPDW - The Fastest Pedestrian Detector in the West". [Online]. Available at http://vision.ucsd.edu/~pdollar/research.html.

[211] "ViBE - A Powerfull Technique for Background Detection and Subtraction in Video Sequences". [Online]. Available at http://www2.ulg.ac.be/telecom/research/vibe/.

[212] JORGE, C. A. F., SEIXAS, J. M., SILVA, E. A. B., et al. "People Detection in Nuclear Plants by Video Processing for Safety Purpose". In: *2011 International Nuclear Atlantic Conference (INAC 2011)*, v. DVD-ROM, 2011.

[213] JORGE, C. A. F., SEIXAS, J. M., SILVA, E. A. B., et al. "Improved People Detection in Nuclear Plants by Video Processing for Safety Purpose". In: *2013 International Nuclear Atlantic Conference (INAC 2013)*, v. DVD-ROM, 2013.

[214] JACCARD, P. "Distribution de la Flore Alpine Dans le Bassin des Dranses et Dans Quelques Régions Voisines", *Bulletin de la Societé Voudoise des Sciences Naturelles*, v. 37, pp. 241 – 272, 1901.

[215] ZHANG, Z., GUNES, H., PICCARDI, M. "Head detection for video surveillance based on categorical hair and skin colour models". In: *16th IEEE International Conference on Image Processing (ICIP 2009)*, 2009.

[216] ZHANG, Z. *Video Tracking of People under Severe Occlusions*. PHD Thesis, University of Technology, Sydney, Australia, 2010.

[217] HU, Y., ZHOU, P., ZHOU, H. "A New Fast and Robust Method Based on Head Detection for People-Flow Counting System", *International Journal of Information Engineering (IJIE)*, v. 1, n. 1, pp. 31 – 40, 2011.

[218] LAZZARI, G. I. *A Method for Detecting People Orientation in Video Sequences (in Portuguese)*. Master Thesis, Applied Computing - Universidade do Vale do Rio dos Sinos, São Leopoldo, Brazil, 2008.

[219] DO AMARAL, W. M. *Tracking People in Video (in Portuguese)*. Master Thesis, Electrical Engineering - Universidade Estadual de Campinas, Campinas, Brazil, 2013.

Appendix A

Flash Synchronization Circuit

The flash synchronization circuit was designed using the timer IC 555 in monostable mode. Figure A.1 shows the circuit schematics and its printed circuit board design. Just a click on its shooting button and the circuit shot the power leds for the chosen time interval - approximately the 1/30 s duration of a frame. Precision fixed-value resistors were used, and also a precise variable one, labelled R8, was used for fine adjustment of the flash duration time. The resistors typically used in series with the transistors' collectors to limit current were not used. Voltage drop over each power led is around 2 V, summing a 8-V drop, and a 9-V battery was used; thus the resistors' values would be very small, and resistor at transistors' base was sufficient. Two batteries were used in parallel due to the needed current for the four parallel leds' branches; the CI 7812T was used in conjunction with the capacitor labelled C4 to keep voltage stable during the flash shootings.

Figure A.2 shows two views of this circuit mounted in its package, in frontal view at top and in oblique view at bottom to show the shooting button. Figure A.3 shows two view of this circuit open.

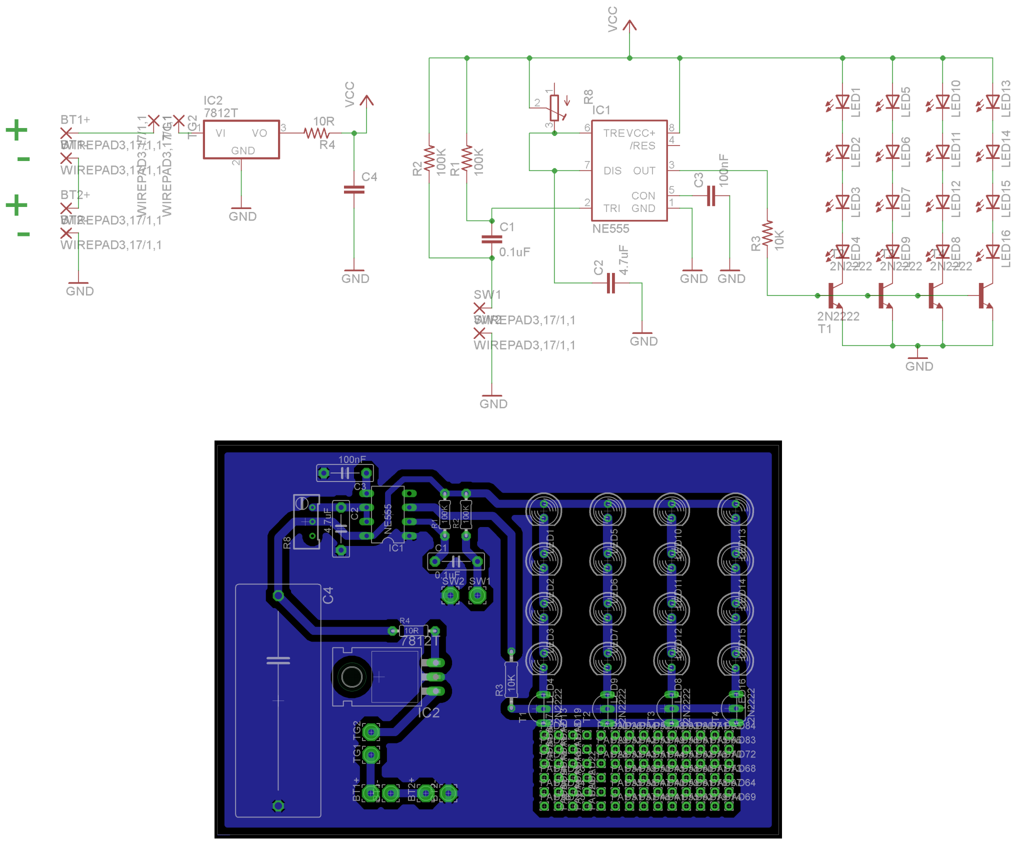

Figure A.1: Flash synchronization circuit schematics and printed circuit board design.

Figure A.2: Flash synchronization circuit in two views.

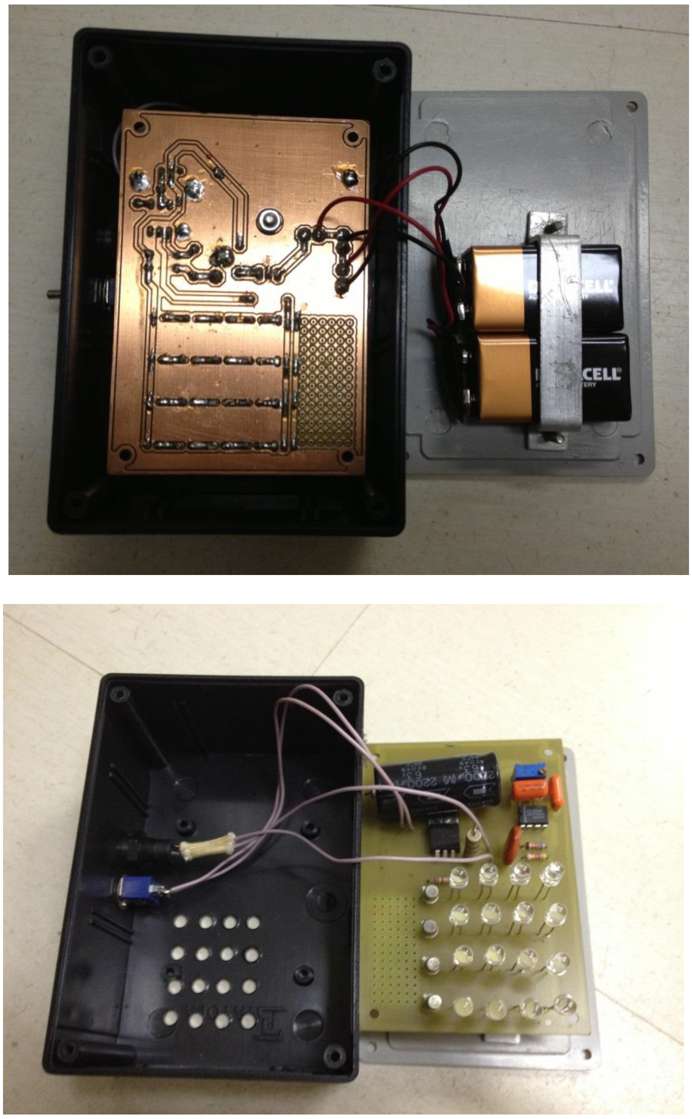

Figure A.3: Flash synchronization circuit in two views.