Universitat
de les Illes Balears

DOCTORAL THESIS
2016

SCENE MODELLING FOR VISION-BASED INTERACTIVE SYSTEMS IN REHABILITATION CONTEXTS

Gabriel Moyà Alcover

Universitat de les Illes Balears

DOCTORAL THESIS
2016

Doctoral Programme in Information and Communications Technology

SCENE MODELLING FOR VISION-BASED INTERACTIVE SYSTEMS IN REHABILITATION CONTEXTS

Gabriel Moyà Alcover

Thesis Supervisor: Javier Varona Gómez
Thesis Supervisor: Antoni Jaume-i-Capó
Thesis tutor: Javier Varona Gómez

Doctor by the Universitat de les Illes Balears

Gabriel Moyà
Scene modelling for vision-based interactive systems in rehabilitation contexts
Documentation, July 4, 2016
Supervisors: Javier Varona Gómez and Antoni Jaume-i-Capó

Universitat de les Illes Balears
Departament de Matemàtiques i Informàtica
UGIVIA

Cra. de Valldemossa, km 7.5.
E-07122 , Palma, Illes Balears

Acronyms

ABDEM Balear Multiple Sclerosis Association.

ADL Activities of Daily Living.

ADO Absent Depth Observations.

ASPACE Spanish Association of Cerebral Palsy Centre.

BBS Berg Balance Scale.

BOS Base of Support.

CI Confidence intervals.

COM Center of Mass.

CP Cerebral Palsy.

FRT Functional Reach Test.

GSM Generic Scene Model.

KDE Kernel Density Estimation.

MMSE Mini-mental state examination.

SINA System of Natural and Advanced Interaction.

TBS Tinetti Balance Section.

TGS Tinetti Gait Section.

TOF Time-of-flight sensors.

TTS Tinetti Total Score.

UGIVIA Computer Graphics, Vision, and Artificial Intelligence Group.

VBI Vision-Based Interfaces.

VR Virtual Reality.

Universitat
de les Illes Balears

Dr Javier Varona Gómez of Universitat de les Illes Balears

I DECLARE:

That the thesis titles *Scene modelling for vision-based interactive systems in rehabilitation contexts*, presented by Gabriel Moyà Alcover to obtain a doctoral degree, has been completed under my supervision and meets the requirements to opt for an International Doctorate.

For all intents and purposes, I hereby sign this document.

Signature

Palma de Mallorca, Monday July 4, 2016.

Universitat
de les Illes Balears

Dr Antoni Jaume-i-Capó of Universitat de les Illes Balears

I DECLARE:

That the thesis titles *Scene modelling for vision-based interactive systems in rehabilitation contexts*, presented by Gabriel Moyà Alcover to obtain a doctoral degree, has been completed under my supervision and meets the requirements to opt for an International Doctorate.

For all intents and purposes, I hereby sign this document.

Signature

Palma de Mallorca, Monday July 4, 2016.

Publications and contributions

Journals

- Alejandro Reyes-Amaro; Yanet Fadraga-González; Oscar Luis Vera-Pérez; Elizabeth Domínguez-Campillo; Jenny Nodarse-Ravelo; Alejandro Mesejo- Chiong, Gabriel Moyà-Alcover; Antoni Jaume- i -Capó. *Rehabilitation of patients with motor disabilities using computer vision based techniques*; Journal of accessibility and design for all, 2012.

  Contribution: Gabriel Moyà-Alcover participated in the serious game design and participated in the system development. Gabriel Moyà-Alcover also contributed to writing the paper.

- Jaume-i-Capó, A.; Martínez-Bueso, P.; Moyà-Alcover, G. ; Varona, J.; *Interactive Rehabilitation System for Improvement of Balance Therapies in People With Cerebral Palsy*; IEEE Transactions On Neural Systems and Rehabilitation Engineering, 2014; Impact Index: 3.188 (JCR 2014).

  Contribution: Gabriel Moyà-Alcover participated in the system design, was responsible for the system development and participated in the experiment execution. Gabriel Moyà-Alcover also contributed to writing parts of section II and revised the paper. In this paper authors are listed in alphabetical order.

- Jaume-i-Capó, A.; Martínez-Bueso, P.; Moyà-Alcover, G.; Varona, J.; *Improving vision-based motor rehabilitation interactive systems for users with disabilities using mirror feedback*; The Scientific World Journal, 2014; Impact Index: 1.219 (JCR 2013).

  Contribution: Gabriel Moyà-Alcover participated in the experimental system design, was responsible for all the experimental development used in this work and participated in the experiment execution and system evaluation. Gabriel Moyà-Alcover contributed to writing the paper. In this paper authors are listed in alphabetical order.

- Gabriel Moyà-Alcover, Ahmed Elgammal, Antoni Jaume-i-Capó, Javier Varona; *Modelling depth for nonparametric foreground segmentation using RGBD devices*; Pattern Recognition Letters; Elsevier; [Minor Revision, 9th June 2016].

Contribution: Gabriel Moyà-Alcover was responsible for the algorithm design, the implementation and the algorithm evaluation. Gabriel Moyà-Alcover wrote the core of the paper.

- Ines Ayed, Gabriel Moyà-Alcover, Pau Martínez-Bueso, Javier Varona, Adel Ghazel, Antoni Jaume-i-Capó; *Validación de dispositivos RGBD para medir terapéuticamente el equilibrio: el test de alcance funcional con Microsoft Kinect*; Revista Iberoamericana de Automática e Informática Industrial; Elsevier; 2016; Impact Index: 0.475 (JCR 2015);

  Contribution: Gabriel Moyà-Alcover participated in the system design and performing the experiments. Gabriel Moyà-Alcover also contributed to writing the paper.

Proceedings

- Moyà-Alcover, G.; Jaume-i-Capó, A.; Varona, J.; Martínez-Bueso, P.; Mesejo Chiong, A; *Use of serious games for motivational balance rehabilitation of cerebral palsy patients*; 13th international ACM SIGACCESS conference on computers and accessibility, 2012.

  Contribution: Gabriel Moyà-Alcover participated in the serious game design and was responsible for all the developments used in this work. Gabriel Moyà-Alcover also contributed to writing the paper and made the oral presentation of this article.

- Jaume-i-Capó, A.; Moyà-Alcover, G.; Varona, J.; Martínez-Bueso, P.; Mesejo Chiong, A.; *Motivational rehabilitation using vision-based serious games*; Ninth IASTED International Conference on Biomedical Engineering, 2012.

  Contribution: Gabriel Moyà-Alcover participated in the game design, was responsible for all the developments used in this work and participating in writing sections 3 and 4. Gabriel Moyà-Alcover also revised the paper and made the oral presentation of this article.

- Ines Ayed, Gabriel Moyà-Alcover, Pau Martínez-Bueso, Adel Ghazel, Javier Varona, Antoni Jaume-i-Capó, Francisco J. Perales; *RGBD-based Serious Games for Fall Prevention in Elderly People*; Cognitive Area Networks, vol. 1, no 3, Junio 2016.

Contribution: Gabriel Moyà-Alcover participated in the system design, the model definition and performing the experiments. Gabriel Moyà-Alcover also contributed to writing the paper.

Book chapters

- Manresa Yee, C.; Mas, R.; Moyà-Alcover, G.; Abásolo, M.J.; Giacomantone, J.; *Interactive Multi-sensory Environment to Control Stereotypy Behaviours*; Computer Science and Technology Series: XVII Argentina Congress of Computer Science. Selected Papers; *Ed*; Edulp; 2012.

  Contribution: Gabriel Moyà-Alcover contributed to this work developing the interaction system, including the interaction mechanism and the user movements detection. Gabriel Moyà-Alcover also revised the book chapter.

- Antoni Jaume-i-Capó, Gabriel Moyà-Alcover, Javier Varona; *Design Issues for Vision-Based Motor-Rehabilitation Serious Games*; Technologies of Inclusive Well-Being; *Ed:* Springer; 2014.

  Contribution: Gabriel Moyà-Alcover contributed to this work helping to write and revising the book chapter.

Internship

- CBIM (Computational Biomedicine Imaging and Modelling Center), Rutgers, New Jersey USA. September 21th December 19th, 2014.

Projects

- Aplicaciones basadas en visión para la rehabilitación. A/023109/09. Responsable Antoni Jaume-i-Capó. 2009-2010. Cuba.

- Formación de recursos humanos para el desarrollo de actividades docentes, investigativas y laborales en aplicación de tecnicas computacionales, especialmente la visión computacional, para la rehabilitación; PCIB - Programa de Cooperación Interuniversitaria e Investigación Científica entre España e Iberoamérica (PCI-Iberoamérica); A2/037538/11; IP: Antoni Jaume-i-Capó. 2012.

- Aplicaciones basadas en visión para la rehabilitación A/030033/10. Responsable: Antoni Jaume-i-Capó. 2011-2012 Cuba.

- Experiencias de diseño y desarrollo de interfaces naturales en industria, educación y rehabilitación; PNIF - Programa Nacional de Investigación Fundamental No Orientada; TIN2012-35427; IP: Javier Varona; 2013-2016.

- Assessorament en tasques de I+D dins del projecte NEOTEC, Plataforma de desarrollo de Aplicaciones Basadas en Interacción Natural; IDI-20140183; Inisle Interactive Technologies S.L. IP: Javier Varona; 2015.

- Formación en técnicas de procesamiento y análisis de imágenes como soporte en el diagnóstico de la Sicklemia y los transtornos neurofisiológicos asociados; Projectes de Cooperació al desenvolupament. UIB; OCDS-CUD2014/10; IP: Antoni Jaume-i-Capó; 2014-2015.

Agraïments

Aquests darrers anys han estat molt intensos, una etapa d'aprenentatge en la que un deixa de ser un jove que es passeja pel món amb els calçons baixos, esdevé adult i ha de començar a assumir les decisions que pren. Fent cap enrere, em consider molt afortunat, les decisions que he pres fins ara m'han dut aquí, un lloc on no em pensava arribar. També és vera que tot això no ho he fet jo sol i just ara és el moment en el qual puc agrair l'ajuda rebuda.

En primer lloc vull destacar la dedicació d'en Xavi i en Toni. Els que han estat els meus guies en aquesta aventura i amb els que fa més de 6 anys que he estat aprenent, no només en l'àmbit de la investigació. Dues persones que han esdevengut quelcom més que companys de feina i han aconseguit que la realització d'aquesta tesi hagi estat un procés que he gaudit enormement. Moltíssimes gràcies!

Voldria agrair als meus pares tot allò que han fet per possibilitar que jo avui estigui escrivint aquestes línies. A na Marta, que has fet molt més del que et puc agrair, a la teva paciència. També voldria tenir un record per na Queta, i en Ramón. M'agradaria enviar una besada a n'Ona, esper que algun dia ho pugui llegir i li faci gràcia. Al meu padrí jove en Guillem que sempre em demana per la tesi i escolta amb atenció tot el que li dic. És clar que també a la resta de família. A na Marga per què n'és una més. Als que per desgràcia ja no hi són.

Aquests anys a la universitat he conegut gent prou important per al meu creixement personal, per aquest motiu voldria tenir un record cap als que han estat els meus companys de feina. En particular a en Jose Maria, que com a company de despatx n'ha hagut d'aguantar moltes. A na Pau, peça indispensable del nostre equip de feina i consellera esportiva. També als que em veieu cada dia i heu compartit amb mi tots els bons moments, al bar fent un cafè, als congressos o en el metro de tornada a casa: en Paco, na Sílvia, en Miquel, en Manolo, na Cristina, en Ramón i n'Esperança. També als altres companys doctorands i professors del departament.

Aquest treball no hauria estat possible sense la col·laboració de tota la gent que ha treballat amb nosaltres a ASPACE i ABDEM, no només els professionals que ens han posat totes les facilitats del món. Als pacients que sempre tengueren una actitud positiva envers de les nostres visites.

Els amics... Als meus amics de la vida: en Juanfran, Chuso, Jordi i Jaume que han suportat des de fa molts anys les meves dèries setmana rere setmana, gràcies!. Als amics de la universitat amb els quals tot va ser més fàcil (i divertit): Alvaro, Dome, Miki, Pepote i Xavi, els quals em varen animar i engrescar per començar amb aquest camí. També als altres companys amb els quals vaig compartir el temps a la universitat.

Hi ha moments que marquen vides, diria que un d'aquests moments el vaig viure durant el meu període a l'esplai. Voldria agrair tot allò que hem fet plegats i que d'altra manera no hagués pogut fer mai: a n'Angel, Mike, Colman, Cloquell, Faust, Veli, Cati, Meli, Marga, Laura... Vos estim a tots igual! Als companys amb els quals he participat en el projecte contrainfo.cat, perquè no tota feina es tesi en aquesta vida, en especial en Gaspar amb el qual vaig compartir laboratori una bona temporada.

He gaudit de moltes experiències aquests quatre anys, però potser poques comparables a l'estada que vaig realitzar als Estats Units, i encara que sé que mai ho llegiran, voldria agrair al grup d'investigació del CBIM de Rutgers el tracte que em dispensaren durant el temps que vaig treballar amb ells. També voldria agrair de tot cor als meus dos companys de pis David i Will que em tractaren com un més de la família cada un dels dies que vaig viure amb ells. També a totes les persones que vaig conèixer allà i que em permeteren viure un dels millors períodes de la meva vida, en especial en Manuel.

A en Teix que fa dos mesos que és a casa i l'estimam com si fes 10 anys, a en Quin perquè sempre el recordarem

Contents

Abstract

The aim of this thesis is to improve vision-based interaction in rehabilitation systems. We defined and evaluated a new method of modelling scenes in the RGBD space in order to generalize the use of this kind of systems in real environments.

This process has been conducted through the design and development of a vision-based interactive system in a context of rehabilitation application. We first defined a vision-based interaction system, then we developed a motivational video game. After first testing, therapists concluded that the system implemented the balance rehabilitation therapy. Preliminary results showed that users performed the rehabilitation activity in 13.5 % less time. Using the presented video game, no users have abandoned and they showed interested in continuing the rehabilitation process.

We also designed an experiment in order to test the feasibility and benefit of including the mirror feedback mechanism in vision-based rehabilitation systems, because during the system development we noticed that users could not match their movements with what they saw on the screen. We define mirror feedback as the visual representation of the users inside the application as the interaction feedback. Results confirmed that in case of people with disabilities the mirror feedback facilitates the interaction.

Finally, we clinically evaluate the designed system. The evaluation consisted on a 24-week physiotherapy intervention program conducted with 9 adults from a cerebral palsy center who exercised weekly in 20-minute sessions. Findings demonstrated a significant increase in balance and gait function scores resulting in indicators of greater independence for our participating adults. Scores improved from 16 to 21 points in a scale of 28, according to the Tinetti Scale for risk of falls, moving from *high fall risk* to *moderate fall risk*. From obtained results we can conclude that our experimental system is feasible for balance rehabilitation therapy.

During this process we noticed that environmental conditions had a big influence in the video game performance so we have developed a scene modelling algorithm using RGBD information in order to segment the user from scene. We constructed our background model using a Kernel Density Estimation (KDE) process with a Gaussian Kernel. We used a three dimensional kernel, one dimension to model depth information and two for normalised chromaticity coordinates. We modelled sensor *absent depth observations* using a

probabilistic strategy in order to distinguish which belongs to the background model and which are provoked by foreground objects in order to detected those ones that are induced by foreground objects. That pixels cannot be classified as background or foreground so we used a third classification class, we called undefined, in order to classify that pixels. In addition, we developed an algorithm to detect changing background objects in the same frame they move based on the cdf of the pixel model. Two strategies are described in order to adapt the update phase to the different nature of the color and depth information, considering color information as short-term model and depth as long-term one.

After evaluating the proposed scene modelling approach with two datasets, results showed that it can handle several practical situations and obtain good results in all cases.

Resum

L'objectiu d'aquesta tesi és millorar la interacció basada en visió en sistemes de rehabilitació. Per aquest motiu hem construït i avaluat un nou mètode de modelat d'escenes en l'espai RGBD, per tal de generalitzar l'ús d'aquests sistemes en entorns reals.

Aquest procés s'ha dut a terme mitjançant el disseny i desenvolupament d'un sistema interactiu basat en visió per computador en el context d'una aplicació de rehabilitació. En primer lloc, hem definit un sistema d'interacció basat en visió, en segon lloc hem desenvolupat un videojoc motivacional.

Després d'una primera fase de proves, els terapeutes varen concloure que el sistema implementa la teràpia de rehabilitació de l'equilibri que es volia transferir. Els resultats preliminars també varen mostrar que els usuaris duien a terme l'activitat de rehabilitació amb un 13,5 % menys de temps. Amb l'ús el videojoc presentat, cap dels usuaris va abandonar la teràpia i tots es mostraren interessats a continuar el procés de rehabilitació amb el nostre sistema.

A continuació, es va dissenyar un experiment per posar a prova la viabilitat i el benefici d'incloure el mecanisme de retroalimentació de mirall en els sistemes de rehabilitació basats en visió. Aquest experiment, va ser motivat per què durant el desenvolupament del sistema abans descrit, ens adonarem que els usuaris no podien equiparar els seus moviments amb el que succeïa a la pantalla. Definim retroalimentació de mirall com la representació visual dels usuaris dins l'aplicació. Els resultats d'aquest experiment varen confirmar que en cas de persones amb discapacitat, la retroalimentació de mirall facilita la interacció.

Finalment, es va avaluar clínicament el sistema proposat. L'avaluació va consistir en un programa d'intervenció de fisioteràpia durant 24 setmanes. Va ser realitzat amb 9 adults d'un centre de paràlisi cerebral, realitzant sessions setmanals de 20 minuts. Els resultats varen demostrar un augment significatiu en les puntuacions d'equilibri i funció de la marxa indicant una major independència per als pacients. Les puntuacions varen millorar de 16 a 21 punts en una escala de 28, d'acord amb l'Escala de Tinetti per al risc de caigudes, passant de *risc alt de caiguda* a *risc moderat de caiguda*. Dels resultats obtinguts es pot concloure que el nostre sistema experimental és viable per a la teràpia de rehabilitació de la marxa i l'equilibri.

Durant el procés abans descrit, ens adonarem que les condicions ambientals tenen una gran influència en el rendiment del videojoc, per aquest motiu, hem desenvolupat un algoritme de modelatge d'escenes utilitzant informació RGBD per tal de segmentar l'usuari de l'escena. Hem desenvolupat un nou enfocament no paramètric que unifica els diversos canals d'informació del dispositiu. També s'ha desenvolupat un model probabilístic que permet manejar aquella informació inexistent en la imatge de profunditat. Finalment s'ha creat un algorisme per detectar aquelles zones de la imatge en les que un objecte de l'escena és mogut i d'aquesta manera evitar errors de classificació.

Després d'avaluar l'algorisme proposat utilitzant dos conjunts de dades, els resultats mostraren que l'algorisme pot manejar un gran ventall de situacions pràctiques i obté bons resultats en tots els casos provats.

1 Introduction

I begin with an idea and then it becomes something else.

— Pablo Picasso

1.1 Motivation

The Computer Graphics, Vision, and Artificial Intelligence Group (UGIVIA) that belongs to the Mathematics and Computer Science Department at the Universitat de les Illes Balears is the research unit where we developed this work. They have been working on vision-based human computer interaction projects for a long time. As the most prominent examples, they developed the System of Natural and Advanced Interaction (SINA), a pedagogic system in order to make computer accessible to people with disabilities. This accessibility is achieved through computer vision techniques, capturing images of the real world using a common webcam and installing a small application that detects and follows the movement of the users' nose, so the user can control the mouse movement and its events. This work was tested in the Spanish Association of Cerebral Palsy Centre (ASPACE) and Balear Multiple Sclerosis Association (ABDEM) with very good outcomes. As a result of these successes a new branch of SINA project emerged: SINAeyes pursuing the same objective of the original SINA project, but changing the nose tracking algorithm for a new eye tracker. Using this software, researchers discovered that SINAeyes provides a tool for rehabilitation, since they observed that people who could not keep the trunk aligned, could do it after using the program for a certain period of time.

Each year, a part of patients working with ASPACE abandon their therapy due to loss of motivation. Cerebral Palsy (CP) is a term used to describe a group of chronic conditions affecting body movement and muscle coordination. CP is the most common cause of disabling conditions in children due to the increased survival of low birth-weight infants [60]. The population of adults with CP is growing, as a result of increased longevity, inspiring new research to improve available therapies to achieve better functional abilities. With adequate treatment, their quality of life can be improved.

Physiotherapy treatment in balance control is important for competence in the performance of most functional skills. This treatment can help individuals suffering from CP to recover from instability when performing an action that is close to the edge of the stability limit. The objectives of medical intervention and physical therapy in CP patients are to improve balance and postural control, to prevent dependence and to conserve autonomy.

Physiotherapy exercises must be repeated weekly, so in long-term rehabilitation programs, patients become demotivated [16, 82, 34, 47]. As a result, patients lose focus on the therapy program, and the therapy loses effectiveness. We know that rehabilitation results are better when patients are motivated [67]. Additionally, demotivation can cause resignation [16] when focusing on rehabilitation for maintaining patient abilities in situations where patients rarely improve.

Nonetheless, if exercises for postural control and balance are based on small achievements in a game and provide continuous feedback to the patient, patients can achieve better motor control [87].

Scientists in our research group also developed a vision-based natural gesture recognition algorithm that enables tracking the users pose in real time [20, 105]. This lets to think ASPACE physiotherapists in combination with UGIVIA researchers that it was possible to use computer vision programs in order to perform rehabilitation exercises, that is, a set of movements of one or more of the body parts under a set of guidelines in order to achieve the highest level of function, independence, and quality of life possible.

1.2 Vision-based interaction

Vision-Based Interfaces (VBI) [104] seeks to provide a wider and more expressive range of input capabilities by using computer vision techniques to process sensor data from one or more cameras in real-time, in order to reliably estimate relevant visual information about the user.

Visual information from the performance of patient actions is a good capture method in motor rehabilitation for two reasons: first, motor rehabilitation consists of body movements that can be recorded; second, VBI is non-invasive and can be used for patients who have difficulties in holding physical devices. In general, VBI systems aim to provide reliable computer methods to detect and analyse human movements. The process is repeated over time, enabling monitoring of the user's actions. Depending on the computer vision technique chosen (e.g., interaction with the silhouette, arms or hands), different levels of precision can be achieved for the user interaction.

In rehabilitation systems using VBI is critical to provide feedback to users in order to feel in control and helping them to understand what is happening [16, 52, 54]. In VBI there is not contact with the interface by means of an interaction device of reference. The user, therefore, always should know when interaction is taking place using visual and audible feedback.

1.3 Serious games for rehabilitation

Video games are a part of our lives, and similar to other technologies, these games may have multiple goals. Such games can be used for entertainment and other purposes. When a video game is designed to allow the user to reach a specific goal (e.g., in education, health, public policy, strategic communications or the military), with entertainment and engagement elicited by the experience of playing, it is known as a serious game [71]. The primary purpose of a serious game is different from the purpose of pure entertainment, and the cognitive and motor activities required by serious games engage users' attention [76] and this helps to distract the user from the task [41, 59].

Research studies show serious games help to motivate users in rehabilitation processes [84]. The cognitive and motor activity required by video games engage the user's attention. In addition, users focus their attention on the game and this helps them in forgetting that they are in therapy.

1.4 Thesis outline

The main objective of this work is defined by the next two tasks:

- Design, develop and clinically evaluate vision-based interactive systems in a context of rehabilitation applications.
- Create and evaluate a new method of modelling scenes in the RGBD space in order to generalize the use of vision-based interactive systems in real environments.

1.4.1 Interactive systems for balance rehabilitation

In order to test if serious games can be used for motivational balance rehabilitation in cerebral palsy patients, we develop a serious game and objectively investigate the game's clinical usefulness to improve therapy. We present an experimental system that consist on transferring the ASPACE balance therapy tasks to a serious game in order to experiment if it is valid for motivational rehabilitation.

The advantage of observation and imitation for learning is well studied [109, 50], and mirror movements and imitation learning is recommended in motor rehabilitation [15]. Motor control amends the motion by interaction between visual feedback that recognizes the external space or movement of oneself through vision feedback that refers information about movement and position of body [23]. Moreover, there is evidence that action observation facilitate motor activity [38]. For this reason, mirrors equip motor therapy rooms and they allow the patients see themselves in order to perform correctly the therapy. Our objective is to explore how mirror feedback through interaction can be included into game interaction design in order to observe whether it is possible to improve results in rehabilitation sessions. We perform a user study testing to explore mirror feedback in vision-based video games.

Furthermore, we study and rigorously evaluate the effects of physiotherapy treatment on balance and gait function of adult subjects with cerebral palsy undergoing our experimental system.

Finally, from our experience of implementing vision-based motor-rehabilitation serious games, we present a set of design issues defining an interaction model adapted to the user's capabilities and following the desirable features for rehabilitation serious games.

1.4.2 Modelling depth for nonparametric foreground segmentation

During the video game development we have noticed that environmental conditions have a big influence on the game. First, due to the dependence on real-world conditions (e.g., lighting, distances, and clothes), the interaction environment limit the techniques that can be used. Second, users were sometimes distracted by what was happening around them and did not pay attention to the game. For those reasons, we developed a computer vision algorithm, in order to segment the user from the scene background.

Scene modelling is a widely used technique for detecting moving foreground objects in image sequences. Foreground segmentation, provide an important cue for numerous applications in computer vision as: surveillance, tracking, recognition, human poses estimation among others. The main objective is to detect objects that do not belong to the scene, by comparing the current observation with previous references.

We present a new scene modelling approach, that uses both depth and color information from RGBD sensors. We construct a unified nonparametric background model for each pixel of the scene and we estimate the probability that a newly observed pixel value belongs to that model.

Interactive systems for postural control and balance rehabilitation

2

Research studies show that serious games help to motivate users in rehabilitation and therapy is better when users are motivated. In this chapter we describe the process of experimenting with serious games for cerebral palsy patients, who rarely show capacity increases with therapy, which causes them demotivation. For this reason, we have designed, implemented and evaluated balance rehabilitation video games for this group of patients. The employed interaction technology is based on computer vision.

2.1 Introduction

This chapter covers a new experimental system designed to improve the balance and postural control of adults with cerebral palsy. Balance control is important for competence in the performance of most functional skills. A lack of or inappropriate skills for this control results in the risk of falls and/or diminished quality of life. Physiotherapy treatment can help individuals suffering from CP to recover from instability when performing an action that is close to the edge of the stability limit. The general objective of the rehabilitation in balance and postural control is to achieve motor automatisms (automatic movements) that enable the patient to have autonomous motor behaviour. This system is based on a serious game to transfer a balance rehabilitation therapy, designed using the prototype development paradigm and features for rehabilitation with serious games: feedback, adaptability, motivational elements and monitoring. In addition, the employed interaction technology is based on computer vision because motor rehabilitation consists of body movements that can be recorded, and because vision capture technology is noninvasive and can be used for patients who have difficulties in holding physical devices.

During this process we noticed that the feedback of current serious games was not enough for understanding the game play by users with disabilities.

For this reason we designed an experiment in order to test the feasibility and benefit of including the mirror feedback mechanism in vision-based rehabilitation systems. We defined mirror feedback as the visual representation of the users inside the application such as interaction feedback.

We rigorously evaluated the effects of physiotherapy treatment on balance and gait function of adult subjects with cerebral palsy undergoing our experimental system. A 24-week physiotherapy intervention program was conducted with 9 adults from a cerebral palsy center who exercised weekly in 20-minute sessions. Findings demonstrated a significant increase in balance and gait function scores resulting in indicators of greater independence for our participating adults.

Finally, we present an implementation guidelines for developing serious-games as motivational tool for rehabilitation therapies.

This chapter is organized as follows: next section is dedicated to describe the related work. Second, we explain how we developed a motivational vision-based serious game. Next we explain how we improved the interaction, validating the mirror feedback. In fourth section, we describe the process of clinical evaluation of the interactive system. Using our experience in vision-based serious-games we present a set of guidelines in fifth section.

2.2 Related work

This section reports the related work in three different themes: first, we describe the relation between serious games and rehabilitation processes. Second, we describe the vision based interaction and the used technology in our interactive systems. Finally, we describe the existing development recommendations in rehabilitation interactive systems.

2.2.1 Serious Games and Rehabilitation

In [111], a taxonomy including the terms *Game*, *Video game* and *Serious Game* is reported. A *Game* is a physical or mental contest played according to specific rules, with the goal of amusing or rewarding the participant. A *Video game* is a mental contest played with a computer according to certain rules and for amusement, for recreation or to win a stake. A *Serious Game* is a mental contest played with a computer according to specific rules that uses entertainment to further government, corporate training, education, health, public policy or strategic communication objectives. The cognitive and motor activities required by serious games engage users' attention [76] and this helps to distract the user from the task [41, 59].

Healthcare-related serious games can be focused on treatment, recovery and rehabilitation. Research studies show that serious games are highly promising in rehabilitation processes [88]. Specifically, in [72], it is demonstrated that serious games help to motivate patients in therapy sessions. This motivation is particularly important in long-term rehabilitation for maintaining motor abilities. In this case, demotivation is frequent in chronic patients because therapy usually consists of repetitive and intensive activities that become boring after hundreds of sessions. As a result, the patient does not focus on the therapy program, thereby risking losing the benefits of the therapy.

In previous work, we can find different serious games for different types of rehabilitation. Regarding upper limb rehabilitation, [66] presented a serious-game based movement therapy which aims to encourage stroke patients with upper limb motor disorders to practice physical exercises, [17] showed VR system for stroke patients, [16] designed several serious games which use low-cost webcams as input technology to capture data of user's movements, [28] created a simple game in which the patient tried to move a coloured circle from an initial position to a goal position using a robotic device designed for arm rehabilitation, in [51] implemented a haptic glove serious game for finger flexion and extension therapy, [1] presented several home-based serious games which use a webcam and a $Wiimote^{TM}$, and [2] designed a low-cost Virtual Reality (VR)-based system using $Wiimote^{TM}$.

Various serious games have been presented for balance rehabilitation. In a report by Betker et al. [9], three serious games were controlled by use of center-of-pressure signal for the maintenance of balance in a short-sitting position in cases of spinal cord and head injuries. The center-of-pressure was acquired via a flexible pressure mat implemented by a $16 \quad 16$ grid of piezoelectricity-resistive sensors. The games were evaluated using a questionnaire administered after the exercises and stability measurements obtained during a set of tasks performed before and after exercise. The patients increased time per session and also number of sessions per week, showing an increment of attention during training with the game-based tool. Observations also indicated that the serious game can have a substantial positive effect by improving dynamic short-sitting balance. Nevertheless, no clinical or analytical study of the results was presented. This review discussed the lack of evidence in clinical evaluation in studies [9]. Although there is evidence that balance training improves postural control, there is a lack of studies that use balance measures, whether static or dynamic, after a serious game intervention therapy program for middle-aged adults with CP [6].

2.2.2 Vision Based Interaction

Video game console technologies focus on motion-based inputs, designed to track body motions or body reactions (e.g. $EyeToy^{TM}$, $Wiimote^{TM}$, $Kinect^{TM}$, $Xtion^{TM}$ and $Move^{TM}$), are becoming popular and low-cost [97, 99]. These sensors can capture motions of the motor therapy and different studies validated $Kinect^{TM}$ sensor for rehabilitation purposes such as postural control [25], clinical functional analysis and rehabilitation [11], gait retraining [26], activities of daily living rehabilitation [27], and coaching of elderly population [77].

Different studies concluded that existing commercials motion-based video games were difficult to use in rehabilitation therapy, because they were designed for users with full capabilities [2, 89]. Therefore, researchers developed motion-based video games for motor rehabilitation using the existing commercial motion-based devices: pressure mat based for maintenance of balance in a short-sitting position following spinal cord and head injuries [9]; vision-based for upper limb stroke rehabilitation [16], for chronic stroke recovery [42, 47, 83] and to improve the balance and postural control of adults with cerebral palsy [52]; $Wiimote^{TM}$-based for postural control and functional mobility of cerebral palsy patients [31]; $Kinect^{TM}$-based to guide and correct of therapeutic movements [30], to train static balance [61] and to improve the motor proficiency and quality of life [22]; and haptic-based for stroke rehabilitation [82]. Also, literature reviews about motion-based rehabilitation system were published in the last years [84, 88, 63, 69]. In particular, $Kinect^{TM}$ sensor captures the visual information of the performance of user motions, then it can also considered a vision-based interaction sensor (VBI) [104].

In addition, in rehabilitation systems using VBI is critical to provide feedback to users in order to feel in control and helping them to understand what is happening [16, 52, 54]. In VBI there is not contact with the interface by means of an interaction device of reference. The user, therefore, always should know when interaction is taking place using visual and audible feedback.

The advantage of observation and imitation for learning is well studied [109, 50], and mirror movements and imitation learning is recommended in motor rehabilitation [15]. Motor control amends the motion by interaction between visual feedback that recognizes the external space or movement of oneself through vision feedback that refers information about movement and position of body [23]. Moreover, there is evidence that action observation facilitate motor activity [38]. For this reason, mirrors equip motor therapy rooms and they allow the patients see themselves in order to perform correctly the therapy.

In fact, some exiting VBI rehabilitation systems allow the patient see themselves on screen, mirror feedback, due to the users stand in front of a screen and interact with the system using their movements [16, 42, 22]. It was demonstrated that own image of the user suggested more realism and sense of presence than an avatar figure [73]. The more sense of presence, the more aware of their position and orientation with respect to the interaction elements the users are. Nevertheless, other VBI rehabilitation systems do not implement the mirror feedback [47, 30, 61] because game-based rehabilitation systems designers frequently overemphasize the video game rather than the user interaction. When these games are designed for people with disabilities, the interaction design issues are fundamental to achieve a high patient's motivation. In addition, game interaction design is usually defined without taking into account user's perceptions with regard to their actions in order to achieve the rehabilitation goals.

Different researchers studied the importance and the effectiveness of the augmented feedback in the therapy (information that cannot be elaborated without an external source such as a therapist or a device) [106, 91]. They discovered that visual augmented feedback could improve the performance of the patients on complex motor tasks. However, to our knowledge, it did not exist any study about the importance of the mirror feedback in vision-based rehabilitation systems. Concretely, Sigrist's survey [91] reviewed different types of natural visualization feedback (such as superposition, side-by-side 3D perspective, end-effector movements and third- and first-person perspective) and only introduced the mirrors in the case of mirror therapy [21].

2.2.3 Development recommendations in rehabilitation interactive systems

Recent research studies have proposed what features are desirable for rehabilitation serious games. [54] proposed target audience, visibility and feedback as important human factors, [16] identified two principles of game design theory which have particular relevance to rehabilitation: meaningful play, the relationship between player's interactions and system reaction, and challenge, maintaining an optimum difficulty is important in order to engage the player. [84] identified as important main criteria for the classification of serious games in the rehabilitation area: application area, interaction technology, game interface, number of players, game genre, adaptability, performance feedback, progress monitoring and game portability. [1] concluded that serious games must ensure that patients are correctly performing and must provide a motivating context for therapy, in order to have maximum impact on the rehabilitation process.

A common conclusion of previous studies is that existing commercial video games are difficult to use in rehabilitation therapy because the games were designed for users with full abilities. Anderson et al. [2] specifically enumerate the problems associated with commercial video games when used in rehabilitation; that is, the games mainly target upper-body gross motor function, and lack support for task customizations, grading and quantitative measurements. In addition, Sandlund et al. [88] state that patient interest in gaming slightly fades over time, indicating that there is a need for flexible games that adapt to the changing ability of the patient and offer a continuous challenge to maintain interest. To solve these problems, researchers have designed their own serious games for different therapies [51, 84, 66, 17, 28, 1].

2.3 Serious Games for Motivational Balance Rehabilitation of Cerebral Palsy Patients

This section explains how we transferred a conventional therapy to a serious game to improve the balance and postural control of adults with CP in a motivational way. First, we present the balance rehabilitation therapy objectives and how the objectives were transferred to the serious game, along with details of the implementation and tests.

2.3.1 Balance rehabilitation therapy

The standing human posture is innately unstable. The Center of Mass (COM) of the body is located at approximately five-ninths of the body height from the ground, over a narrow Base of Support (BOS). The active control of body alignment is the skill of maintaining balance inside the BOS. Activities of Daily Living (ADL) require control of the position of the COM over the BOS. A lack of or inappropriate skills for this control results in the risk of falls and/or diminished quality of life.

Exercises for this active control include functional strengthening, balance activity to improve tone and spatial orientation in postural control. The therapy also aims to strengthen the muscles of the neck, back and upper limbs which are used in balance and to coordinate the upper limbs with the visual environment. In a standard therapy session, the following exercises are performed [57]:

- To start the therapy, the patient is asked to stand up (with his/her own technical aids if needed) from the sitting position.

- Standing up, the patient performs COM movements for 5 minutes, using large and coordinated movements to displace the COM with speed, safety and balance.

- In that position, the patient makes 20 to 30 repetitions of forward-reaching exercises toward an object.

- In that position, the patient makes 20 to 30 repetitions of left side-reaching exercises toward an object.

- The same reaching exercises are performed to the right side.

The physiotherapist can adjust the amount of exercise according to the subject's level of physical function to provide the optimal amount of exercise for each subject, according to the patients' weekly tolerance for management of fatigue. Each participant has his/her exercise program updated or revised every week to improve his/her abilities in range of balance, motion, speed and functional strengthening for postural control.

2.3.2 System design

We selected the therapy defined in section 2.3.1 to be transferred to a serious game to improve balance, increase motivation in patients and achieve higher adherence to this long-term therapy. The users must interact with objects that cannot be reached without moving the COM beyond the BOS. More specifically, users must remove individual items that appear on the screen by reaching each item with one hand.

The serious game is designed using the prototype development paradigm, following requirements indicated by physiotherapists and considering the desirable features for rehabilitation with serious games [16], as follows:

- **Meaningful play:** the relationship between player's interactions and system responses.
- **Challenge:** maintaining an optimal difficulty, which is important to engage the player.

Moreover, the serious game allows for the inclusion of motivational elements to increase playing engagement. Monitoring mechanisms to simplify the therapist's work are also included.

We propose a system configuration in which users stand in front of the screen and interact with the video game using their movements (see Figure 2.1). In addition, because users may have difficulty in holding devices, the designed game is markerless and device free. With this configuration, users can see the serious game while interacting with the game.

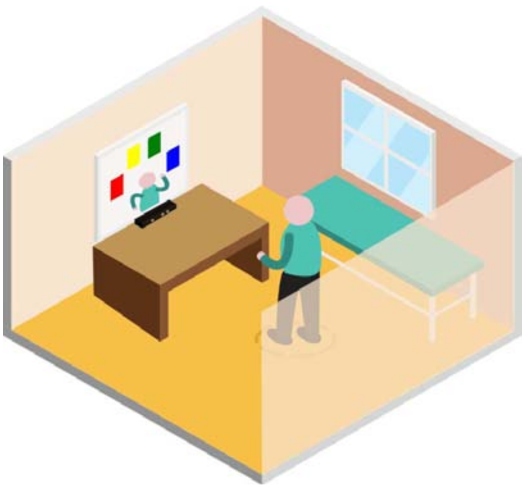


Fig. 2.1: System environment configuration. Important elements are: the user, the capturing device in front of the user, the screen and the scene background.

Motivation elements

The objects with which the participant interacts can be easily changed to show images that increase motivation by displaying themes of particular interest to each user (see Figure 2.2).

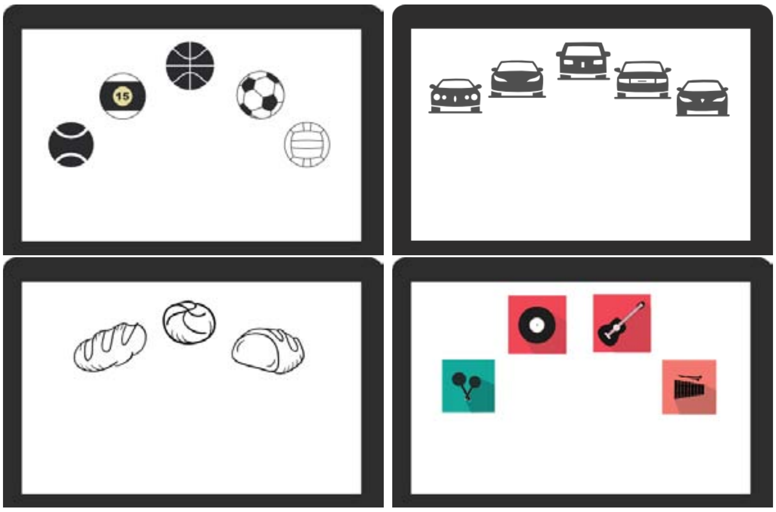


Fig. 2.2: Interaction screens with themes for motivating users and various templates.

Feedback

The serious game presents two types of feedback: visual and auditory. The visual feedback allows users to see themselves on-screen at all times, so the player's position relative to interaction objects is always known. Moreover, a pointer is projected on the user's hand, and the part of the interaction object that intersects with the pointer is erased (see Figure 2.3). When the interaction object is completely deleted from the screen, auditory feedback is played. Finally, when the game ends, the user receives different types of visual and auditory feedback, depending on the end game conditions. With these feedback mechanisms, we ensure the *meaningful play* [16] feature for rehabilitation with serious games.

Adaptability

To make rehabilitation sessions adaptable to different users, a set of templates were created that define the size and position of the interaction objects (see Figure 2.2). This way, we can define different levels in the game, depending on the skills and progression challenges for each user. In addition to the pattern templates, we defined configuration parameters to customize the

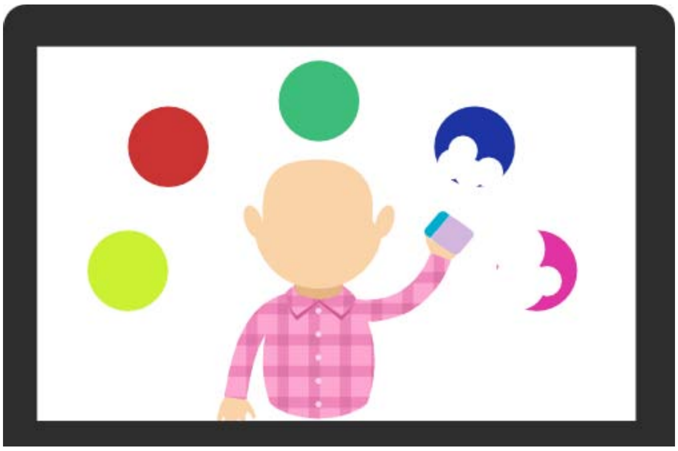


Fig. 2.3: Visual feedback example.

game and adapt the game to different users. The tunable conditions are the following:

- **Maximum playing time**: The therapist can set a time limit for each session.
- **Inverse effect**: To increase the game difficulty, the game screen can be reversed. Thus, when users move their right hand, users see their left hand moving.
- **Contact time**: The therapist can customize how long a player must be in contact with an element to erase that object.
- **User distance**: The distance between the user and screen in meters can be modified. The larger the distance from the screen, the larger the COM change needed.

With these adaptation mechanisms we ensure the *challenge* [16] feature for rehabilitation with serious games

Monitoring

To simplify the monitoring of patients by therapists, the rehabilitation system saves and maintains an *xml* file for each patient. The file is easy to parse and analyse, and configuration parameters and a dataset for each session are stored. The data consist of: the date of the session, level pattern, playing time, removed percentage, patient distance from the screen and contact time. This way, the monitoring of patients by therapists is simplified. This design has the potential to be used in other rehabilitation systems. The system is flexible and can accommodate two types of users: the patient and the therapist, who have different interaction objectives with the system. Figure 2.4 describes the feedback the system gives to each user, matching subject needs.

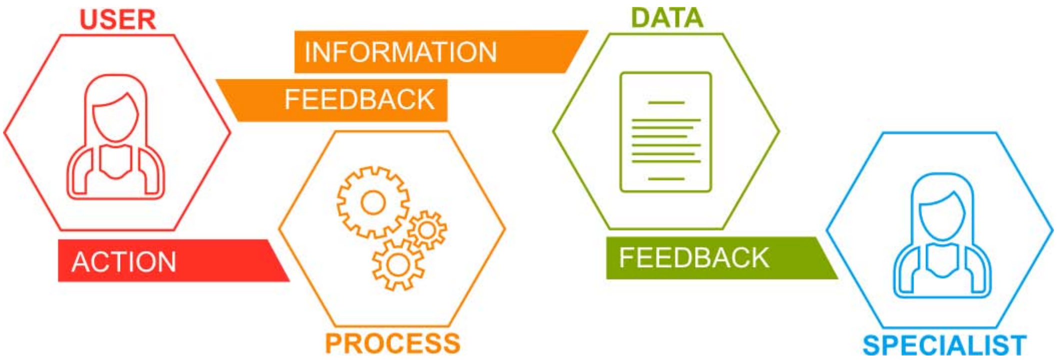


Fig. 2.4: Monitoring scheme.

2.3.3 System development

In general, VBI systems aim to provide reliable computer methods to detect and analyse human movements. The process is repeated over time, enabling monitoring of the user's actions. Depending on the computer vision technique chosen (e.g., interaction with the silhouette, arms or hands), different levels of precision are achieved for the user interaction. In addition, due to the dependence on real-world conditions (e.g., lighting, distances, and clothes), the interaction environment limits the techniques that can be used. Figure 2.5 depicts one example of interaction consisting of 'touching' a virtual object, which can be implemented by detecting the user's silhouette, skin color or hand motion.

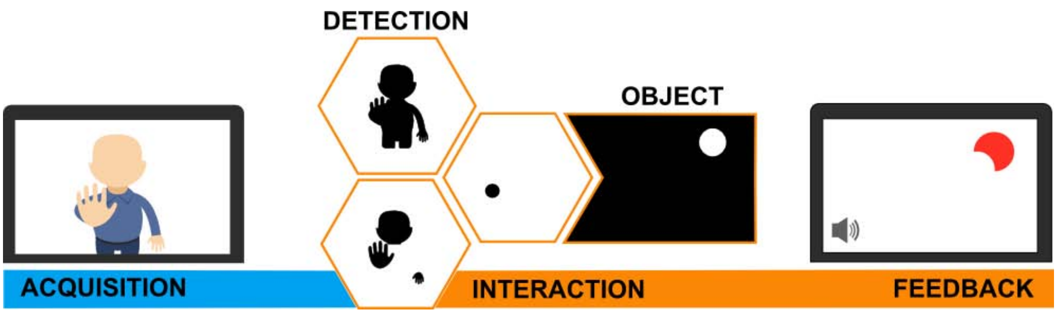


Fig. 2.5: Scheme of touching virtual objects using various computer vision techniques.

Background replacement

The user can be easily distracted by the scene background as there can be moving elements as the physiotherapists, or moving objects through a window. Another source of problems are background objects that can be mislead with the interaction objects due to similar colors or shapes. In order to avoid these problems a background subtraction process is applied and the background is substituted with a predefined image.

Tracking algorithm

In this experimental system, we specifically use the described tracking for the $Microsoft\, Kinect^{TM}$ SDK one, in order to avoid the user to hold a device. As the tracking initialization of the $Microsoft\, Kinect^{TM}$ algorithm was a wave gesture and the patients cannot perform it, we changed and the tracking starts when the user hand was in a certain screen position. Real world hand positions are mapped to the system coordinates due the application of a SDK function that transforms 3D world measures to pixel measures (height, width), maintaining the depth value in world coordinates. As the interaction is performed in the 2D space we obviate this depth coordinate.

Interaction mechanism

By considering interaction as the users ability to perform actions that are recognized by the system and to make interaction independent of the computer vision technique, 'contact' with the image interaction objects defines the interaction model of our system. The implementation is based on the *mask* concept. The designer creates a mask, with the same screen resolution, that contains the interaction regions marked in white and the unimportant regions marked in black. This way, for a game design, each object with the potential for interaction is represented by white regions in the mask (see Figure 2.6). Once the appropriate mask for each interaction screen is defined, the computer vision techniques are applied to determine whether there is any interaction. The method is easy to understand if it is assumed that the results of computer vision techniques can be represented by an image that is operated with a mask image to return the interaction event.

Resources

All tests were performed using a PC with this configuration:

- Intel Core2 Duo CPU P8400 @2.26 GHz

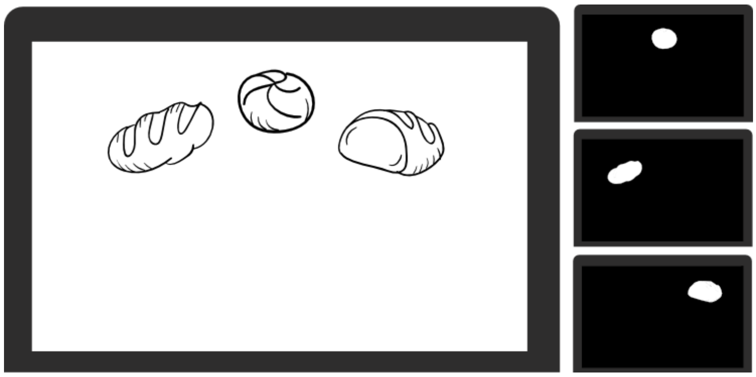


Fig. 2.6: Mask definition for objects with potential for interaction. The system only detects movements within the mask.

- 3034 MB RAM
- Graphic card Mesa DRI Mobile Intel GM45 Express
- Ubuntu 9.10
- $Microsoft\, Kinect^{TM}$

The serious game was developed using C++ programming language, OpenCV and OpenNI as a computer vision library and Qt as the graphical user interface library. The capture process, image processing and image visualization are performed by means of an OpenCV library. The system performance is 30 fps. This result ensures a real-time response from the system.

In $Microsoft\, Kinect^{TM}$ the field of view is $63°$ horizontal and $50°$ vertical [92]. Based on trigonometric properties, the maximum captured magnitude of the system with the user distance from the camera d is as follows:

$$m_{vertical} = 2.3835d$$

$$m_{horizontal} = 3.92522d$$

2.3.4 Preliminary validation

Alternative system testing was undertaken before the start of the clinical study. This testing lasted 6 months, at which point the therapists concluded that the system implemented the balance rehabilitation therapy defined at section 2.3.1. To check whether images of particular interest to each user motivated the users, we performed an experiment with 10 patients with CP during the last 12 weeks of preliminary testing:

- The patients played the serious game for 15 minutes each week.
- The patients played a sequence of one week with themes of particular interest and the following week without, until completing 12 weeks.
- Each week, the system stored the time taken to perform the session
- Each week, the configuration of the system was the same except for the motivational elements.

When the interaction objects were related to patient interests, the patients performed the rehabilitation activity in 13.5% (standard deviation of 4.3%) less time than when the objects did not represent such interests. The results are expressed in percentages because the time required to perform the task is dependent on the specific abilities of each patient.

System motions

Once the experiment finished, the therapist concluded that user the designed system the user can perform the following actions to interact with the system:

- Flexion, extension, lateral flexion (both sides) and rotation of the neck.
- Flexion, extension, lateral flexion (both sides) and rotation of the trunk.
- Flexion, extension, adduction, abduction, internal and external rotation and circumduction of the shoulder.
- Flexion, extension of the elbow.
- Pronation and supination of the elbow-wrist complex.
- Flexion, extension, adduction, abduction and circumduction of the wrist.

The system allows the therapist to define the dimension of the interaction objects to adapt the serious game to the user's abilities. Also, the therapist can decide which hand the patient should use for the therapy. Another variable used to adapt the serious game to the user's abilities is the distance between the user and the camera. The further that users stand from the camera, the more that users have to change their COM to reach the interaction elements located on the screen perimeter (see Figure 2.7).

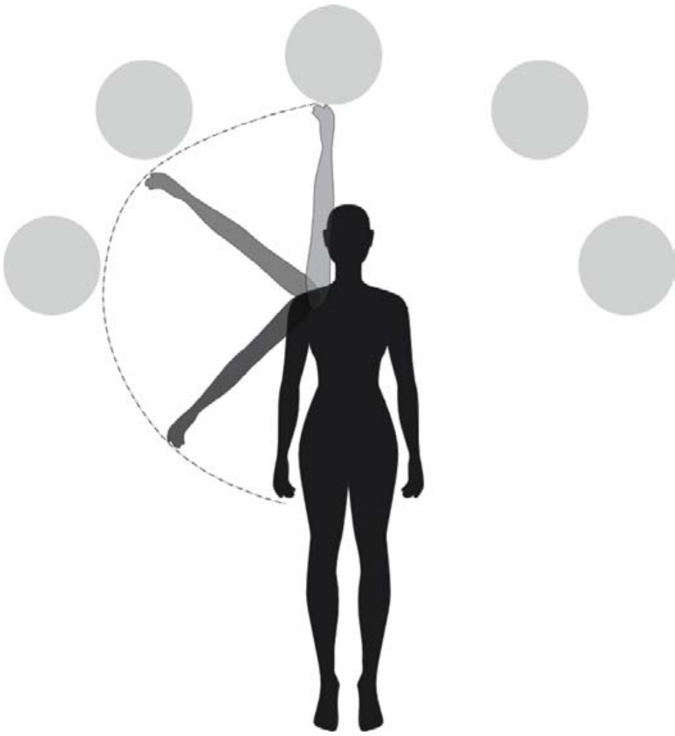


Fig. 2.7: User reach (arm span) without changing center of mass, in grey.

We experimented if serious games for rehabilitation can be used for motivational balance rehabilitation in cerebral palsy patients. The presented video game tries to promote a specific body movement in order to change the users' gravity center. Results show that users improved their balance slowly; improvements were also detected in individual items. With regards to motivation, in previous years the set of users had abandoned their therapeutic plans. Using the presented video game, no users have abandoned and they showed interested in continuing the rehabilitation process with the video games.

2.4 Mirror feedback validation

During the development of the interaction system described in previous section we noticed that users had problems to perform the exercise because they could not match their movements with what they saw on the screen. We added the image of the user into the game, so they could see themselves into the game in order to simplify the interaction. We called that feedback, *mirror feedback*. In order to explore the importance of mirror feedback in vision-based motor rehabilitation interactive systems, we conducted a user study testing. Specifically, the users tested their own visual representation such as interaction feedback of the video game for rehabilitation.

Different articles reviewed the importance of feedback in motor learning and rehabilitation. To our knowledge, it did not exist any study about the importance of the mirror feedback in vision-based rehabilitation systems. In fact, on one hand [106] indicated that feedback might enhance motor leaning but there were many areas as yet not examined as the case of mirror feedback. On the other hand, in [91] reviewed different types of natural visual feedback such as superposition, side-by-side 3D perspective, end-effector movements and third- and first-person perspective. However, they did not reference any work about mirrors as natural visual feedback. For this reason, with the aim of seeking deeply articles related to mirror feedback, we also searched at Google Scholar, Web of Science, IEEE Explorer and ACM Digital Library different combinations of the following key words: *feedback, motor learning, augmented feedback, extrinsic feedback, rehabilitation and mirror*. We neither found any article related to mirrors as natural visual feedback.

2.4.1 Participants

Adults diagnosed with CP and with limited voluntary motor control of one or both arms and legs and of the trunk were recruited from ASPACE. These subjects had mild to moderate cognitive impairment, as shown in Table 2.1. We used the Mini-mental state examination (MMSE) to classify their cognitive impairment because it is a brief and objective screening test, and also because it is valid and reliable across a variety of clinical, epidemiological, and community survey studies [102]. The inclusion criteria were as follows:

- Aged 20 to 65 years.
- No participation in clinical study published in [52].
- Ability to walk with or without technical aids (GMFCS I and II) [1].
- Ability to understand, learn and follow simple instructions.
- Voluntary agreement to participate in the clinical study.

The exclusion criteria were as follows:

[1] Gross Motor Function Classification System

- Severe cognitive impairment.
- Profound bilateral hearing loss with the use of hearing aids.
- Severe visual impairment.
- Serious or uncontrolled epilepsy.
- Serious or recurring medical complications.

The research team made a request to all adults in ASPACE. The final study population included 8 adults (7 male), aged 22 to 41 (Mean (M) = 33), with CP. Their families signed an informed consent, as legal proxies. Characteristics of the participants are presented in Table 2.1.

We also included a control group composed of 32 non-paid volunteers (14 female) aged 19 to 25 (M = 20.4), with no disabilities.

User	Age	Physical diagnosis	MMSE
1	22	cerebral palsy	moderate
		spastic tetraparesis	cognitive impairment
2	27	cerebral palsy	mild
		spastic tetraparesis	cognitive impairment
3	32	cerebral palsy	moderate
		spastic tetraparesis	cognitive impairment
4	32	cerebral palsy	mild
		mixed spastic tetraparesis	cognitive impairment
5	34	cerebral palsy	mild
		spastic tetraparesis	cognitive impairment
6	37	head trauma	mild
		spastic tetraparesis	cognitive impairment
7	39	cerebral palsy	moderate
		mixed spastic tetraparesis	cognitive impairment
8	41	cerebral palsy	mild
		ataxic tetraparesis	cognitive impairment

Tab. 2.1: Characteristics of participants.

2.4.2 Procedure

We were interested in the different users performances in the game interaction enabling or not the explained mirror feedback mechanism, i.e., the no-mirror feedback condition was characterized by the absence of such visual feedback. Other feedback mechanisms defined in Section 2.3.2 were activated.

We used a within-subjects design with the two previously defined feedback conditions:

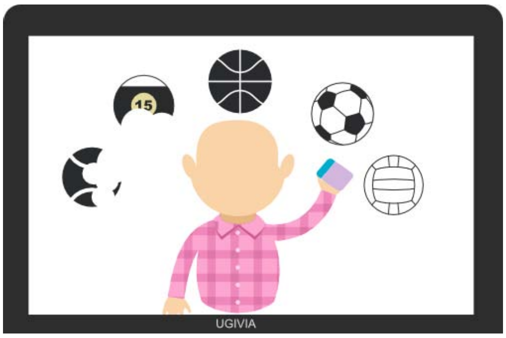
MF : Mirror feedback (including the own visual representation).

NM : No-mirror feedback (absence of such visual feedback).

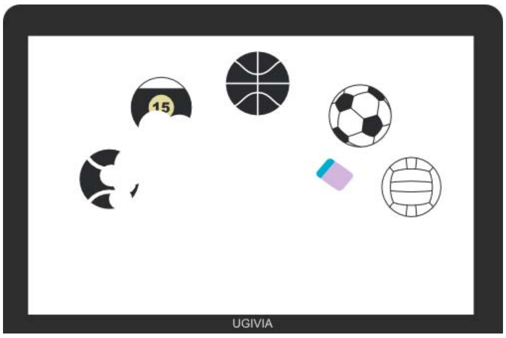
In Figure 2.8 it is possible to observe the feedback for the two interaction conditions. The user study was divided into two experiments with two different groups of users:

C : Control group (users without disabilities).

D : Users with disabilities.



(a) Mirror feedback.



(b) No-mirror feedback.

Fig. 2.8: Experimental feedback conditions.

For the control group, the user study started with a brief introduction and a demonstration, together with a demographic questionnaire asking about age and previous use of vision-based interaction applications. Participants played two sessions of the designed computer game with the same conditions that the user with disabilities, i.e., only moving the upper body part to delete the virtual objects with their hands. For each session the order of conditions (mirror feedback vs. no-mirror feedback) were randomly selected so as to balance both interaction conditions across participants.

For users with disabilities, the game was previously tested on a pilot scheme for a two month period, attending the rehabilitation center once a week. They practiced with the game for at least 20 minutes only with no-mirror feedback condition, and the number of repetitions varied according to participants' tolerance and the physiotherapist's prescription to better manage fatigue. These two months of training were important to ensure a correct understanding of the game and to learn how to carry it out, as well as ensuring a correct parameter adaptation to each user. Once the users correctly understood the game play, participants played two sessions of the designed computer applying the same procedure as for the control group. Figure 2.12 shows real performance of the system in ASPACE rehabilitation room using mirror feedback.

2.4.3 Measurements

Quantitative measures included logged *time-to-start* (T_s) and *time-to-complete* (T_c) times. The *time-to-start* measured the time the user interacted with the first virtual object. We interpreted this time as the time taken by the users to orientate their motions with the game interactions. This measure was derived from the observations in the pre-test sessions performed with the pilot. In these sessions, users with disabilities had greater difficulty in attaining orientation, and they had trouble knowing their position during play, relative to the interaction objects. This fact was more clearly observable when they had to delete the first virtual object.

The *time-to-complete* measured the time that users needed to complete the deletion of all virtual objects. In the experiment with the group of users with disabilities, the virtual objects were properly located in order to ensure that all the performances achieved the complete deletion goal. Furthermore, in a final questionnaire, the participants selected their preferred interaction feedback for playing the game.

time-to-start is related to effectiveness and *time-to-complete* is related to efficiency of interaction task. According to usability definition [10], it has three aspects: satisfaction, effectiveness and efficiency. Satisfaction's measures include users' preferences: we had demonstrated experimental system improved the balance and postural control [52], that is the user's objective and the final questionnaire indicated they preferred interaction feedback for playing the game. Effectiveness's measures include quality of solution: *time-to-start* implies first interaction, and users are not able to complete the task if they do not understand the game mechanics and, therefore, start to play. Then, it also has a direct correlation with the task completion and the quality of solution. Efficiency's measures include use of time: tasks completion time (*time-to-complete*).

2.4.4 Results

Table 2.2 shows the measured *Time-to-start* and *Time-to-complete* for users with disabilities using the feedback conditions defined by the experiment. Mirror feedback had better results on the measured times for users with disabilities. Users with moderate cognitive impairment had bigger differences between feedback conditions (mirror vs. no-mirror) than users with mild cognitive impairment.

User	T_s	T_s	T_c	T_c
	MF	NM	MF	NM
1	5	23	150	245
2	2	11	129	160
3	5	26	132	226
4	2	10	126	154
5	2	10	160	174
6	3	15	132	176
7	4	19	148	218
8	2	11	121	148

Tab. 2.2: Measured *Time-to-start* (T_s) and *time-to-complete* (T_c) for users with disabilities. Mirror feedback (MF) and No-mirror feedback (NM)

Figure 2.9 and 2.10 show mean completed times for both measures, with error bars indicating 95% Confidence intervals (CI). They will be further discussed in the following section, based on the interaction feedback. Table 2.3 summarize the influence of mirror feedback in the mean measured times for T_s and T_t .

	Time-to-start, T_s	*Time-to-complete*, T_c
Control group (**C**)	$t(31) = -1.74, p = 0.09$	$t(31) = -0.87, p = 0.389$
Users with disabilities (**D**)	$t(7) = 7.09, \mathbf{p < 0.001}$	$t(7) = 4.48, \mathbf{p < 0.005}$

Tab. 2.3: Overview of the influence of mirror feedback as interaction feedback for each user group on the mean of the defined time measures. Significant results are printed in bold.

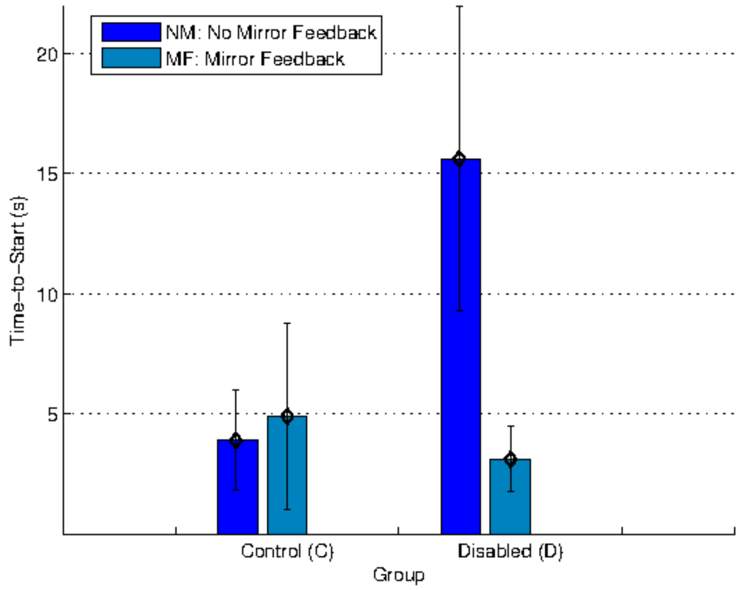


Fig. 2.9: Overview of mean times for the *time-to-start* measure (T_s).

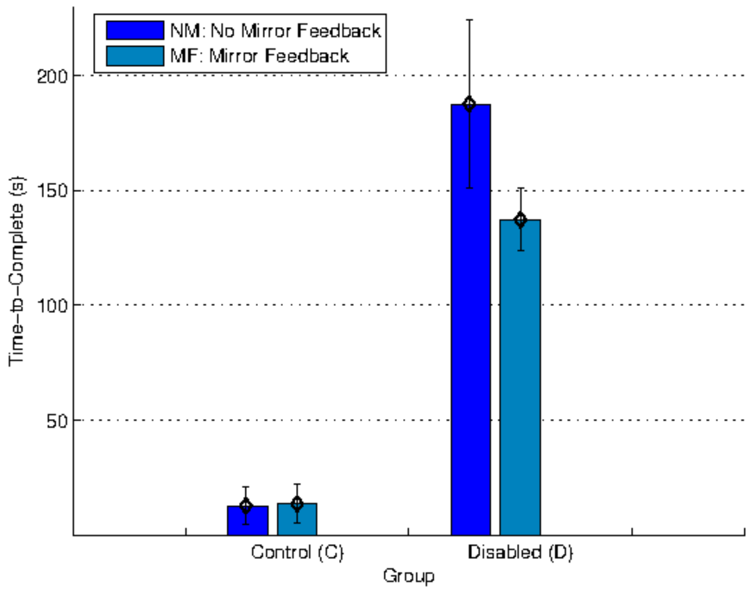


Fig. 2.10: Overview of mean times for the *time-to-complete* measure (T_c).

A two-tailed paired samples t-test was conducted to evaluate the impact of mirror feedback for each measure in both user groups. Mirror feedback had a highly significant impact on the measured times for users with disabilities (cf. Table 2.3). The measured times were significantly worse in the absence of the own visual feedback. The control group, on the other hand, completed the experiment with both feedback with no significant performance differences. However, in the final questionnaire results about interaction feedback preferences, 24 participants of the control group preferred the mirror feedback. For the group disabilities, seven participants answered that by including mirror feedback mechanisms they gained more control.

Before using the paired t-test, we applied a Kolgomorov-Smirnov test of normality ($D = 0.2587, p = 0.1187$) and the Wilcoxon signed rank test with continuity correction obtaining a p-value less than 0.01 ($p = 0.007015$).

2.4.5 Discussion

Results confirmed our hypothesis that in case of disabilities the mirror feedback mechanisms facilitated the interaction in vision-based systems for rehabilitation. They demonstrated that the implementation of mirror feedback by giving patients the possibility of seeing themselves on screen, means that they were conscious at all times of the actions performed relative to the video game. In the user study presented, we proved this claim by means of experiments, showing a significant improvement of users with disabilities results in the game play. We also observed that users with moderate cognitive impairment had bigger differences between feedback conditions than users with mild cognitive impairment (see Tables 2.1 and 2.2).

Finally, we want to discuss the relationship between the mirror feedback and the *game feel* definition used for game design [98]. Game feel is the sensation of the system's response to the player: the kinesthesic qualities of the experience created by coupling with the input device, and seeing what happens in the game as a result. In our experience, the input device is the user's own body, therefore, it makes sense that feedback should be related to it also. In this sense, it is interesting to point out that the results obtained for the control group users (users without disabilities) could be interpreted to mean the mirror feedback mechanism is not significant. However, it should be taken into account that the video game was specifically designed for users with disabilities and its game play was too easy for users without disabilities. It may be interesting to perform another user study with more complex games based on vision-based interaction, in order to properly explore if the introduction of mirror feedback can improve the user experience of vision-based interaction.

A limitation of our study was the sample size of users with disabilities (8 subjects). However, a post-hoc power analysis indicated that with 8 subjects there would be a 90% chance (for $\alpha = 0.05$) that the statistics would have detected a difference greater than 6.6 points in the *time-to-start* measure (we obtained a mean difference of 12.5), and greater that 50 points in the *time-to-complete* measure (we obtained a mean difference of 50.38).

We observed that users with cognitive impairment had bigger differences between feedback conditions. The higher cognitive impairment user had, the more important feedback was in order to perform correctly the therapy. Results confirmed our hypothesis that in case of disabilities the mirror feedback mechanisms facilitated the interaction in the vision-based systems for rehabilitation. Once demonstrated the positive effects of mirror feedback mechanism in rehabilitation system, it may potentially be extended to individuals with disabilities, in order to help achieving better functional abilities.

2.5 Clinical validation

There is evidence that balance training improves postural control. But, analysing the state-of-art there is a lack of studies that use balance measures, whether static or dynamic, after a serious game intervention therapy program. In this section a clinical study is presented using the vision-based interaction system in Section 2.3. We describe the participants, procedure, measurements and results after a 24-week therapy period using the system.

2.5.1 Participants

Adults diagnosed with CP and with limited voluntary motor control of one or both arms and legs and of the trunk were recruited from ASPACE. The inclusion criteria were as follows:

- Aged 20 to 65 years.
- No adherence to physiotherapy treatment after attending the center's program as a long-term therapy.

Participant	Age	Diagnosis
1	34	CP with spastic tetraparesis and hydrocephalus
2	36	CP with spastic tetraplegia
3	27	CP with spastic tetraparesis
4	57	Mixed type CP with spastic tetraparesis and athetosis
5	43	CP with moderate to severe psychomotor retardation
6	37	CP with spastic tetraparesis
7	32	CP with ataxic tetraparesis
8	32	CP with spastic tetraparesis
9	33	CP with spastic tetraplegia

Tab. 2.4: Characteristics of the participants

- Ability to walk with or without technical aids (GMFCS I and II) [2]
- Ability to understand, learn and follow simple instructions.
- Voluntary agreement to participate in the clinical study.

The exclusion criteria were as follows:

- Severe cognitive deterioration.
- Profound bilateral hearing loss with the use of hearing aids.
- Severe visual impairment.
- Serious or uncontrolled epilepsy.
- Serious or recurring medical complications.

The research team made a request to all adults in the ASPACE who met the inclusion criteria, and 90% agreed to participate in the study. The final study population included nine adults (seven male) with a mean age of 37 years and 4 months (range 27 to 57 years). The characteristics of the participants are presented in Table 2.4. The participants' families signed an informed consent form, as legal proxies.

[2] Gross Motor Function Classification System

2.5.2 Procedure

The Physiotherapy Department was provided with the system described in section 2.3.3.

One of the inclusion criteria was not adhering to physiotherapy treatment after attending the conventional program of the center as a long-term therapy. Therefore, the patients in this study were undergoing rehabilitation only within our experimental system. This rehabilitation program consisted of one session per week for 24 weeks. The patients practiced with the game for at least 20 minutes, and the number of repetitions varied according to participants' tolerance and the physiotherapist's prescription to better manage fatigue. The participants had their exercise program updated or revised each week to improve their abilities in range of motion, speed and functional strengthening for postural control.

Before and after the 24-week therapy period, the users were pre- and post-assessed, as described in Figure 2.11. To obtain the information needed to perform this study and monitor the users, clinical records were obtained from the ASPACE and further records were obtained from the interdisciplinary team: a physiotherapist and an occupational therapist and/or a rehabilitation physician.

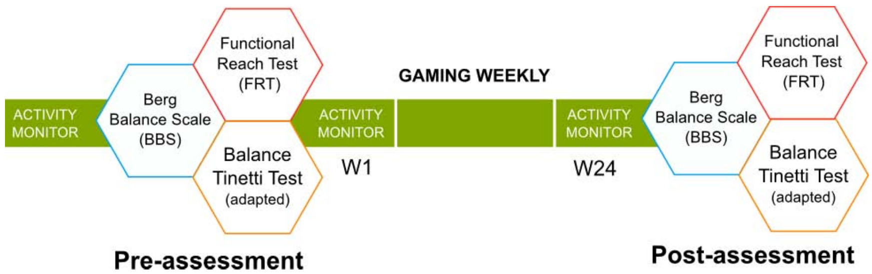


Fig. 2.11: Scheme of entire clinical intervention.

2.5.3 Measurements

Gaming was assisted by a physiotherapist and monitored by the research team. In a case report form, the team recorded how participants felt about the game that day, whether the participants were tired, which games were played and whether the participants' did not want to play on that particular day and why. It was also noted if participants needed verbal commands to adjust their postural control and when these commands were required. Additionally, fatigue after therapy was assessed on an analogue 0 to 10 scale, and whether fatigue appeared immediately after therapy sessions or during sessions was noted.

Fatigue is very changeable from day to day in people affected with CP. Because the system registered difficulty levels according to the parameters cited previously, the research team's first task each day was to adjust the parameters to the patient's fatigue level. We wanted to increase the difficulty levels of the tasks, but constant changes in the patients' fatigue made it impossible to measure the challenge to balance.

As stated before, all subjects were clinically evaluated prior to the intervention program and again at the conclusion to assess the hypothesis/objectives.

The parameters used to functionally assess balance were the following:

- Berg Balance Scale (BBS)
- Functional Reach Test (FRT) (FRT, included in the BBS)
- Balance Tinetti Test (adapted)

The BBS was chosen because this scale evaluates the static and dynamic balance of a patient. It comprises a set of 14 simple balance related tasks, ranging from standing up to a sitting position. The degree of success in achieving each task is given a score of zero (unable) to four (independent), and the final measure is the sum of all of the scores. We consider this scale to be important because these tasks are based on the patients' abilities to perform the tasks and maintain an adequate position to do so. The maximum score is 56 points, and when the value is under 46, the score predicts that the patient will suffer multiple falls. The scale has been used to quantify balance function in neurological disorders. The scale results are interpreted as follows: from 0 to 20, a wheelchair is required; from 21 to 40, assistance is needed for walking; and from 41 to 56, the individual is independent while walking. It takes approximately 15 minutes to perform this test, which is considered to be a sensitive measure because the test differentiates between patients with various technical aids for mobility [8].

The FRT was selected because this test notes limitations in ADL and indicates the risk of falling. This test has been validated as a predictive measure of repetitive falls [33]. To perform the test, the physiotherapist marked the shoulder height of each patient on a large sheet attached to a wall. In a standing position comfortable for the patients, the patients were asked to lift an arm to 90 degrees and to then stretch out the fingers and reach forward as far as they could, bending at the trunk. Functional reach was the distance between the initial and final positions. Reaching less than 10 inches (25 cm) indicates a limitation in ADL and a risk of falls.

The Balance Tinetti Test was selected because this evaluates static and dynamic balance using several items. However, the Balance Tinetti Test had to be adapted to our patients because certain items were not applicable to our study [101]. Scoring of the Tinetti Assessment Tool is performed on a three-point ordinal scale, with a range of 0 to 2. A score of 0 represents the most impaired, whereas a 2 represents the independence. The individual scores are then combined to form three measures: an overall gait assessment score Tinetti Gait Section (TGS) an overall balance assessment score Tinetti Balance Section (TBS) and a gait and balance score (). The maximum score for the TBS is 16 points. The maximum score for the TGS is 12 points. The maximum score for the Tinetti Total Score (TTS) is 28 points (TBS + TGS). In general, patients who the TTS score below 19 are at a high risk of falls. A TTS score in the range of 19 to 24 indicates that the patient has a risk of falls [74]. To functionally evaluate gait, the TGS was used.

2.5.4 Results

All participating subjects were able to practice with the motion interactive game and needed help to set up the game (see Figure 2.12). One subject was excluded during the study due to tendinitis in one knee.

Fig. 2.12: Actual performance of the system in the ASPACE rehabilitation room.

		pre-assessment (mean	SD)	post-assessment (mean	SD)	P-value
BBS		29.5	3.9	34.1	2.2	0.002
FRT						
	FRT for the Right arm	8.6	1.4	10.1	2.0	0.007
	FRT for the Left arm	8.3	2.0	10.1	3.7	0.052
TTS		16.0	4.0	21.0	2.8	0.010
	TBS	9.6	2.8	12.5	1.9	0.007
	TGS	6.3	1.5	8.5	2.1	0.061

Tab. 2.5: Results of the pre and post 24-week intervention. SD, standard deviation; BBS, Berg Balance Scale; FRT, Functional Reach Test; TTS, Tinetti Total Score; TBS, Tinetti Balance Section; TGS, Tinetti Gait Section.

All statistical analyses were performed using SPSS version 19.0 (SPSS Inc., IBM, Chicago, IL 60606, USA) and the means of the measurements before (pre) and after (post) the 24 weeks of intervention were compared using a paired-samples t test. The results are summarized in Table 2.5 (mean standard deviation).

Impact on balance

Eight subjects were included in the analysis. According to the BBS, when comparing the pre- ($29.5 \quad 3.9$) and post-assessment ($34.1 \quad 2.2$) results demonstrated a significant functional improvement ($p = 0.002$).

A comparison of functional balance revealed significant differences in the FRT before and after intervention:

- In the right upper limb (pre- ($8.6 \quad 1.4$) and post-assessment ($10.1 \quad 2.0$)).
- In the left upper limb (pre- ($8.3 \quad 2.0$) and post-assessment ($10.1 \quad 3.7$)).

Upon completion of the 24-week intervention program, the results for the TBS were significantly improved ($p < 0.007$). In the pre-assessment, the results were ($9.6 \quad 2.8$), and in the post-assessment, the results were ($12.5 \quad 1.9$). Seven subjects improved, and one kept the same score.

Most of the subjects did not reach 10 points in the TBS's pre-assessment (only two scored over 10). In the post-assessment, all participants reached or passed a score of 10, and one even reached the maximum of 16 points.

Impact on gait

Upon completion of the 24-week therapy program, a significant difference was noted for the pre- and post-assessment measures. Most of the subjects did not reach 9 points in the TGS's pre-assessment (only three scored 8). However, in the post-assessment, all patients improved, and two surpassed a score of 10. One participant even reached the maximum score, or 12 points. Post-hoc comparisons showed that balance changed, demonstrating a significant improvement on the items of *Step Symmetry* and *Path*.

A significant difference between the pre- ($16.0 \quad 4.0$) and post-assessment ($21.0 \quad 2.8$) measures ($p = 0.010$) in the TTS score resulted at the end of the 24-week intervention period. The degree to which the program influenced the outcome was due to the balance section.

2.5.5 Discussion

The present study was designed to prove the clinical validity of interactive system, providing objective measures of the improvement of balance function according to the types of functional exercises and by means of the presented experimental system. This system was implemented and investigated during a 24-week physiotherapy intervention program at ASPACE in the Balearic Islands. All participants significantly increased their balance and gait function from 16 to 21 points of a total of 28 possible points, according to the TTS for the risk of falls ($p = 0.01$).

Objective measurements of balance depend on several physiological systems. Balance allows us to undertake a wide range of ADL because it enables us to anticipate changes and coordinate our muscle activity to maintain stability.

Our results show that the experimental system presented improves static and dynamic postural balance and gait function. The improvement in dynamic postural balance and gait function is significant after a 24-week intervention.

Some items, involving the physical performance, did not change in the post-assessment. These tasks may have been too difficult for our population, and therefore, the experimental manipulation had little effect.

The degree to which the therapy system influenced the outcome of the FRT was moderate as an improvement because not reaching 10 inches (25 cm) indicates limitations in ADL and a risk of falls.

In the post-assessment of the TBS, all participants reached or passed a score of 10, and one even reached the maximum of 16 points. This increase is due to improvements on items 3, 6 and 7 on the TBS:

- Item 3: attempts to rise (0 = unable without help; 1 = able, but requires more than one attempt; 2 = able to rise with one attempt).

- Item 6: nudge, subject at maximum position, with feet as close together as possible, and examiner pushes lightly on subject's sternum with palm of hand three times (0 = begins to fall; 1 = staggers and grabs, but catches self; 2 = steady).

- Item 7: eyes closed, at maximum position in item 6 (0 = unsteady; 2 = steady)

One limitation of our study is the sample size (eight subjects), especially because we lacked a control group (subjects not under treatment). However, a post-hoc power analysis indicated that with eight subjects, there would be an 80% chance that the statistics would have detected a difference greater than 4.6 points in the TTS (for $\alpha = 0.05$). This difference is clinically significant in this study because the patients transition from a *High Fall Risk* (mean of the pre-test equal to 16) to a *Moderate Fall Risk* (mean of the post-test equal to 21). The TTS calculation was based on the assumption that the magnitude of the difference would be equal to the standard deviation of the difference between the pre- and post-intervention. Future studies with larger samples should be performed to gain a more complete picture of the improvements in balance and gait abilities observed in this study.

One of the advantages of the presented interactive system is that it is not necessary to take any measurements. With an adequate combination of the computer vision and the virtual view, the patient feels that his/her movements in the real world cause actions in the virtual view. The patients are able to understand how their movements perform game actions during the rehabilitation session, thus facilitating the game's correct performance to achieve the rehabilitation goals.

2.6 Design issues for vision-based motor rehabilitation serious games

Such a conclusion of our previous described experience of implementing vision-based motor-rehabilitation serious games [53, 72, 85], defining an interaction model adapted to the user's capabilities and following the desirables features for rehabilitation serious games presented, we can state that success of using this type of serious game depends on 7 design issues: the development paradigm, the interaction mechanism, the interactive elements, the feedback, the adaptability, the monitoring, and the clinical evaluation. These design issues are presented in more detail in the next section.

2.6.1 Development paradigm

After first meetings with physiotherapists, we discovered that engineering technical language is totally different to physiotherapy. To ensure objectives, we decided to develop the game using the prototype development paradigm [80], which facilitated communication between engineers and physiotherapists. This paradigm ensures all the necessary information to perform the different tasks is provided in a clear and understandable way.

2.6.2 Interaction mechanism

A serious game should not develop a new rehabilitation therapy. It is more suitable to transfer existing therapy to a serious game, where the selected rehabilitation therapy is the means of interaction with the serious game. As a major number of patients with motor disease cannot hold a device, a camera can be used as an input device in order to define the interaction model adapted to the user's capabilities [75]. Recent technological advances have created the possibility to enhance naturally and significantly the interface perception by means of visual inputs [105], the so-called VBI.

In general, vision-based interaction systems aim to provide reliable computer methods to detect and analyse human movements. The process is repeated over time, allowing for monitoring of the user's interacting actions. According to the computer vision technique used it is possible to achieve different levels of detail. In addition, due to the dependence on real conditions (lighting, distances, clothes...) the interaction environment limits the techniques that can be used. Figure 2.5 depicts one example of vision-based interaction which can be implemented by detecting the user's silhouette, the skin colour or the hand motion, for an implementation details see Section 2.3.3.

2.6.3 Interaction elements

Interaction objects must be selected in order to show the users images that achieve an optimal level of motivation, by choosing themes of particular interest to each user, see Section 2.3.2 for a detailed example. When the interaction objects were related to some of these interests [53], the patients performed the rehabilitation activity faster.

2.6.4 Feedback

The game must respond to the actions of the user through different types of feedback, in order to user be aware of their current state. In general, when using VBI with a system [16, 54], providing feedback is critical for users to feel in control and helping them to understand what is happening. Especially if there is not contact with the interface by means of an interaction device. A significant problem of vision-based interaction is that users have no interaction device of reference. The user, therefore, always should know when interaction is taking place using visual and audible feedback, see Section 2.3.2.

2.6.5 Adaptability

Rehabilitation sessions must adapt to the characteristics of the different users [16, 54, 84]. As the difficult level of therapy depends on the user interaction (motions), physiotherapists should create a set of templates which define the position where interaction elements must have, in order to define different levels in the game depending on the skills and the evolution of each user. In addition, the game should define different configurations parameters to customize games and adapt them to different users, see Figure 2.2.

2.6.6 Monitoring

The system must be able to archive different information about each user, configuration parameters and a dataset for each session consisting of patients performance, in order to simplify a patient's progression as monitored by the specialists [1, 54, 84]. Therefore, the system has two class of users, the patient and the specialist. Each are in pursuit of different objectives of the interaction with the system, see Figure 2.4.

2.6.7 Clinical evaluation

The clinical evaluation aims to quantify the improvement of the rehabilitation according to the kinds of functional exercises. In order to perform a successful clinical evaluation, it should define the experiment, the participants, and the measurements depending on the final goals and type of therapy. However, from our experience, we recommend to design the whole intervention through a pre-assessment and post-assessment of every measurement, see Figure 2.11. Optimally, the success of clinical evaluation increase when one can include a control group and the largest number of measurements. See Section 2.5 for a entire clinical evaluation description.

2.6.8 Concluding remarks

During the firsts iterations of the development of the experimental system, we specifically started using the algorithm *Continuously Adaptive Mean Shift* (camshift) [14], based on a mean-shift algorithm using a simple webcam. This iterative method allows the tracking of an object based on color as a fundamental property. To avoid difficulties with tracking in variable light conditions, a yellow glove was used to make the algorithm more robust and to avoid color confusion.

Then, we changed the described tracking for the $Microsoft\,Kinect^{TM}$ SDK one, in order to avoid the user to put a glove. As the tracking initialization of the $Microsoft\,Kinect^{TM}$ algorithm was a wave gesture and the patients cannot perform it, we changed and the tracking starts when the user hand is in a certain position of the screen. During all the experimentation period we used that tracking.

During the video game development we have noticed that environmental conditions have a big influence on the game. First, due to the dependence on real-world conditions (e.g., lighting, distances, and clothes), the interaction environment limit the techniques that can be used. Second, we noticed that user sometimes was distracted by what was happening around him. For that reasons, we developed a computer vision algorithm in order to segment the user from the scene background.

3 Modelling depth for nonparametric foreground segmentation

The problem of detecting changes in a scene and segmenting the foreground from background is still challenging, despite previous work. New RGBD capturing devices including depth cues could be incorporated to improve foreground segmentation. In this chapter, we present a new nonparametric approach where a uni ed model mixes the device multiple information cues. In order to unify all the device channel cues, a new probabilistic depth data model is also proposed where we show how handle the inaccurate data to improve foreground segmentation.

3.1 Introduction

Background subtraction is a widely used technique for detecting moving foreground objects in image sequences. It is considered the first step in many computer vision algorithms. In addition, foreground segmentation, provide an important cue for numerous applications in computer vision such as: surveillance, tracking, recognition, human poses estimation, among others. The main objective is to detect objects that do not belong to the scene by comparing the current observation with previous references. This reference can be a single image or a more complex model of the real scene, also called *scene model*. Typically, this model is a statistical representation of the scene and it is updated in order to be adapted to the variations of its conditions.

This problem has been widely addressed in the literature. Reviews can be found in [13, 36]. Despite this previous research, there is no universal technique covering all requirements of applications that need to detect the foreground of a scene [12]. In [103] several important challenges of background subtraction were described. Some of them are strongly related to the nature of color information, such as: shadows, changes in the scene illumination, camouflage or foreground aperture. These problems continue to be challenging to the modern approaches as described in [93], where 29 different algorithms were evaluated and compared. We can conclude that we need to continue working in this kind of algorithm in order to solve the previous described problems.

Our hypothesis is overcome these problems adding physical information to the background model. Different approaches to obtain 3D information of the scene were proposed using stereo devices or camera networks [29]. This depth measures provide geometrical information about the scene where each pixel value represents the distance from the device to the point in the real world. In order to obtain an accurate dense map of correlations between two stereo images, time-consuming stereo algorithms are required. Without specialized hardware, most of these algorithms are too slow for real time background subtraction. In addition, multi-camera networks introduce other problems such as: camera installation, calibration and data fusion.

Nowadays, low-cost RGBD devices that are able to capture depth and color images simultaneously at frame rate up to 30 fps are available off the shelf. These devices have certain limitations such as lower sensitivity at high distances, producing depth camouflage or absent observations due to the scene characteristics.

We present a new per-pixel scene modelling approach using both depth and color information, using this kind of noisy depth information in an unified model mixing all the device multiple information cues. We propose a model that keeps a sample for each pixel of the scene and estimates the probability that a newly observed pixel value belongs to the background. The model:

- Estimates the probability of pixel being foreground independently for each new frame.
- Is updated each iteration of the algorithm depending on partial results.
- Adapts itself to changes in the background process and detect targets with high sensitivity.

There are multiple proposals for scene modelling. In particular, KDE has been already used in *state-of-the-art* techniques [37] only with color information with good results [48]. Although, KDE allows to add new channels to the model without increasing computational complexity. By these two previous described reasons, we decided to construct our model using a KDE process. When using a Gaussian Kernel, the probability density function can be thought of as a generalization of the Gaussian Mixture Model (GMM), where each single sample is considered to be a Gaussian distribution by itself. However unlike GMM, in KDE no mixture parameters need to be estimated. This allows us to estimate the density function more accurately, without assumptions about the density model, depending only on recent information from the sequence.

Adding a depth channel to KDE background model is not an obvious process because the depth channel defers in its characteristics from color channels. In particular the depth channel has significant amount of missing information where the sensor is unable to estimate the depth at certain pixels. In this chapter, we show how to handle the inaccurate depth data in the proposed nonparametric scene model. With this purpose we properly define the absent depth observations in order to include them in the scene model. The key idea is that pixels that cannot be classified as background or foreground are classified in a new class, which we called *unde ned*. Therefore, absent observations could be handled in an unified manner. In addition, by introducing depth data, the proposed scene model is capable of instantly detecting the changes of background objects.

This chapter is organized as follows: first, we describe the related work of background subtraction algorithms using depth information. Second, we enumerate and describe the challenges of using depth data. Next, we describe the theoretical model we use to segment the foreground from background in image sequences. In Section 3.5 we describe the model update process. Finally, we present a concrete implementation of our background subtraction algorithm using a particular device.

3.2 Related Work

There is a large literature on the subject of background subtraction. We refer to some comprehensive surveys about this subject [13, 36, 29, 7, 93]. We focus here on approaches that fused color and depth information. Most of these techniques modify traditional background subtraction approaches by adding one extra channel to depth besides the color channels, and suggesting some heuristics to address the heterogeneous characteristics of these different cues, Table 3.1 shows an overview of the different background scene modelling techniques using depth and color information.

Gordon *et al.* [46] proposed an approximation to Gaussian Mixture Modelling (GMM) to describe the recent history of color and depth scene observations at each pixel. A multidimensional Gaussian mixture distribution is constructed, with three components in a luminance-normalized color space, and one depth channel. Special processing is made in order to deal with absent depth pixels, allowing foreground decisions when depth model for a pixel is invalid but its latest depth observation is valid and it is connected to regions where foreground decisions have been made in the presence of valid background data. No update phase is described, so this algorithm only can be used in static scenes. A similar approach is presented in [62] using ToF cameras.

In [95] a new mixture of Gaussians approach is proposed, where depth and infra-red data are combined to detect foreground objects. Two independent background models are built using depth and infra-red information, each pixel is classified by binary combinations of foreground masks. Performance of this approach is limited since a failure of one of the models affects the final pixel classification.

An adaptation of the Codebook [56] background subtraction algorithm was proposed by [40] fusing depth and color information to segment foreground regions. A four-channel codebook was used, depth information is also used to bias the distance in chromaticity associated to a pixel, according to the depth measurements. Therefore, when the depth value is invalid, the detection depends entirely on color information.

Clapés et al. [24], presented a background subtraction technique where a four dimensional Gaussian distribution was used as first step of user identification and object recognition surveillance system. No special processing was made in order to deal with absent depth observation pixels. As they used a single Gaussian approximation, the algorithm was not able to manage multi-modal backgrounds, this similar problem can be observed in other approaches like [46] and [58].

ViBe is a *per-pixel* algorithm, based on a Parzen windows-like process [5]. The update is done by a random process that substitutes old pixel values with new ones, sampling the spatial neighbourhood for refining the per-pixel estimation is done. ViBe has acceptable detection results in many scenarios, it is problematic with challenging scenarios such as darker background, shadows, and frequent background changes [12]. In [64], a new ViBe approach is presented using RGB and ToF (*Time-of-Flight*) cameras. Each model is processed independently and the foreground masks are combined with logical operations and then post-processed with morphological operators.

In [78] used separate GMM for color and depth information. Algorithm consider a foreground mask calculated using the GMM with RGB information given its great reliability in contours definition. Then a compensation factor (CF) is added, this is obtained by both color and depth information. Once CF has been evaluated, the final foreground mask is obtained after a logical OR and a final step for noise removal. Same authors present a similar approach in [79] changing the GMM for ViBE .

Song *et al.* [94] developed a method based on the GMM using gray scale and depth information to solve the limitations of color-based algorithms, especially in color camouflage situation. They built a probabilistic background model for combining color and depth information. In addition, they made a new dataset for evaluating the BGS algorithm using color and depth.

In [44], a foreground segmentation system that combines color and depth sensors information in a Bayesian Logarithmic Opinion Pool framework is presented. They presented a Spatial Color GMM and a Spatial Depth GMM to model the foreground, and two pixel-wise Gaussian models to model the color and the depth background domains. Those models are combined by using the Logarithmic Opinion Pool and the Hellinger distance in order to achieve a correct and robust classification of the pixels of the scene. The system is robust in front of color and depth camouflage problems, and also improves the segmentation in the area of the objects' contours. The quality of the foreground segmentation depends on its correct initialization and the correct modelling of all the regions of the object.

The method presented in [45] exploits new RGBD devices to make the background and foreground models more robust to effects such as camouflage and illumination changes. First, a preprocessing stage for aligning color and depth data and for filtering/filling noisy depth measurements is done. Then, they models the scene's background and foreground with a Kernel Density Estimation approach in a quantized x-y-hue-saturation-depth space. This approach is not suitable for real-time applications.

Camplani and Salgado [18] proposed a *per-pixel* background modelling approach that fuses different statistical classifiers based on depth and color data by means of a weighted average combination that takes into account the characteristics of depth and color data. A mixture of Gaussians distributions is used to model the background pixels and an uniform distribution is used for the modelling of the foreground. Same authors presented another approach in [19] based on the fusion of multiple region-based classifiers. Foreground objects are detected by combining a region-based foreground on depth data prediction with different background models, providing color and depth descriptions of the scene at pixel and region level. The information given by these modules is fused in a mixture of experts fashion to improve the foreground detection accuracy.

In the different approaches we described in this Section normally don't use depth and color information in an unified way, other solutions tends to use depth information as first decision filter before applying color information. Furthermore absent depth observations from sensor are ignored without any special treatment.

3.2.1 Datasets

Related to the different algorithms there are different datasets that are designed to verify the feasibility and the quality of the constructed algorithms and enables direct comparison between them.

In [40] they provided a dataset, called CITIC RGB-D Dataset, for evaluation of background subtraction techniques using the Microsoft Kinect sensor. They presented four sequences, including both RGB and depth images. Some frames have been hand-segmented to provide ground truth information. For each frame in the dataset, depth information was normalised from 0 to 255, where 255 is the maximum depth value in that frame, with the resulting loss of information.

In [43] they recorded a nine single person sequences with a Microsoft Kinect device. The aim of this dataset is to show depth and color camouflage situations that are prone to errors in color-depth scenarios.

In [96] they present a RGB-D Rigid Multi-Body Dataset that consists of 3 RGB-D videos of objects with different sizes. The sequences have been recorded using an Asus Xtion Pro Live camera. Ground truth for the camera pose has been obtained with an OptiTrack Motion Capture system. They also manually annotated the moving objects in frames at every 5 seconds. Each dataset contains 1100 frames.

Authors	Background model	Maintenance Scheme	Features	Information Combination	Depth Acquisition
Gordon and Darrell (1999) [46]	GMM	Online K-means approximation	YUV depth	Combined in the same model	Two Cameras
S. Ottonelli et al. (2013) [78]	GMM RGB GMM depth Compensation factor	Not defined	RGB Depth	Logical operations	Stereo Camera
S. Ottonelli et al. (2013) [79]	ViBe RGB ViBe depth Compensation factor	Not defined	RGB Depth	Logical operations	Stereo Camera
Langmann et al. (2010) [62]	GMM	New Gaussian creation or parameters update	YCbCr depth	Same Model	ToF
Jaime Gallego and Montse Pardàs (2014) [44]	Background model: two independent Gaussians per pixel. Foreground model: Spatial Color Gaussian Mixture Model (SCGMM) Spatial Depth Gaussian Mixture Model (SDGMM)	Separate update for background and foreground model	RGB Depth	Logarithmic Opinion Pool and the Hellinger distance	ToF
A. Störmer et al. (2010) [95]	GMM	FIFO Iteratively updated	Near infrared Range data	Combined in the same model	ToF
J.Leens et al. (2009) [64]	Motion maps	Not defined	gray scale PMD: distance amplitude intensity	Logical operations	ToF
D. Giordano et al. (2014) [45]	Binned KDE	Iteratively updated	Hue Saturation depth	Combined in the same model	Kinect
Y.Song (2014) [94]	GMM	Not defined	gray scale depth	Probabilistic model based on Gaussian distribution	Kinect
Camplani et al. (2013) [18]	Background: GMM for color and another for depth Foreground: uniform distribution	Iteratively updated	RGB depth	Statistical classifiers	Kinect
Camplani et al. (2014) [19]	Fusion of multiple region-based classifiers. Different background models, based on GMM	Iteratively updated	RGB depth	Mixture of experts	Kinect
Enrique J. Fernandez-Sanchez et al. (2013) [40]	Codebook	Not defined	RGB depth	Codeword	Kinect
Clapés et al (2013) [24]	Four dimensional Gaussian distribution	Not defined	RGB depth	Combined in the same model	Kinect

Tab. 3.1: Background subtraction algorithms using color and depth information: An overview.

In [18] the authors present a six-video RGBD dataset recorded using a Microsoft Kinect device with hand labelled ground truth in a indoor scenario.

In [81] present a dataset featuring a total of 5724 annotated frames divided in three indoor scenes. For each scene, a total of three persons are interacting, reading, walking, sitting, reading, etc. Every person is annotated with a unique ID in the scene on a pixel-level in the RGB modality. For the thermal and depth modalities, annotations are transferred from the RGB images using a registration algorithm.

3.3 Challenges of depth data

Depth sensors provide partial geometrical information about the scene, where each pixel depth value is proportional to the estimated distance from the device plane to the point in the real world. Among several technologies, recently there are two main types of consumer depth sensors that have become widely popular and accessible, based on Structured light and Time-of-flight.

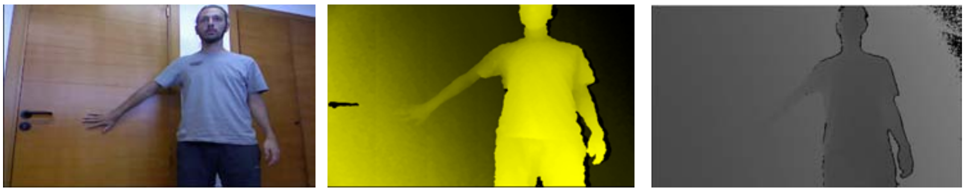
Structured light sensors: consist of an infra-red (IR) emitter and an IR camera. It estimates depth by structured light coding technology. Its IR emitter projects an IR speckle pattern onto the scene. IR camera captures the reflected pattern and correlates it against a stored reference pattern of a plane. These sensors have a lack of sensitivity and they are not able to estimate depth at all pixels in the scene. The noise of depth measurements increases quadratically with increasing distance to the sensor [55]. In consequence, between consecutive values gaps are found.

Time-of-flight sensors (TOF): resolve the distance based on the known speed of light. Depth is proportional to the time needed by the active illumination source to travel from emitter to target. Typically IR light is used for this purpose. This technology provides better accuracy than structured light sensor and is less susceptible to generate shadows in the scene. Noise can be approximated well by means of a normal distribution [39].

Independently of the used technology, estimated depth data by these devices suffer from several problems which we describe here.

Depth Camouflage

Due to the sensor lack of sensitivity, when foreground and background are close at depth, the sensor gives the same depth data values. This makes it hard to segment the foreground from the background based on depth. In Fig. 3.1 there is a example of depth camouflage.

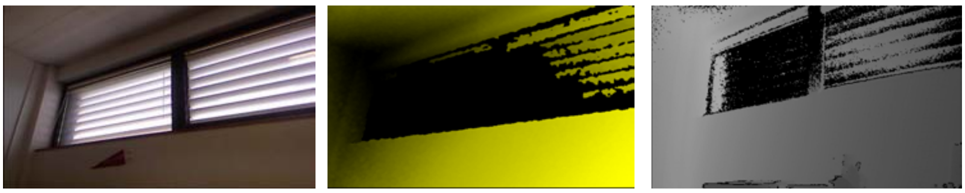


(a)RGB image. **(b)**Structured light sensor. **(c)**TOF sensor.

Fig. 3.1: Depth camouflage.

Specular materials

Rays from a single incoming direction are reflected back into a single outgoing direction without causing the diffusion needed in order to obtain depth information. This effect is described by the law of reflection, which states that the direction of incoming light (the incident ray), and the direction of outgoing light reflected (the reflected ray) make the same angle with respect to the surface normal. It is typically produced in smooth materials as glass or plastic. In Fig. 3.2 we can see the effect when the camera is pointing a window.



(a)RGB image. **(b)**Structured light sensor. **(c)**TOF.

Fig. 3.2: Specular materials.

Near objects

Sensors have minimum depth working distance. Due to the proximity of the foreground objects, the sensor is unable to have any depth measurement of that objects. Typically both structured light sensors and time-of-flight sensors have a near depth limit of 0.5 meters. See Fig. 3.3.

(a)RGB image. **(b)**Structured light sensor. **(c)**TOF sensor.

Fig. 3.3: Near objects.

Remote parts of the scene

Sensors have maximum depth distance detection. Parts of the scene farther from this distance appear as gaps in depth images. Typically both structured light sensors and time-of-flight sensors have a depth limit between 4 and 5 meters. See Fig. 3.4.

(a)RGB image. **(b)**Structured light sensor. **(c)**TOF sensor.

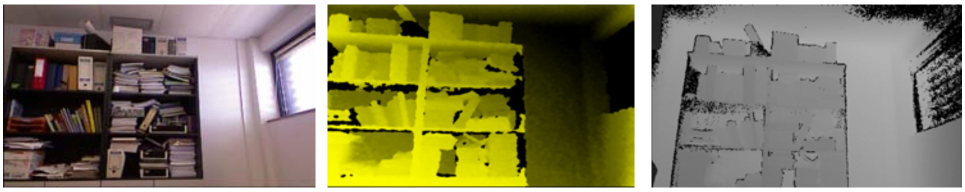
Fig. 3.4: Remote parts of the scene.

Non reachable areas

Depending on the imaging geometry and the sensor position, parts of the background may be occluded. This makes the sensor unable to estimate the depth at these locations. This effect normally takes place on the objects inside corners, as Fig. 3.5 illustrates.

Shadows

Foreground objects blocks the active light emitted by the sensor from reaching the background, which causes shadow cast on the background. Therefore the sensors are unable to estimate the depth at these blocked regions. Therefore RGBD sensors exhibit two different types of shadows: visible-light shadows in



(a)RGB image. (b)Structured light sensor. (c)TOF sensor.

Fig. 3.5: Non reachable areas.

the RGB channels, and IR shadows in the depth channel. These two different types of shadows are different in their geometry and in their spatial extent in the image. See Fig. 3.6.

(a)RGB image. (b)Structured light sensor. (c)TOF sensor.

Fig. 3.6: Shadows.

Absent depth observations

When depth can not be measured at a given pixel the sensor return a special non-value code to indicate its inability to measure depth. Such pixels appears as holes in the images with absence of depth value. In this work we denote these pixels as Absent Depth Observations (ADO), usually the depth channel value for an ADO pixel is 0. See black regions in Figs. 3.1-3.6.

3.4 Nonparametric scene model

In this section, we describe the theoretical background model. As we explained in the introduction our objective is to propose an unified probabilistic model that takes a sample for each pixel of the scene and estimates the probability that a newly observed pixel value belongs to the background.

3.4.1 Statistical model

The model is based on recent scene information. That is, given the last n observations of a pixel, denoted by $\mathbf{x}_i$, $i = 1, \ldots, n$, it is possible to estimate the probability density function (pdf) of each pixel with respect to all previously observed values [37, 90]

$$P(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} K \quad (\mathbf{x} - \mathbf{x}_i) \;, \tag{3.1}$$

where K is a kernel, that is, a non-negative function satisfying $\int K(x)dx = 1$ and $K(u) \quad 0$.   is the bandwidth parameter that controls the smoothness of the estimation.

Kernel density estimation allow us to estimate the pdf by averaging the effect of a set of kernel functions centred at each data observation. Kernel density estimators asymptotically converge to any density function with sufficient samples [90, 32]. This property makes the technique quite general for estimating the density of any distribution.

For multidimensional observations, we can generalize Eq. 3.1 as

$$P(\mathbf{x}) = \frac{1}{n} |\mathbf{H}|^{-\frac{1}{2}} \sum_{i=1}^{n} K(\mathbf{H}^{-\frac{1}{2}}(\mathbf{x} - \mathbf{x}_i)) \;, \tag{3.2}$$

where K is a multivariate kernel. **H** is the bandwidth matrix, which is a symmetric positive $d \quad d$-matrix in the d-dimensional observation space R^d, which enclose all the sensor data values.

The choice of the bandwidth matrix **H** is the single most important factor affecting the estimation accuracy, since it controls the amount of and orientation of smoothing induced [108]. Diagonal matrix bandwidth kernels allows different amounts of smoothing in each of the dimensions and are the most widespread due to computational reasons [107].

$$\mathbf{H} = \begin{pmatrix} {}^2_1 & 0 & & 0 \\ 0 & {}^2_2 & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & {}^2_d \end{pmatrix}$$

where σ_i^2 is bandwith of the kernel in the *i-th* dimension, i.e. independence between the different channels is assumed.

Instead of the matrix products of Eq. 3.2 independent products of one-dimensional kernels [90] can be used

$$P(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d} K_{\sigma_j}(x_j - x_{ij})\ , \tag{3.3}$$

A variety of kernel functions with different properties have been used in the literature. Typically, the Gaussian kernel is used for its continuity, differentiability, and locality properties. Note that choosing the Gaussian as a kernel function is different from fitting the distribution to a Gaussian model (normal distribution). Here, the Gaussian is only used as a function to weight the data points. Unlike parametric fitting of a mixture of Gaussian functions, kernel density estimation is a more general approach that does not assume any specific shape for the density function. In our approach $N(0, \mathbf{H})$ is selected. The final probability density function can be written as

$$P(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^{n}\prod_{j=1}^{d}\frac{1}{\sqrt{2\pi\sigma_j^2}}e^{-\frac{1}{2}\frac{(x_j - x_{ij})^2}{\sigma_j^2}}\ . \tag{3.4}$$

Given this estimate at each pixel, a pixel is considered foreground if its probability is under a certain threshold, τ, that is, If $P(x_t) < \tau$.

To estimate the kernel bandwidth σ_j^2 for the jth dimension for a given pixel, similar to [37], the median absolute deviation over the data for consecutive values of each pixel is computed. That is, the median, m_j, of each consecutive pair (x_i, x_{i+1}) in the data is calculated independently for each dimension. Since we are measuring deviations between two consecutive values. Each pair (x_i, x_{i+1}) usually comes from the same local-in-time distribution and only few pairs are expected to come from cross distributions. Assuming that this local in-time distribution is Normal $N(\mu, \sigma)$ then the deviation (x_i - x_{i+1}) is Normal $N(0, 2\sigma_j^2)$. Since this distribution is symmetric, the median of the absolute deviations is equivalent to the quarter percentile of the deviation distribution

$$\Pr(N(0, 2\sigma_j) > m_j) = 0.25\ . \tag{3.5}$$

So, the standard deviation of the first distribution can be estimated as:

$$
{}_j = \frac{m_j}{0.68\sqrt{2}} \; . \tag{3.6}
$$

Finally, in order to make a fast implementation of the algorithm, the probability is estimated given the pixel value difference and the Kernel function bandwidth using a pre-calculated lookup table.

3.4.2 Depth data modelling

We want to use the previous described scene model with color and depth information in an unified way. However, the model cannot be applied in a standard way because the sensor's ADO need a special treatment where depth is just treated as a fourth channel besides RGB. These ADO can introduce errors to our model as well as to any background model. In addition, a pixel can be ADO all over the sequence or switch in a random manner between ADO and a valid value. We classify two categories of ADO:

- Ones provoked by the scene physical configuration, and therefore they belong to the background, even in absence of foreground objects. These include specular background materials, remote parts of the scenes, and non reachable areas.

- Ones caused by the foreground objects. These include near objects, specular foreground objects and shadows.

We want to differentiate these two classes of ADO-pixels in order to model the ones that belong to scene in order to add some information to the background model. We propose a probabilistic model, we call ADO-model, that is the probability of a pixel being ADO and belonging to the background model, denoted by P_A. This probability is calculated for the depth component of each pixel, D, of last observations. ADO-model is updated for each pixel during the training stage. P_A is calculated recursively as follows:

$$
\begin{aligned}
P_A(D_0) &= 0 \qquad (3.7)\\
P_A(D_t) &= \alpha * mask_t + (1-\alpha) * P_A(D_{t-1}) \; ,
\end{aligned}
$$

where $\alpha \in [0, 1]$ is the update rate, and $mask_t$ is a binary value corresponding to an ADO-mask, where each pixel have value of 1 if D_t is ADO and 0 if not.

In order to avoid adding ADO-pixels to the background model, we selected a strategy based on overwriting the pixel with its previous one. Let D_t, $t = 1, \ldots, n$ be the n recently sampled depth values at a given pixel, D_t is calculated as follows:

$$D_t = \begin{cases} D_{t-1}, & \text{if } D_t = ADO \\ D_t, & otherwise \end{cases}, \tag{3.8}$$

where for D_0, we use the suggested inpainting strategy in [49], applying an initial image reconstruction algorithm by [100] that tries to estimate the correct values for ADO regions (see Fig. 3.7). For the following depth frames each ADO pixel is overwritten by a previous value (see Eq. 3.8). Once this process is done, background model is updated.

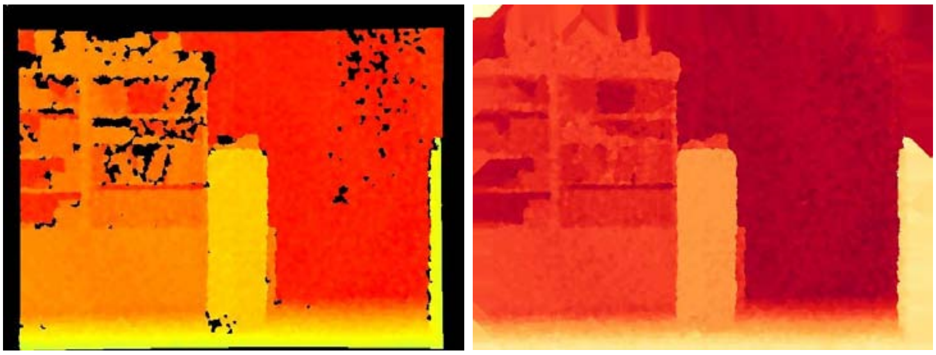


Fig. 3.7: Inpainting process: Left image depicts first sequence frame after inpainting process. In the right image it can be observed that holes appearing in left images have disappeared due the process.

This ADO-model is calculated for each pixel during the training phase. Fig. 3.8 depicts the ADO-model computation process: Pixels with P_A higher than a threshold, , are overwritten with a previous value and incorporated to the background model. The other ones rest as undefined pixels and then are classified as the undefined class: If $D_t = ADO$ and $P_A(D_t) <$.

3.4.3 Background moving object detection

Depth information can be useful to address different problems than using color information due its different nature. In real scenes a background object

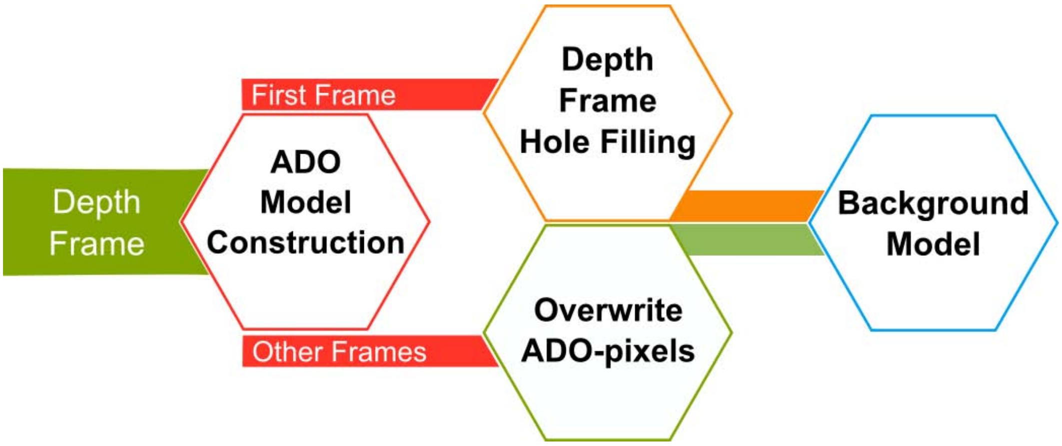


Fig. 3.8: Training step: In the first frame we apply an inpainting process. For the following depth frames each ADO pixel is overwritten by a previous value. Once this process is done pixels are added to background model.

can be moved. Such area should not be considered part of the foreground forever after, so the scene model has to adapt and understand that the scene layout may be physically updated [29]. Typically in background subtraction algorithms the new background is incorporated to the model at different speed depending on the corresponding update rates, see Eq. 3.9. By introducing depth data we present a new approach, allowing instantaneous pixel classification.

$$pixel_t \quad = \quad \alpha * mask + (1-\alpha) * pixel_{t-1} \; . \tag{3.9}$$

The idea is based on the fact that if a new depth observation is located farther than the modelled values, this is probably because it becomes part of the background because some object has been removed from the scene. In order to detect these changes we compare the difference of this new observation with the background model and we evaluate if that difference is bigger than the modelled one for each pixel. The cumulative density function (cdf) over the absolute difference of two consecutive observations of a pixel permits to formalize this idea.

Given the last observations $D_1, \ldots, D_n$ the *ith* component of vector **V** is defined as

$$V_i \;= \mid D_i - D_{i-1} \mid, \; \forall i \;=\; [1 \ldots n], \tag{3.10}$$

for all k possible sensor values, k $\in \{0 \ldots L\}$, where L is the maximum number of depth levels.

Then, we define $P(k) = \frac{1}{n}\# \ V_i : V_i = k \ $,
and
$$F_x(k) = \sum_{j=1}^{k} P(j) \ . \tag{3.11}$$

Finally, given a new observation D_t and the observations $D_1, \ldots, D_n$, we define the the *ith* component of evaluation set C_{D_t} as the threshold to zero of the difference. That is

$$C_{D_{t,i}} = \begin{cases} D_t - D_i, & \text{if } D_t - D_i > 0 \\ 0, & otherwise \end{cases} , \tag{3.12}$$

$\forall \ i = [1 \ldots n]$. The evaluation function for background moving objects detection is

$$U(D_t) = \frac{\sum \ k \ \ C_{D_t} \ F_x(k)}{n} \ . \tag{3.13}$$

If $U(D_t)$ is higher than a predefined threshold, , the pixel is considered part of the background.

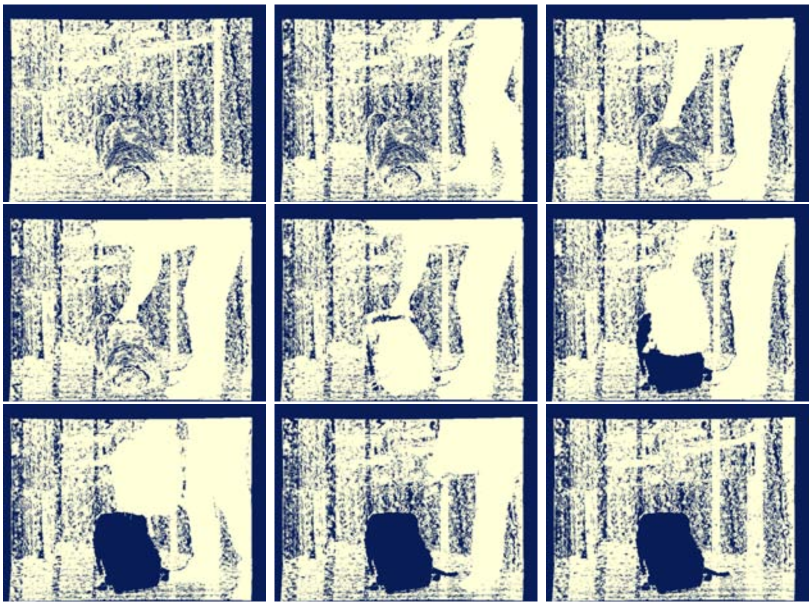
This detection is very relevant, so physical changes in the scene are detected when occur. The 3.9a contains a sequence illustrating an example of background object moving detection, Fig 3.9b depicts the same sequence once the algorithm has been applied.

3.5 Model update

In previous sections we detailed how to detect foreground regions of a scene given a recent history sample for each pixel. This model needs to be dynamically updated in order to properly respond to changes in the scene. As the kernel bandwidth estimation requires all the samples to be close in time, the update is performed using a first-in first-out queue, the oldest sample is discarded and a new sample is added to the model.

Different update strategies are used in order to keep the model updated. On one hand, color information tends to have quick variations due to shadows and variant luminance so we consider it as an unstable model. On the other hand, depth information tends to be more stable.

(a) Set of color frames background from a moving object sequence from an existing dataset.



(b) Same sequence once the background moving objects algorithm has been applied. Blue pixels are those classified as background.

Fig. 3.9: Graphical explanation of background moving objects algorithm.

Color update

The intensity distribution of the color information can change dramatically over very short periods of time [37]. For each frame the color model is updated so the model can adapt very quickly to changes in the background process. A new observation is added to the model only if it is classified as a background sample. If a pixel is updated with the foreground color value, the error will be propagated over time and misclassification problem will be raised. In order to avoid the model adaptation to the foreground objects characteristics, a higher threshold is proposed in order to have a less strict condition and avoid updating pixels that are very close to belong to the foreground.

Depth update

Unlike color, depth information represents a stable long-term description of the scene. Therefore, it is not necessary to update the model each frame as pixel values do not change as fast as color values. Pixels detected as a part of a background moving object are automatically classified as background and their models are updated. In fact, updating depth model is very related to physical changes in the real world scene, so pixels detected as background moving objects (see Section 3.4.3) are the selected pixels to be updated. In addition the ADO-model is updated for these pixels during this update phase.

3.6 Generic Scene Modelling (GSM)

In this Section we describe an experimental configuration of the scene modelling algorithm. Prior to the use of the scene model it is necessary to perform a training stage where models of color and depth information are learned and the bandwidth of each used channel is calculated at each pixel. In Table.3.2 there are the complete algorithm details.

As it is explained in Section 3.4.2 the algorithm classifies pixels in three different classes: background, foreground and undefined. Undefined pixels are those ones that are ADO and we cannot ensure that belong to the background (see Table.3.2 for details), the final decision of decide what to do with pixels that belong to the undefined class is left to the final user depending the application purposes and environmental conditions. In order to have a complete evaluation we used two implementations of the proposed GSM algorithm: GSM_{UB} if undefined pixels are considered as background and GSM_{UF} if undefined pixels are considered as foreground.

Let define an observation $\mathbf{x} = \quad r, g, D \quad$, therefore d = 3. θ, γ and ξ are constant values over all algorithm, where: $\gamma = 10^{-8}$, $\theta = 0.0050$ and $\xi = 0.6$.

Training stage

Initialization step:

- Image inpainting algorithm to compute D_0.
- $P_A(D_0) = 0$.

Let ($\mathbf{x}_1, \ldots, \mathbf{x}_i$, $\mathbf{x}_n$) the observations used for modelling the scene for each $D_i \in \mathbf{x}_i$

Depth treatment:

- Compute the ADO-model: $P_A(D_i) = \alpha * mask + (1 - \alpha) * P_A(D_{i-1})$.
- *If* D_i is ADO-pixel *then* $\hat{D}_i = \hat{D}_{i-1}$ *else* $\hat{D}_i = D_i$.
- Substitute $\hat{D}_i$ for D_i in $\mathbf{x}_i$.

Calculate the kernel bandwidth for each dimension *d*.

- $\forall j \in [1..d]$ compute the median absolute deviation m_j for consecutive values of observations:

$$m_j =\mid x_{ji} - x_{ji+1} \mid \forall\, \boldsymbol{i} \in [1 \ldots t].$$

- Calculate de bandwith $\sigma_j = \frac{m_j}{0\ 68\sqrt{2}}$.

Segmentation stage

Let $\mathbf{x}_t$ a new observation

Depth treatment:

- *If* D_t is ADO-pixel *then* $\hat{D}_t = \hat{D}_{t-1}$ *else* $\hat{D}_t = D_t$.
- Measure the probability of a pixel being part of a background moving object:

$$U(\hat{D}_t) = \frac{1}{n} \sum_{\forall k \in C_{\hat{D}_t}} F_x(k).$$

- *If* $U(\hat{D}_t) > \xi$ *then* $\mathbf{x}_t \in$ **background**.

Calculate the probability of a pixel being background for all dimensions, d:

- $P(\mathbf{x}_t) = \frac{1}{n} \sum_{i=1}^{n} \prod_{j=1}^{d} \frac{1}{\overline{2 \quad {}_j^2}} e^{-\frac{1}{2} \frac{(x_{nj} - x_{ij})^2}{\sigma_j^2}}$.
- *If* $P(\mathbf{x_t}) < \gamma$ *then* $\mathbf{x}_t \in$ **foreground** *else* $\mathbf{x}_t \in$ **background**.
- Classify $\mathbf{x}_t \in$ as **unde ned** *If* D_t is ADO-pixel *and* $P_A(D_t) < \theta$.

Update background model:

- *If* $\mathbf{x}_t \in$ **background** *then* update color model.
- *If* $U(\hat{D}_t) > \xi$ *then* update depth model.

Tab. 3.2: The generic scene modelling algorithm for RGBD devices.

4 Evaluation

The evaluation of a new method is a very relevant task so as to discover its strengths and weaknesses compared to the other existing proposals. In this chapter, the Generic Scene Model described previously is evaluated using an existing dataset. In addition, a new RGBD video dataset is presented in order to introduce a new standard for comparison purposes of this kind of algorithms. Results show that Generic Scene Model (GSM) can handle all depth issues, obtaining better results than state-of-the-art algorithms.

4.1 Introduction

The GSM algorithm described in the previous chapter needs to be properly evaluated in order to ensure its performance.

In order to evaluate the previously defined scene model algorithm a Microsoft Kinect 1 sensor is used as RGBD device, see Fig 4.1. The device's technology is based on structured light. The image processor of the sensor uses the relative positions of the dots in the pattern to calculate the depth displacement at each pixel position in the image [55]. We use the sensor raw depth information value D, which corresponds to a depth valid range between 0.5 and 4.5 meters. In addition, for this sensor, all ADO have the same value, that is 650.

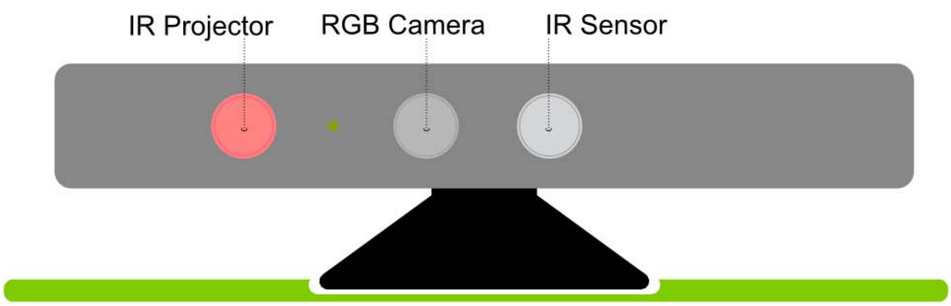


Fig. 4.1: RGBD used sensor scheme.

Usually, color information is useful for suppressing shadows from detection by separating color information from lightness information. In order to construct a robust algorithm independent of illumination variations we separated color information from luminance information using a non luminance dependent space color. Then, color is defined as a combination of luminance, hue and saturation. Chromaticity is an objective specification of the quality of a color regardless of its luminance. It can be described as a combination of hue and saturation. Given the device's three color channels *R, G, B*, the chromaticity coordinates *r, g* and *b* are: $r = R\ (R+G+B)$, $g = G\ (R+G+B)$, $b = B\ (R+G+B)$ where: $r+g+b = 1$ [65]. In our model we use two dimensions: r and g.

The evaluation process was done using two datasets. First, we compared with an existing dataset which emphasizes on camouflage and shadows problems, Camplani dataset. We selected this dataset because facilitates us comparing with other background subtraction algorithms using both color and depth information [18] and up to our knowledge is the first one made using RGBD information. Second, a new RGBD dataset is built inspired by one of the most used color-based background subtraction datasets, the WallFlower dataset [103]. As we will explain, it is designed in order to test all background subtraction issues described in Wallflower besides the new depth challenges described in the previous chapter.

Different metrics can be used to measure the algorithm's performance. For the two performed tests, all the used ones are based on the following basic metrics: **True Positives (TP)**: counts the number of correctly detected foreground pixels. **False positives (FP)**: counts the number of background pixels incorrectly classified as foreground. **True negatives (TN)**: counts the number of correctly classified background pixels. **False negatives (FN)**: counts the number of foreground pixels incorrectly classified as background.

This chapter is organized as follows: first part is dedicated to the evaluation of the presented algorithm using an existing dataset. Second part is dedicated to the evaluation of the algorithm using the dataset we designed. In each part, there is the description of each sequence, the description of the evaluation method and the evaluation results. Finally, there are the conclusions of the chapter.

4.2 The Camplani Dataset

In [18] the authors presented a six-video RGBD dataset with hand labelled ground truth. Up to our knowledge this is the first RGBD benchmark dataset with hand-labelled ground truth that includes sequences with different challenging scenarios for foreground segmentation.

4.2.1 Description

In this section we describe each used sequence of the Camplani dataset, information is summarized in Table 4.1.

Sequence	Number of frames	Ground truth frames	Test Objective
GenSeq	300	39	Test the overall performance of the algorithm in case of general scenes.
DCamSeq	670	102	Analyse the performance of the algorithm when depth camouflage occur.
ColCamSeq	360	45	Test the performance of the algorithms when color camouflage problem occur .
ShSeq	250	25	Highlight the impact that shadows projected by moving objects.
MovedBG	250	-	Evaluate the algorithms where an object from the background is moving.

Tab. 4.1: Characteristics of evaluated sequences from Camplani dataset.

GenSeq

GenSeq is used to test the overall performance of the algorithm in case of general scenes taking into account all the possible error contributions to the scene (see Fig. 4.2). It is composed by 300 frames and the corresponding ground truth is composed by 39 frames spanning 115 frames of the sequence where the moving objects are present, one of every three frames has been labelled.

Fig. 4.2: **GenSeq** sequence. Left frame depicts an example of depth frame for this sequence. Two next frames depict the start and the end of the sequence.

DCamSeq

DCamSeq helps to analyse the performance of the algorithm when depth camouflage occur (see Fig. 4.3). It is designed to evaluate the performance of the algorithm when interactions between foreground and background elements of the scene occur. The sequence contains 670 frames and the ground truth is composed by 102 frames that cover 400 frames of the sequence where the moving object is present. One of every four frames has been labelled as ground truth.

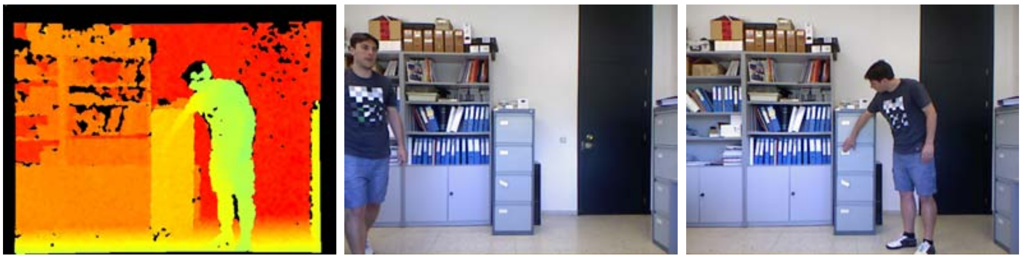


Fig. 4.3: **DCamSeq** sequence. Left frame depicts an example of depth frame for this sequence. Two next frames depict the start and the end of the sequence.

ColCamSeq

Sequence *ColCamSeq* aims to test the performance of the algorithms when the color camouflage problem occur (see Fig. 4.4). It contains 360 frames and the ground truth is composed by 45 frames that cover 240 frames of the sequence that are the one where the moving object is present.

Fig. 4.4: **ColCamSeq** sequence. Left frame depicts an example of depth frame for this sequence. Two next frames depict the start and the end of the sequence.

ShSeq

The goal of this sequence is to highlight the impact that shadows projected by moving objects have on the foreground segmentation algorithm (see Fig. 4.5). It is composed by 250 frames and the corresponding ground truth is composed by 25 frames spanning 120 frames of the sequence where the issue to evaluate is present.

Fig. 4.5: **ShSeq** sequence. Left frame depicts an example of depth frame for this sequence. Two next frames depict the start and the end of the sequence.

MovedBG

The goal of the sequence is to evaluate the algorithms where an object from the background is moving, that is, a waking object (see Fig. 4.6). The sequence is composed by 250 frames. There is not ground truth frames as the dataset authors argue that since the new background is incorporated at different speed by the different algorithms (depending on the corresponding learning rates) they cannot be compared.

Fig. 4.6: **MovedBG** sequence. Left frame depicts depth information. Two next frames depict the start and the end of the sequence.

4.2.2 Metrics

Camplani *et al.* compared eight different scene modelling algorithms: the Camplani algorithm (CL_W), two weak classifiers defined in their paper (CL_C and CL_D), four different implementations of mixture of Gaussian and ViBe. In order to perform the evaluation they used seven metrics: the generic **FN** and **FP**, **Total error (TE)**, that is, the total number of misclassified pixels, normalized with respect to the image size. **Similarity measure (S)** that is a non-linear measure that fuses FP and FN and it is close to 1 if detected foreground regions correspond to the real ones, otherwise its value is close to 0. Finally, a **Similarity measure in object boundaries (S$_B$)**, it is calculated as S, but considering only the regions of the image surrounding the ground truth object boundaries of 10 pixel width. Two different metrics are used to rank the accuracy of the analysed algorithms. **RM**, rank each method for each performance metric for one sequence. Let us define $rank_i(m,\ sq)$ as the rank of the *ith* method for the performance metric m in the sequence sq, the average ranking of the method i in the sequence sq is calculated as:

$$RM_i = \frac{1}{N_m} \sum_m rank_i(m, sq), \tag{4.1}$$

where N_m is the number of performance metrics. **RC**, is defined as a global ranking of algorithms across all sequences, that is computed as the mean of RM for each method across all the sequences. Table. 4.2 summarizes these metrics.

Measure	Description
False Positives (FP)	Number of false positives.
False Negatives (FN)	Number of false negatives.
Total error (TE)	Total number of misclassified pixels, normalized with respect to the image size, that is: $(FP + FN) \quad Image\ Size$.
Similarity measure (S)	Non-linear measure that fuses FP and FN and it is close to 1 if detected foreground regions correspond to the real ones, otherwise its value is close to 0, that is: $(FP \quad FN) \quad (FP \quad FN)$.
Similarity measure in object boundaries (S$_B$)	Calculated as S, but considering only the regions of the image surrounding the ground truth object boundaries of 10 pixel width.
RM	Rank each method for each performance metric for one sequence: $RM_i = \frac{1}{N_m} \sum_m rank_i(m, sq)$.
RC	Global ranking of the algorithms across different sequences, that is: $RC_i = \frac{1}{N_{sq}} \sum_{i=1}^{N_{sq}} RM_i$.

Tab. 4.2: Description of the metrics used to evaluate the algorithm's performance.

In the following section we compare the performance of the following algorithms: the adaptive weighted classifier CL$_W$ and the two weak classifiers CL$_C$ and CL$_D$ presented in [18]; the MoG algorithm proposed in [46] (MOG$_{RGBD}$); the binary combinations of the foreground masks obtained by two independent modules based on depth and color data as proposed in [95] (by using MoG) and in [64] (by using ViBe), we refer to these algorithm as MOG$_{Bin}$

and Vibe$_{Bin}$, the algorithm proposed in [68] (SOM), the MoG algorithm proposed in [110] (MoG$_{Ziv}$). Finally, the two implementation, GSM$_{UB}$ and GSM$_{UF}$, of the proposed GSM algorithm.

4.2.3 Results

To summarize the method's performance, in Fig. 4.7 the ranking of each sequence is shown, that is the RM classification. In addition, it is shown, the RC classification in order to establish a global result. Both GSM and CL$_{W}$ have qualitative best results in RC ranking, for that reason and in order to understand the global results, we analyse the performance of both algorithms for each sequence.

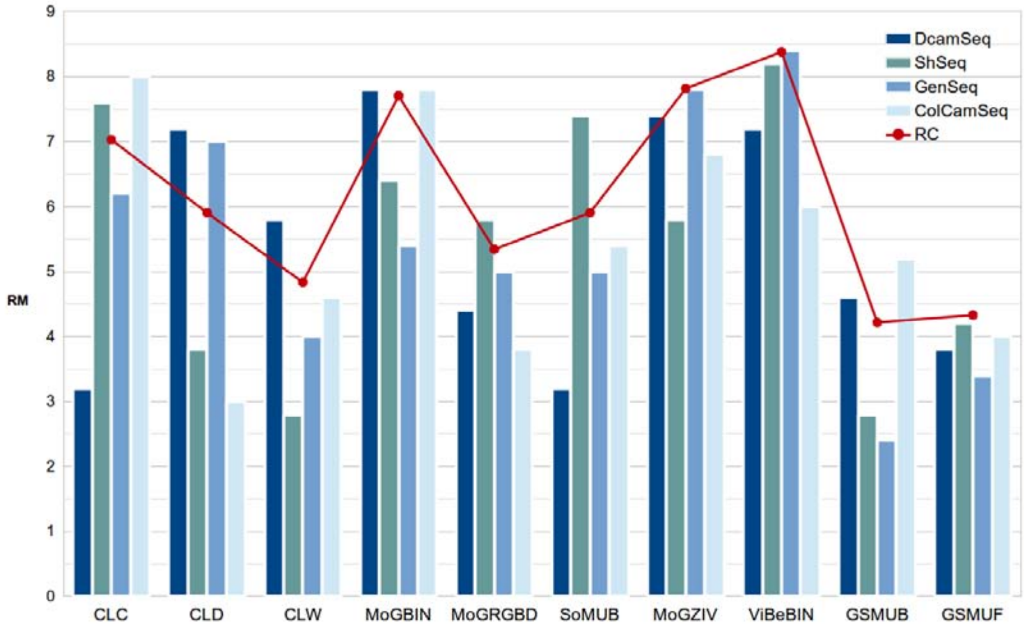


Fig. 4.7: Camplani dataset simulation results. It can be found results of four different sequences for all tested algorithms and final comparisons. As GSM$_{UF}$ and GSM$_{UB}$ are not the best for each sequence, are the most regular ones as it can be seen with the **RC** line, the lower the better.

The measures described in Table 4.3 corresponds to the algorithms when *GenSeq* is evaluated. We can observe that GSM$_{UF}$ has higher FN due to the classification of all undefined pixels-class as foreground. Anyway, both of our proposed solutions have better values in contours (S_B) than the CL$_{W}$ one. In Fig 4.8 it is important to observe the overall image. In last two rows the person leaves the box on the table and we can observe that it perfectly fits on the scene, as the edges does not blend with the background. Finally, as expected, we can note that GSM$_{UB}$ achieves the best global result for this sequence.

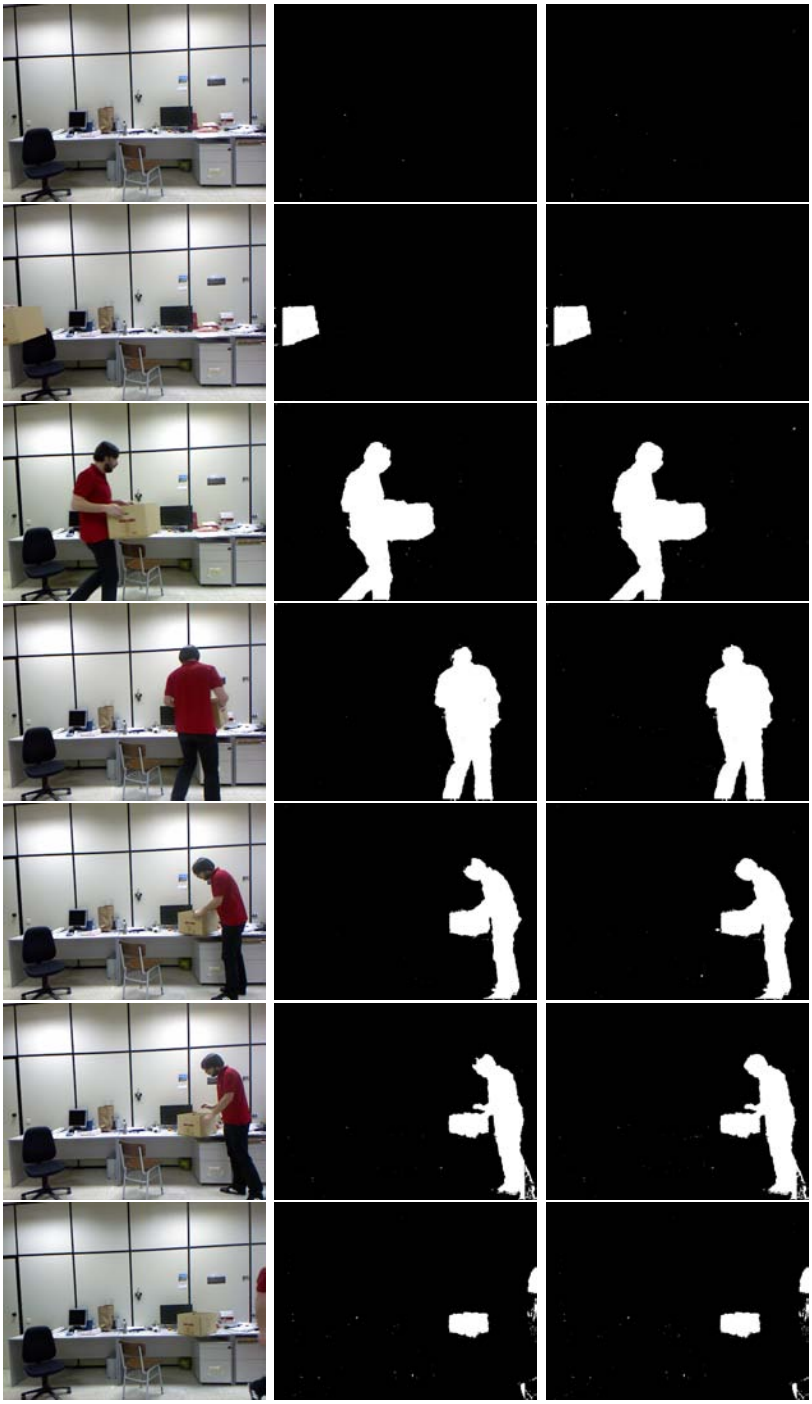


Fig. 4.8: Results of **GenSeq sequence**: Left column corresponds to color images used as reference, right column is the result of the algorithm considering undefined pixels as foreground (GSM_{UF}) and the column at the center is the result of the algorithm considering undefined pixels as background (GSM_{UB}).

	TE		FN		FP		S		S_B		RM
	Avg	**Std**	**Avg**	**Std**	**Avg**	**Std**	**Avg**	**Std**	**Avg**	**Std**	
CL_W	1.30	0.42	1.49	0.002	1.27	0.01	0.83	0.21	0.53	0.14	3.2
GSM_{UB}	1.38	0.56	1.04	0.78	1.44	0.66	0.83	0.2	0.78	0.11	2.6
GSM_{UF}	1.3	0.52	4.08	15.38	1.3	0.6	0.83	0.2	0.78	0.14	3.2

Tab. 4.3: Results for **GenSeq**. **FP**: False positives. **FN**: False negatives. **TE**: Total error. **S**: Similarity measure. **S_B**: Similarity measure in object boundaries.

Sequence *DCamSeq* is very relevant in that dataset as it is the unique sequence that test one of the depth information issues: depth camouflage. As we can see in Fig. 4.9 in the shelf area we do not have any relevant depth information, so all the work is done using the color information. Our algorithm achieve best results these kind of situations (see Table 4.4) due the combination of two kind of information, color and depth, in the same model. This situation is changeling when mixing color and depth information, other state-of-art algorithms emphasize the weight of depth information, so, when this kind of information is not relevant they have misclassification.

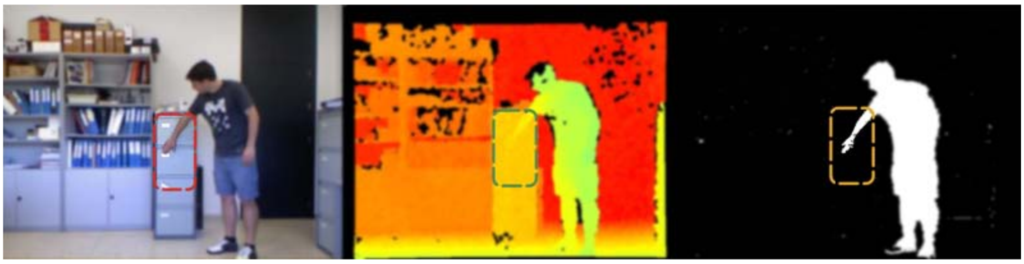


Fig. 4.9: Detail of segmentation when depth camouflage occurs. Second frame depicts there is not relevant depth information. Third frame show the segmentation result.

	TE		FN		FP		S		S_B		RM
	Avg	**Std**	**Avg**	**Std**	**Avg**	**Std**	**Avg**	**Std**	**Avg**	**Std**	
CL_W	2.46	1.82	32.21	0.26	0.66	0.01	0.55	0.14	0.5	10.12	6.2
GSM_{UB}	1.74	1.7	20.45	10.73	0.46	1.57	0.64	0.17	0.54	0.14	3.8
GSM_{UF}	1.65	1.49	22.06	11.6	0.61	1.73	0.65	0.18	0.55	0.14	3.6

Tab. 4.4: Results for **DCamSeq**. **FP**: False positives. **FN**: False negatives. **TE**: Total error. **S**: Similarity measure. **S_B**: Similarity measure in object boundaries.

When analysing color camouflage (*ColCamSeq*), both implementations of GSM have better values in the similarity measure (S) and in contours (S_B) than the CL_W algorithm. Regarding undefined pixels, the sequence *ColCamSeq* (see Table. 4.5) is the opposite case of *GenSeq*, GSM_{UB} has higher FN due to the misclassification of all undefined pixels as background, for that reason that method has worth RM score than CL_W. Finally, we can highlight that GSM_{UF} has the best RM score for that sequence. In Fig. 4.10 we show both algorithms results (GSM$_{UB}$ and GSM$_{UF}$) over the sequence designed to evaluate this issue.

	TE		FN		FP		S		$\mathbf{S}_B$		RM
	Avg	**Std**	**Avg**	**Std**	**Avg**	**Std**	**Avg**	**Std**	**Avg**	**Std**	
CL$_W$	3.20	2.77	3.52	0.09	2.92	0.10	0.89	0.15	0.77	0.16	4.8
GSM$_{UB}$	2.3	2.26	7.1	14.5	3.21	6.3	0.9	0.15	0.52	0.11	5.2
GSM$_{UF}$	2.2	2.27	2.94	5.53	4.36	6.42	0.92	0.08	0.53	0.09	4

Tab. 4.5: Results for **ColCamSeq**. **FP**: False positives. **FN**: False negatives. **TE**: Total error. **S**: Similarity measure. $\mathbf{S}_B$: Similarity measure in object boundaries.

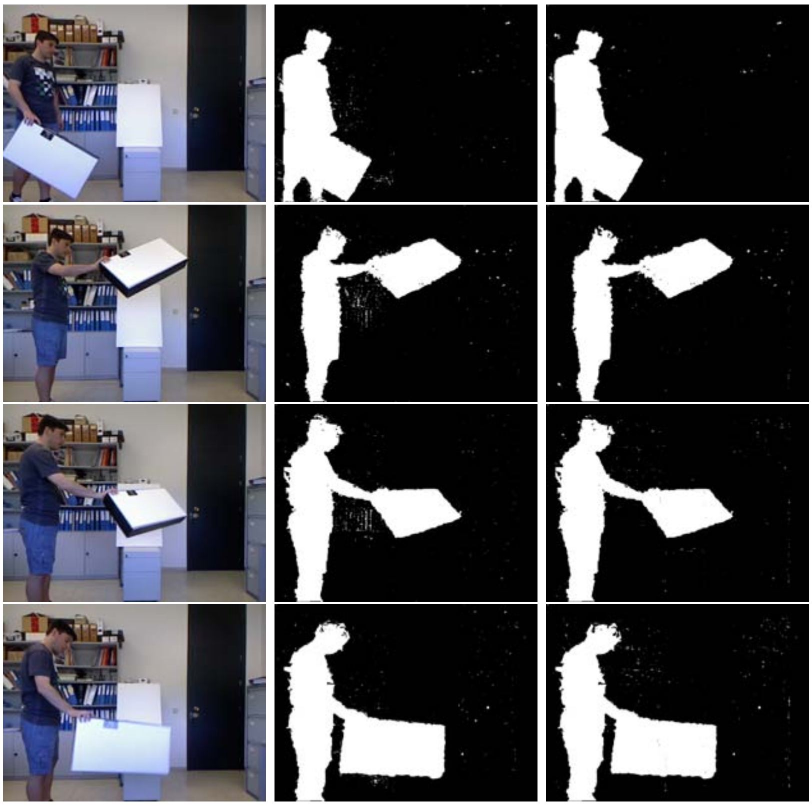


Fig. 4.10: Results of **ColCamSeq sequence**: Left column corresponds to color images used as reference, right column is the result of the algorithm considering undefined pixels as foreground (GSM$_{UF}$) and the column at the center is the result of the algorithm considering undefined pixels as background (GSM$_{UB}$).

Results in Table 4.6 depicts the performance of algorithms in *ShSeq* sequence. We can observe that both GSM$_{UB}$ and CL$_W$ have very similar results, the classification of undefined pixels as foreground class (GSM$_{UF}$), provokes the apparition of a higher number of false positives, due to this reason this algorithm has worst RM score. The importance of this sequence is focused on the shadow apparition on the floor, see Fig. 4.11. As the box approximates to the ground a shadow appears on it. After that, the box rotates on its axis provoking the apparition of new shadows. We can observe that in segmented images shadows does not appear.

	TE		FN		FP		S		S_B		RM
	Avg	**Std**	**Avg**	**Std**	**Avg**	**Std**	**Avg**	**Std**	**Avg**	**Std**	
CL_W	0.81	0.35	1.60	0.05	0.68	0.02	0.94	0.04	0.71	0.07	2.80
GSM_{UB}	0.87	0.33	0.98	0.88	0.88	0.42	0.93	0.03	0.76	0.06	2.80
GSM_{UF}	1.66	0.38	0.14	0.19	1.92	0.44	0.89	0.04	0.65	0.05	4.2

Tab. 4.6: Results for **ShSeq**. **FP**: False positives. **FN**: False negatives. **TE**: Total error. **S**: Similarity measure. $\mathbf{S}_B$: Similarity measure in object boundaries.

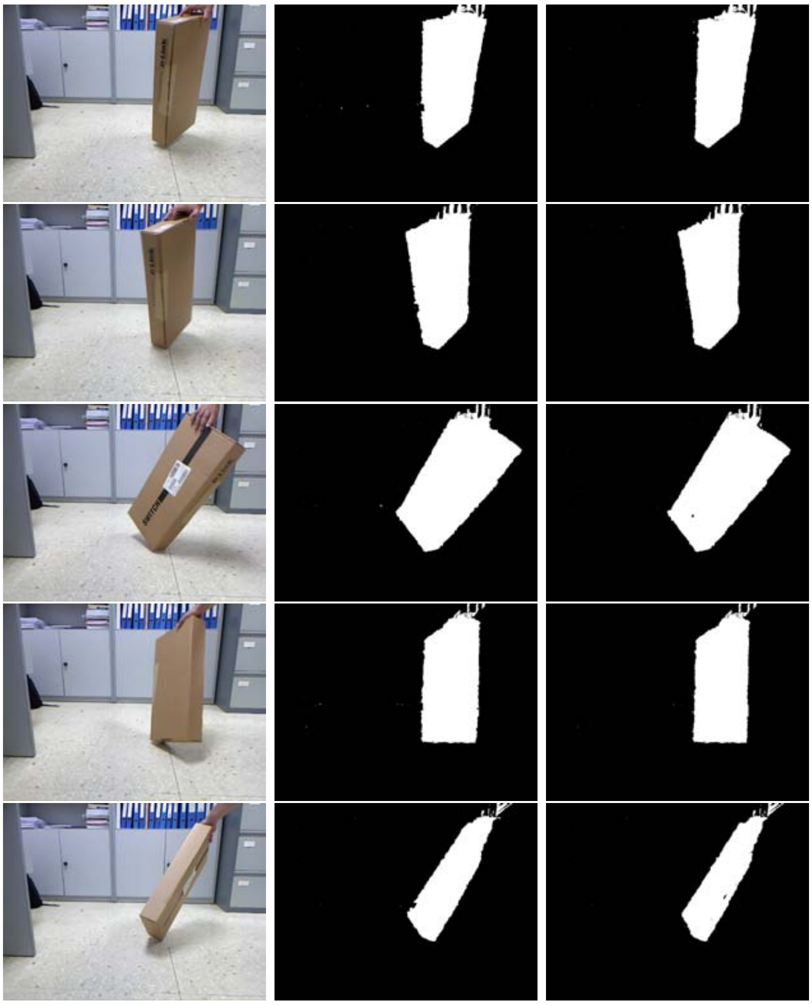


Fig. 4.11: Results of **ShSeq sequence**: Left column corresponds to color images used as reference, right column is the result of the algorithm considering undefined pixels as foreground (GSM_{UF}) and the column at the center is the result of the algorithm considering undefined pixels as background (GSM_{UB}).

As we explained in the previous section, normally background is incorporated at different speed by the different algorithms so they cannot be quantitatively compared, Fig. 4.12 depicts the importance of applying the background moving object algorithm (see Section 3.4.3). The same frame the objects move, that region is classified as background.

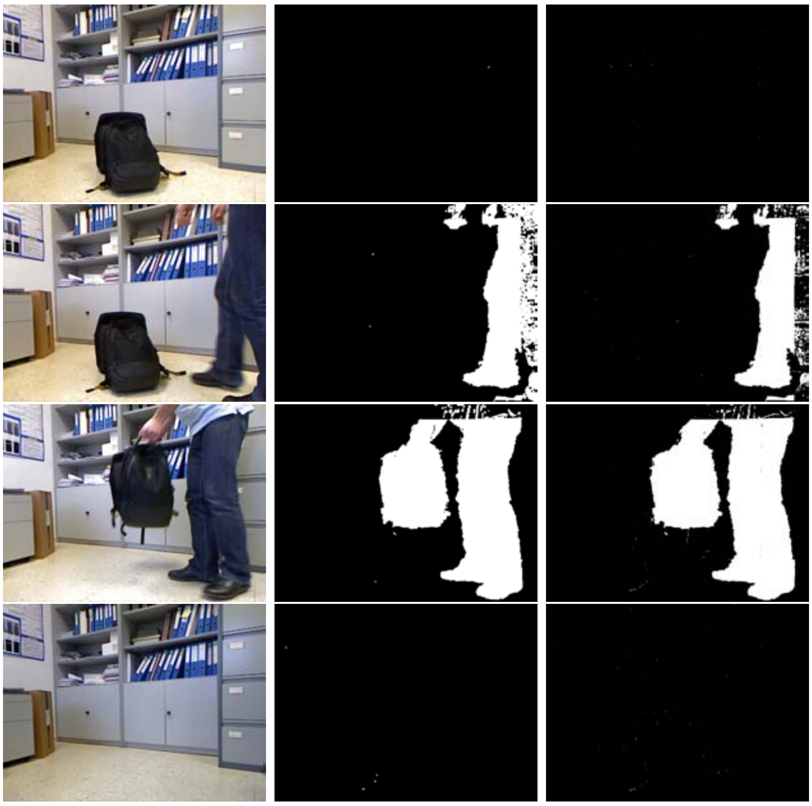


Fig. 4.12: Results of **MovedBG sequence**: Left column corresponds to color images used as reference, right column is the result of the algorithm considering undefined pixels as foreground (GSM$_{UF}$) and the column at the center is the result of the algorithm considering undefined pixels as background (GSM$_{UB}$).

4.3 GSM Dataset

As we explained before, in the previous work there is no general purpose RGBD dataset which covers all the desirable types of sequences in order to properly evaluate a scene modelling algorithm. Each algorithm is evaluated using its own proposed data and different metrics. This fact, as we seen in the Camplani dataset makes very difficult to perform an unified comparison between different methods. For this purpose, we propose a comprehensive dataset that covers all the challenges that occur when combining depth and color information.

4.3.1 Description

The GSM dataset includes 7 different sequences (see Table 4.7) designed to test each of the main problems in scene modelling when both color and depth information are used. Each sequence starts with 100 training frames, and have a foreground hand labelled ground truth.

Sequence	Number of frames	Ground truth frames	Test Objective
Time of day	1231	23	Subtle illumination changes.
Colour Camouflage	428	11	Segmentation when colour information is not relevant.
Depth Camouflage	465	12	Segmentation when colour information is not existing or is not relevant.
Shadows	330	11	Apparition of shadows in the scene background.
Light Switch	407	9	Sudden illumination changes.
Bootstrapping	300	11	Moving objects in training frames.
Waking object	200	10	Detect objects moving from the scene background.

Tab. 4.7: Characteristics of sequences from GSM dataset.

Time of Day

This 1231 frame sequence is designed to evaluate smooth illumination changes in the scene (see Fig. 4.13). After training frames no moving object appears in the scene, subtle illumination changes occur during the sequence. The ground truth is composed by 23 frames that cover the most relevant part of the sequence where there are scene illumination changes.

Fig. 4.13: **Time of day** sequence. Left frame depicts an example of depth frame for this sequence. Two next frames depict the start and the end of the sequence.

Color camouflage

The sequence has 428 frames, is used to evaluate the algorithm when color camouflage take place(see Fig. 4.14). After 100 training frames a person appears and places a folder on a shelf covering other folders of the same color. The ground truth is composed by 11 frames that cover the most relevant part of the sequence where there are moving objects and color camouflage happen.

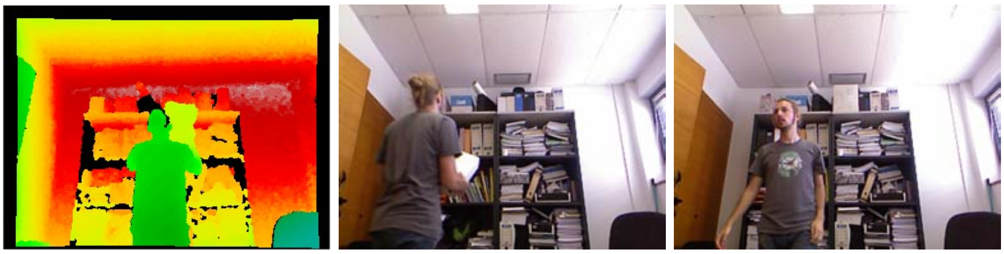


Fig. 4.14: **Color camouflage** sequence. Left frame depicts an example of depth frame for this sequence. Two next frames depict the start and the end of the sequence.

Depth camouflage

This is a 465 frame sequence used to evaluate the presented algorithm when depth camouflage occur (see Fig. 4.15). After 100 training frames a person appears and places a folder on an empty place on a shelf provoking new depth values similar to the old ones. The ground truth is composed by 12 frames that cover the most relevant part of the sequence where there are moving objects and depth camouflage occur.

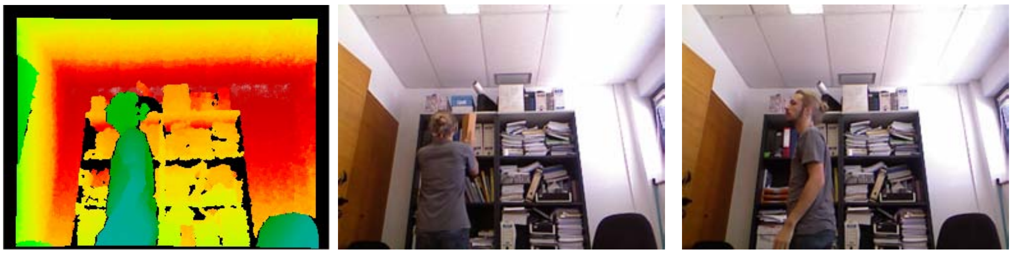


Fig. 4.15: **Depth camouflage** sequence. Left frame depicts an example of depth frame for this sequence. Two next frames depict the start and the end of the sequence.

Shadows

This 330 frame sequence created to evaluate the method's performance when shadows appears (see Fig. 4.16). In this example a person appears in the scene and moves his hand near a wall provoking shadow apparition. The ground truth is composed by 11 frames that cover the most relevant part of the sequence that are the ones where the moving hand and the shadow is present.

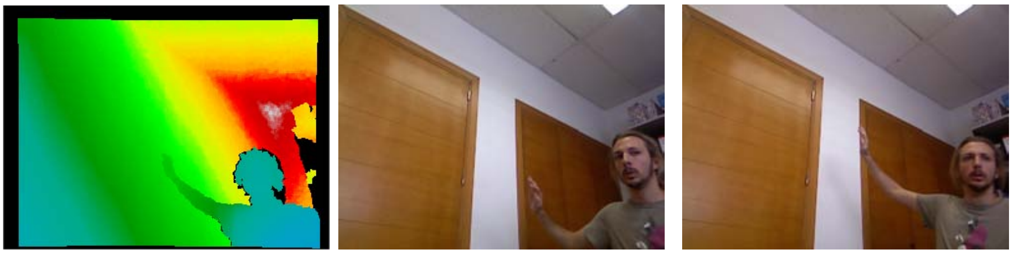


Fig. 4.16: Shadow sequence. Left frame depicts an example of depth frame for this sequence. Two next frames depict the start and the end of the sequence.

Light Switch

The sequence is used to evaluate sudden illumination changes in the scene. No moving object appears (see Fig. 4.17). After the training frames a lamp is turned on. *Light Switch* sequence has 407 frames. The ground truth is composed by 9 frames that cover the most relevant part of the sequence where the illumination changes are produced.

Fig. 4.17: Lightswitch sequence. Left frame depicts an example of depth frame for this sequence. Two next frames depict the start and the end of the sequence.

Bootstrapping

This is a 300 frame sequence used to evaluate the method's when there are moving objects in the training stage (see Fig. 4.18). In this example a person is moving during first 100 frames. This sequence is very challenging because in training stage we have the assumption that depth information is constant during all frames, so its possible to model wrong distributions that leads to misclassification.

Fig. 4.18: **Bootstrapping** sequence. Left frame depicts an example of depth frame for this sequence. Two next frames depict the start and the end of the sequence.

Waking Object

This is a 200 frame sequence used to evaluate the method's performance when an object initially in the background is moved (see Fig. 4.19). In this example a person appears in the scene and moves a chair during 100 frames. The ground truth is composed by 10 frames that cover the most relevant part of the sequence when chair starts moving. Unlike the other dataset we decided to create the ground truth images of this sequence because we think that qualitative comparison between algorithms are very useful in order to understand how they work and how we can improve them.

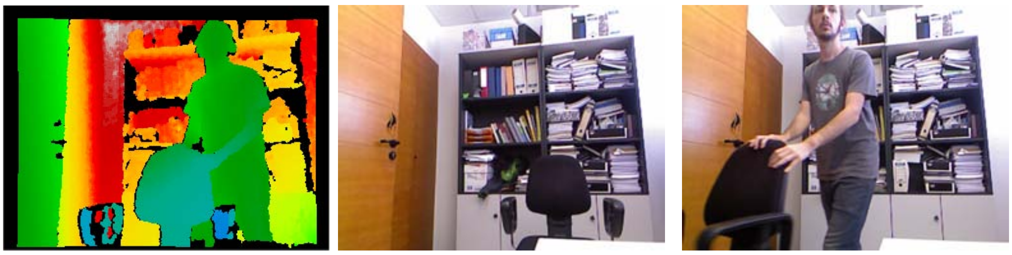


Fig. 4.19: **Waking object** sequence. Left frame depicts an example of depth frame for this sequence. Two next frames depict the start and the end of the sequence.

4.3.2 Metrics

In order to do an exhaustive and standard performance evaluation, we computed the performance measures using the framework proposed on CVPR 2014 *CDnet* challenge [48], which implements the following seven different measures: recall, specificity, false positive ratio (FPR), false negative ratio (FNR), percentage of wrong classifications (PWC), f-measure and precision. See Table 4.8 for details.

Measure	Description
Recall (Re)	Fraction of relevant instances that are retrieved, that is: $TP\ \ (TP+FN)$ a value $\in [0..1]$, the higher the better.
Speci city (Sp)	Proportion of negatives which are correctly identified as such: $TN\ \ (TN+FP)$ a value $\in [0..1]$, the higher the better.
False Positive Rate (FPR)	Probability of falsely rejecting the correct hypothesis: $FP\ \ (FP+TN)$ a value $\in [0..1]$, the lower the better.
False Negative Rate (FNR)	Probability of falsely accepting the invalid hypothesis: $FN\ \ (TP+FN)$ a value $\in [0..1]$, the lower the better.
Percentage of Wrong Classi cations (PWC)	Proportion of misclassified pixels: $100 * (FN + FP)\ \ (TP + FN + FP + TN)$ a value $\in [0..100]$, the lower the better.
F-Measure	Weighted harmonic mean of precision and recall: $(2 * Precision*Recall)\ \ (Precision+Recall)$ a value $\in [0..1]$, the higher the better.
Precision	Fraction of positive instances that are correctly classified $TP\ \ (TP+FP)$ a value $\in [0..1]$, the higher the better.
RM	Rank each method for each performance metric for one sequence.
RC	Global ranking of the algorithms across different sequences. That is, the mean of RM for each method across all sequences.

Tab. 4.8: Description of the metrics used to evaluate the algorithm's performance.

We evaluate our proposed algorithm using three different algorithms: ViBe [5], a mixture of Gaussian (MoG) implementation in Opencv library by Zivkovic [110] and the background subtraction algorithm [37] that uses a Gaussian kernel. Following the *CDnet* rules, each algorithm uses a single set of parameters.

4.3.3 Results

In this section, we analyse the performance of these algorithms for each sequence. As we can see in Tables 4.9 and 4.10, the addition of one geometrical dimension to our model permits to obtain a small advantage in the evaluation of color related issues as Time of Day and Light Switch. Fig. 4.20 illustrates that subtle changes in illumination, it is important to observe in result images, as we expected, we do not have regions marked as foreground.

	FN	FP	TP	TN	Re	Sp	FPR	FNR	PWC	F-Measure	Precision	RM
GSM_{UB}	0.00	0.19	0.00	99.81	0.00	1.00	0.00	0.00	0.19	0.00	0.00	1.00
GSM_{UF}	0.00	0.31	0.00	99.69	0.00	1.00	0.00	0.00	0.31	0.00	0.00	1.43
MOG [110]	0.00	0.82	0.00	99.18	0.00	0.99	0.01	0.00	0.82	0.00	0.00	2.29
ViBe [64]	0.00	2.92	0.00	97.08	0.00	0.97	0.03	0.00	2.92	0.00	0.00	2.71
KDE [36]	0.00	0.33	0.00	99.67	0.00	1.00	0.00	0.00	0.33	0.00	0.00	1.86

Tab. 4.9: Time of day sequence measurements. **FN**: False Negatives. **FP**: False Positives. **TP**: True Positives. **TN**: True Negatives. **Re**: Recall. **Sp**:Specificity. **FPR**: False Positive Rate. **FNR**:False Negative Rate. **PWC**: Percentage of Wrong Classifications.

LightSwitch	FN	FP	TP	TN	Re	Sp	FPR	FNR	PWC	F-Measure	Precision	RM
GSM_{UB}	0.00	0.11	0.00	99.89	0.00	1.00	0.00	0.00	0.11	0.00	0.00	1.00
GSM_{UF}	0.00	0.34	0.00	99.66	0.00	1.00	0.00	0.00	0.34	0.00	0.00	1.43
MOG [110]	0.00	2.60	0.00	97.40	0.00	0.97	0.03	0.00	2.60	0.00	0.00	1.86
ViBe [64]	0.00	4.70	0.00	95.30	0.00	0.95	0.05	0.00	4.70	0.00	0.00	2.71
KDE [36]	0.00	4.15	0.00	95.85	0.00	0.96	0.04	0.00	4.15	0.00	0.00	2.29

Tab. 4.10: Light switch sequence measurements. **FN**: False Negatives. **FP**: False Positives. **TP**: True Positives. **TN**: True Negatives. **Re**: Recall. **Sp**:Specificity. **FPR**: False Positive Rate. **FNR**:False Negative Rate. **PWC**: Percentage of Wrong Classifications.

Fig. 4.20: Results of **Time Of Day** sequence: Left column corresponds to color images used as reference, right column is the result of the algorithm considering undefined pixels as foreground (GSM_{UF}) and the column at the center is the result of the algorithm considering undefined pixels as background (GSM_{UB}).

The evaluation of colour camouflage situations is summarized in Table. 4.11, GSM algorithms improves the other algorithm's results, having best final classification. Fig. 4.21) depicts the results of applying the GSM algorithms in that sequences, the important area to observe is the final position of the box on the shelf.

	FN	FP	TP	TN	Re	Sp	FPR	FNR	PWC	F-Measure	Precision	RM
GSM$_{UB}$	5.88	0.68	126.71	93.99	0.96	0.99	0.01	0.04	2.89	3.85	0.99	2.29
GSM$_{UF}$	2.42	0.92	126.59	94.02	0.98	0.99	0.01	0.02	1.49	3.92	0.99	2.14
MOG [110]	42.09	0.69	117.45	91.90	0.74	0.99	0.01	0.26	16.97	3.19	0.99	4.29
ViBe [64]	33.25	0.26	120.23	91.98	0.78	1.00	0.00	0.22	13.64	3.34	1.00	3.29
KDE [36]	17.08	0.69	123.00	93.50	0.88	0.99	0.01	0.12	7.58	3.62	0.99	3.00

Tab. 4.11: Colour camouflage sequence measurements. **FN**: False Negatives. **FP**: False Positives. **TP**: True Positives. **TN**: True Negatives. **Re**: Recall. **Sp**:Specificity. **FPR**: False Positive Rate. **FNR**:False Negative Rate. **PWC**: Percentage of Wrong Classifications.

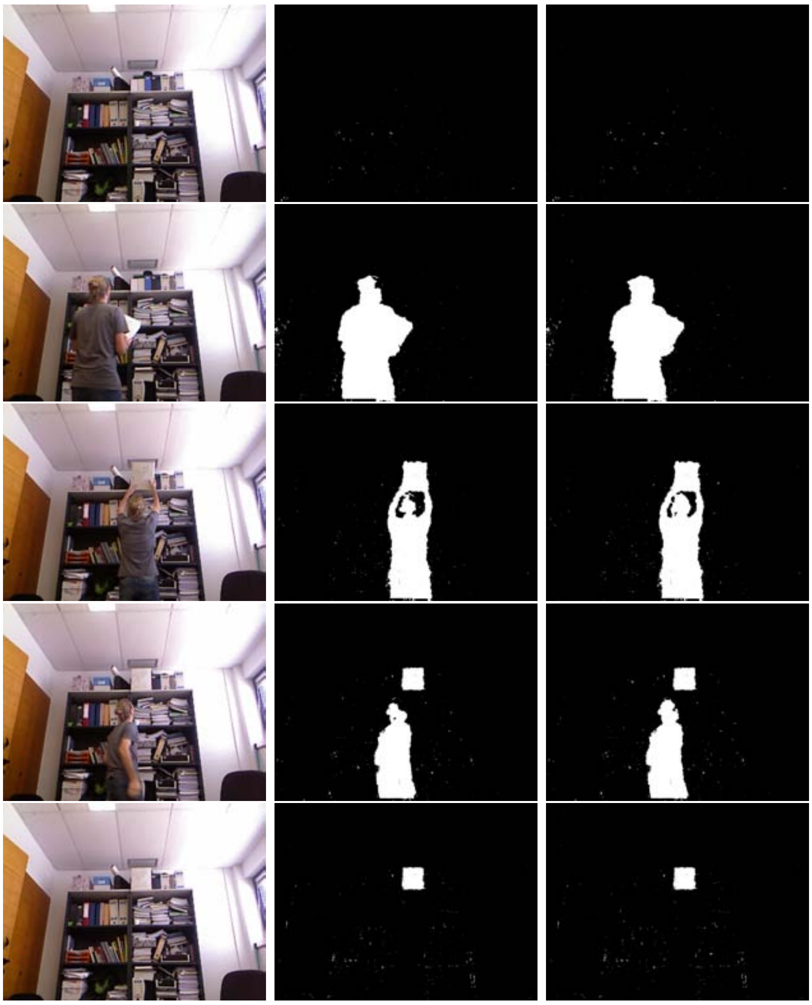


Fig. 4.21: Results of **Colour Camouflage** sequence: Left column corresponds to color images used as reference, right column is the result of the algorithm considering undefined pixels as foreground (GSM$_{UF}$) and the column at the center is the result of the algorithm considering undefined pixels as background (GSM$_{UB}$).

When depth camouflage occurs, GSM algorithms has best rank, due the low number of false negatives comparing to the other algorithms, see Table. 4.12 for details. It is important to notice the difference between GSM$_{UF}$ and GSM$_{UB}$ in this sequence, that is caused by the misclassification of certain parts of the body in first two images of the sequence (see Fig. 4.22) when undefined pixels appears. It is important to notice that results of Elgammal's approach are very similar as the ones of proposed algorithm, that situation is caused because in this sequence the important information is color information and we model that in a similar way. Finally, it is important to notice that the book is perfectly segmented in the last images of the sequence.

	FN	FP	TP	TN	Re	Sp	FPR	FNR	PWC	F-Measure	Precision	RM
GSM$_{UB}$	5.91	0.83	93.56	98.80	0.94	0.99	0.01	0.06	3.39	3.80	0.99	2.86
GSM$_{UF}$	2.82	1.15	94.69	98.76	0.97	1.00	0.00	0.03	0.00	3.88	0.99	1.00
MOG [110]	53.88	0.29	90.64	94.89	0.63	1.00	0.00	0.37	22.60	2.87	1.00	4.14
ViBe [64]	35.88	0.44	92.50	95.85	0.72	1.00	0.00	0.28	16.17	3.15	1.00	3.71
KDE [36]	8.07	0.70	94.46	98.67	0.92	0.99	0.01	0.08	4.34	3.74	0.99	3.29

Tab. 4.12: Depth camouflage sequence measurements. **FN**: False Negatives. **FP**: False Positives. **TP**: True Positives. **TN**: True Negatives. **Re**: Recall. **Sp**:Specificity. **FPR**: False Positive Rate. **FNR**:False Negative Rate. **PWC**: Percentage of Wrong Classifications.

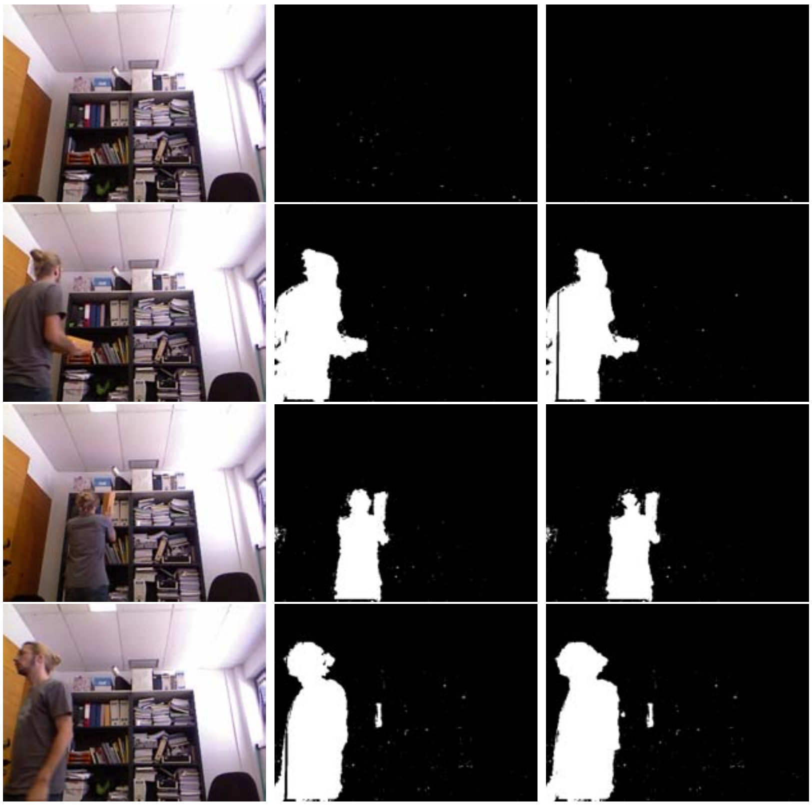


Fig. 4.22: Results of **Depth Camouflage** sequence: Left column corresponds to color images used as reference, right column is the result of the algorithm considering undefined pixels as foreground (GSM$_{UF}$) and the column at the center is the result of the algorithm considering undefined pixels as background (GSM$_{UB}$).

In case of shadow evaluation, GSM$_{UF}$ and GSM$_{UB}$ have the best classifications proving that adding depth information can help to avoid some color problems (see Table 4.13). Elgammal's algorithm has good results due to its special treatment of color information to avoid shadows. In Fig. 4.23 it is important to observe in color images how the hand approximates to the wall and a shadow appears and how our algorithm can avoid classifying that region as foreground.

	FN	FP	TP	TN	Re	Sp	FPR	FNR	PWC	F-Measure	Precision	RM
GSM$_{UB}$	2.98	0.27	79.39	96.82	0.96	1.00	0.00	0.04	1.81	3.88	1.00	2.29
GSM$_{UF}$	1.40	0.46	79.55	96.78	0.98	1.00	0.00	0.02	1.04	3.93	0.99	2.29
MOG [110]	29.55	0.19	67.56	95.24	0.70	1.00	0.00	0.30	15.44	3.08	1.00	3.86
ViBe [64]	25.00	0.43	70.22	95.35	0.74	1.00	0.00	0.26	13.32	3.20	0.99	3.57
KDE [36]	8.54	0.36	77.45	96.37	0.90	1.00	0.00	0.10	4.87	3.69	1.00	3.00

Tab. 4.13: Shadow sequence measurements. **FN**: False Negatives. **FP**: False Positives. **TP**: True Positives. **TN**: True Negatives. **Re**: Recall. **Sp**:Specificity. **FPR**: False Positive Rate. **FNR**:False Negative Rate. **PWC**: Percentage of Wrong Classifications.

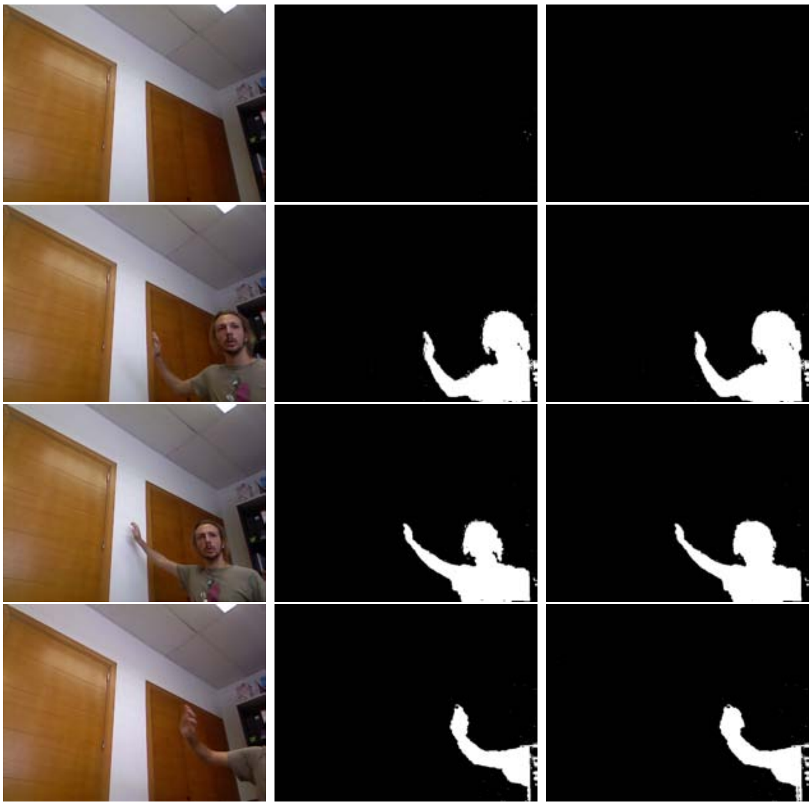


Fig. 4.23: Results of **Shadow** sequence: Left column corresponds to color images used as reference, right column is the result of the algorithm considering undefined pixels as foreground (GSM$_{UF}$) and the column at the center is the result of the algorithm considering undefined pixels as background (GSM$_{UB}$).

GSM have also the best results in Bootstrapping situations (see Table 4.14). This sequence is very challenging because in training stage, we have the assumption that depth information is constant during all frames, so its possible to model wrong distributions that leads to misclassification, see Fig. 4.24 for a visual example.

BootStraping	FN	FP	TP	TN	Re	Sp	FPR	FNR	PWC	F-Measure	Precision	RM
GSM$_{UB}$	9.05	0.42	26.15	100.78	0.74	1.00	0.00	0.26	6.94	3.19	0.98	2.43
GSM$_{UF}$	4.64	0.56	26.25	101.63	0.85	0.99	0.01	0.15	3.91	3.49	0.98	2.00
MOG [110]	27.12	0.20	24.02	100.20	0.47	1.00	0.00	0.53	18.03	2.39	0.99	3.57
ViBe [64]	16.98	12.69	24.29	88.22	0.59	0.87	0.13	0.41	20.87	2.02	0.66	3.86
Elgammal	21.92	0.38	25.27	99.63	0.54	1.00	0.00	0.46	15.15	2.58	0.99	3.14

Tab. 4.14: Bootstrapping sequence measurements. **FN**: False Negatives. **FP**: False Positives. **TP**: True Positives. **TN**: True Negatives. **Re**: Recall. **Sp**:Specificity. **FPR**: False Positive Rate. **FNR**:False Negative Rate. **PWC**: Percentage of Wrong Classifications.

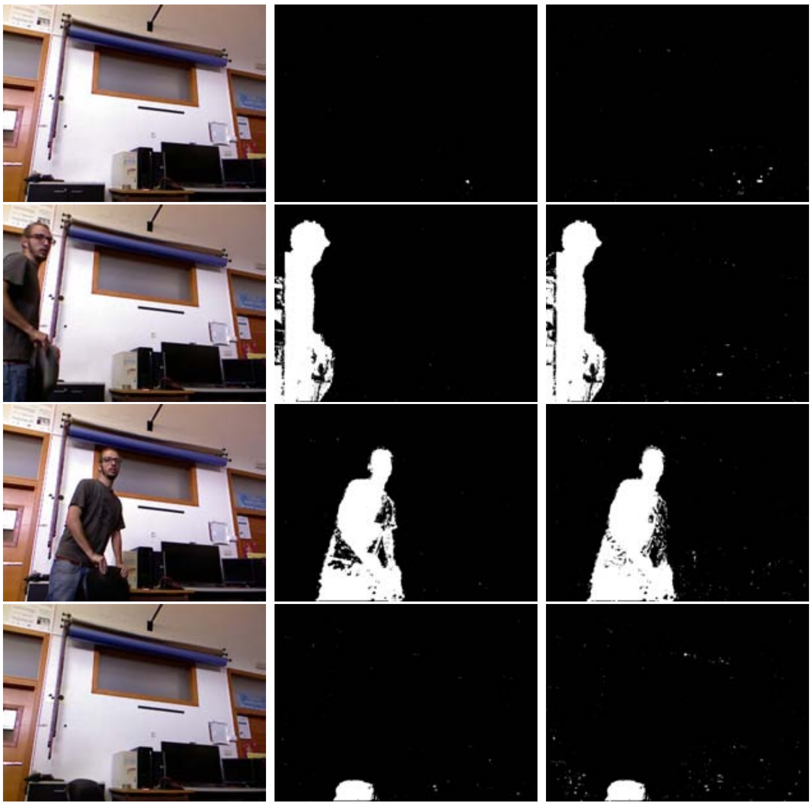


Fig. 4.24: Results of **Bootstrapping** sequence: Left column corresponds to color images used as reference, right column is the result of the algorithm considering undefined pixels as foreground (GSM$_{UF}$) and the column at the center is the result of the algorithm considering undefined pixels as background (GSM$_{UB}$).

The waking object sequence allows us to test the background object moving detection obtaining the expected results. Results from GSM$_{UF}$ are better than GSM$_{UB}$, as it is shown in Table 4.14, that occur because in the last part of the sequence the user is near the sensor provoking the apparition of a big region with ADO-pixels, and consequently provokes misclassification. Fig. 4.25 depicts the results of the GSM algorithms with the Bootstrapping sequence.

	FN	FP	TP	TN	Re	Sp	FPR	FNR	PWC	F-Measure	Precision	RM
GSM_{UB}	17.77	1.49	74.83	91.24	0.81	0.98	0.02	0.19	10.39	3.37	0.98	3.71
GSM_{UF}	3.27	3.85	77.15	94.36	0.96	0.96	0.04	0.04	3.98	3.74	0.95	2.29
MOG [110]	14.26	2.47	73.41	92.68	0.84	0.97	0.03	0.16	9.15	3.43	0.97	3.29
ViBe [64]	15.59	7.27	72176.00	87.37	1.00	0.92	0.08	0.00	0.03	4.00	1.00	2.43
KDE [36]	5.97	9.61	74.65	88.29	0.93	0.90	0.10	0.07	8.72	3.45	0.89	3.29

Tab. 4.15: Waking object sequence measurements. **FN**: False Negatives. **FP**: False Positives. **TP**: True Positives. **TN**: True Negatives. **Re**: Recall. **Sp**:Specificity. **FPR**: False Positive Rate. **FNR**:False Negative Rate. **PWC**: Percentage of Wrong Classifications.

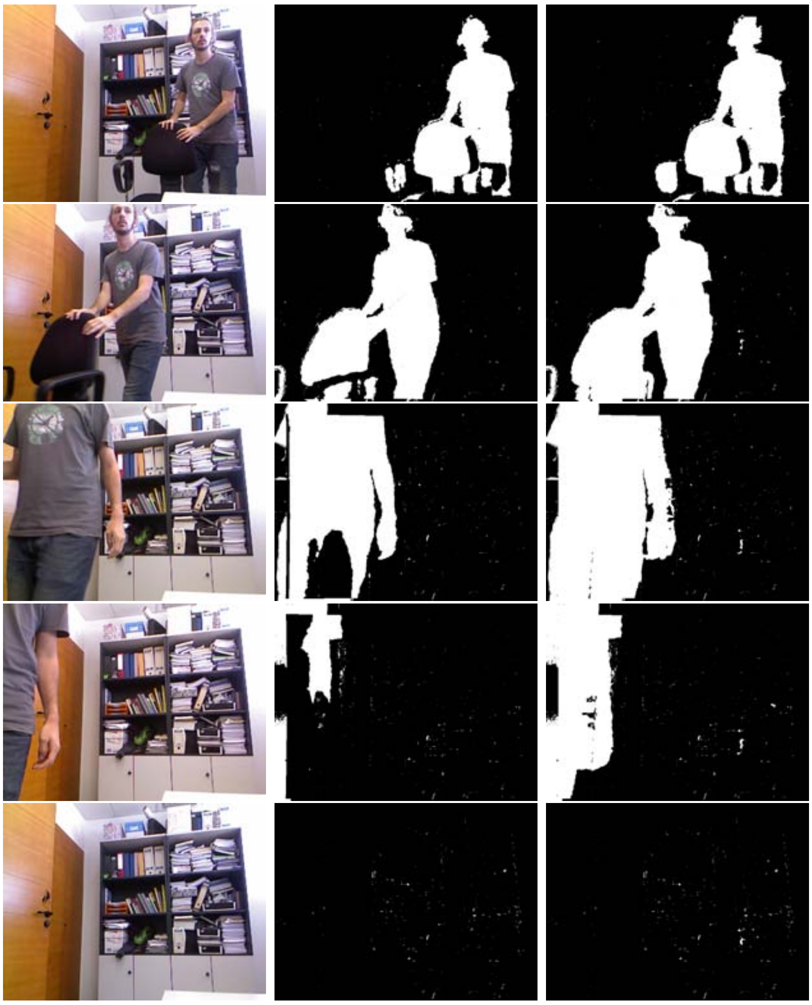


Fig. 4.25: Results of **Waking Object** sequence: Left column corresponds to color images used as reference, right column is the result of the algorithm considering undefined pixels as foreground (GSM_{UF}) and the column at the center is the result of the algorithm considering undefined pixels as background (GSM_{UB}).

In order to understand the global results, they have been summarized in Table. 4.16. A ranking of the tested algorithms is computed (RC) starting from the partial ranks on these measures (RM), see Fig. 4.26 for a visual representation of results. GSM algorithm has good results when we test different color changeling situations. GSM$_{UB}$ and GSM$_{UF}$ prove to be the most stable in sudden illumination changes with significant difference to the other algorithms due the invariance of depth information. As expected, when depth information is relevant, GSM algorithm have best scores in all cases. In Fig. 4.26 global results are summarized in a visual way.

Sequence	Waking Object	Time of Day	Colour Camouflage	Depth Camouflage	Shadows	Light Switch	Bootstrapping	RC
GSM-UB	3	**1**	**2**	*2.857*	**2**	**1**	*2.429*	*2.457*
GSM-UF	**1.857**	*1.429*	*2.714*	**2.714**	2.571	*1.429*	**2**	**2.429**
MOG [110]	*2.571*	2.286	4.571	3.286	3.286	1.857	3.571	3.629
ViBe [64]	4.286	2.714	*2.714*	3.143	4.143	2.714	3.857	3.886
KDE [36]	3.286	1.857	3	3	3	*2.286*	3.143	3.314

Tab. 4.16: Evaluation results for all algorithms averaged over all sequences. Last column shows final average ranking. Bold entries indicate the best result and italics the second one.

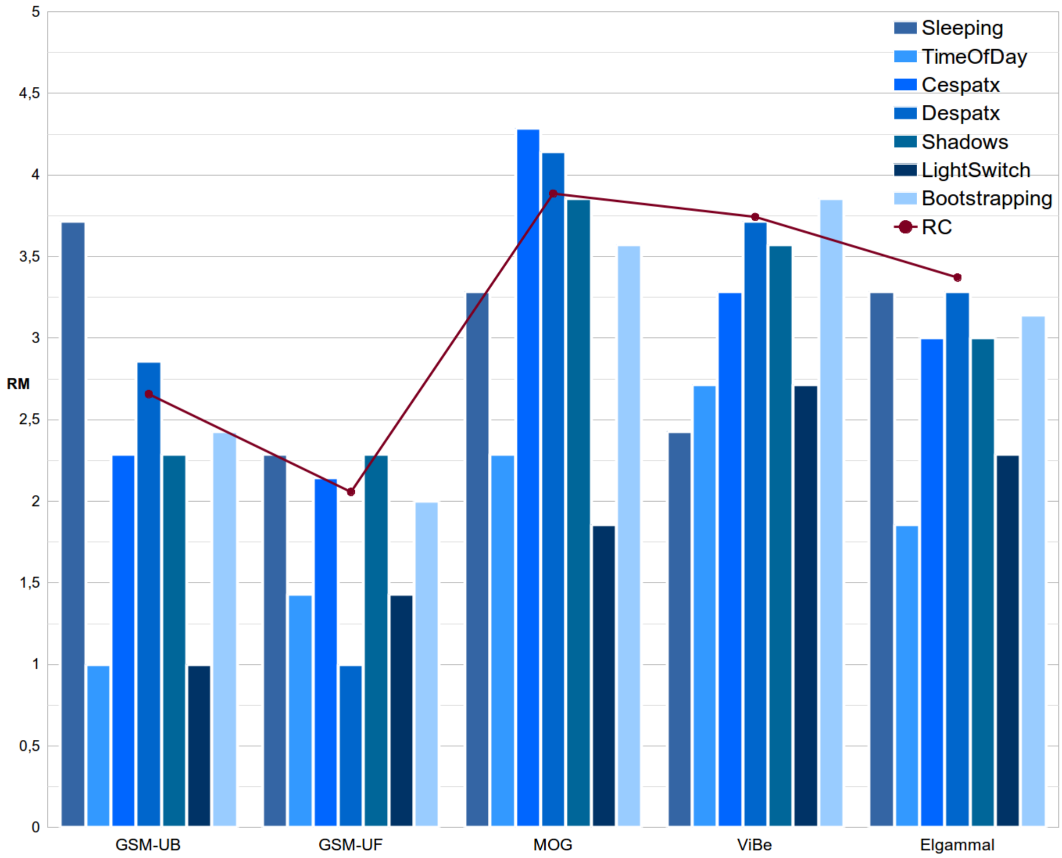


Fig. 4.26: GSM dataset simulation results. It can be found results of seven different sequences for all tested algorithms and final comparisons. As GSM$_{UF}$ and GSM$_{UB}$ are not the best for each sequence, are the most regular ones as it can be seen with the **RC** line, the lower the better.

In order to make available the research done in scene modelling, allow future work based on the GSM algorithm and future comparisons. The presented dataset, the used implementation and the results obtained in this chapter are available at `gsm.uib.es`.

5 Conclusions

5.1 Conclusions

We have developed an interactive system to improve balance and postural control through rehabilitation. The system exercises are focused on center-of-mass movements induced by interacting with virtual objects shown on-screen. The serious game was designed using the prototype development paradigm and features for rehabilitation with serious games: feedback, adaptability, motivational elements and monitoring. The employed interaction technology is based on computer vision because balance rehabilitation consists of body movements that can be recorded using these vision systems and because users can have difficulties holding physical devices.

Through the described system we experimented if serious games for rehabilitation can be used for motivational balance rehabilitation in cerebral palsy patients. Results show that users improved their balance slowly; improvements were also detected in individual items. With regards to motivation, in previous years the set of users had abandoned their therapeutic plans. Using the presented system, no users abandoned and they showed interested in continuing the rehabilitation process with the video games.

We experimented mirror feedback by means of an user study, mirror feedback can be defined as the visual representation of the users inside the application such as interaction feedback. We observed that users with cognitive impairment had bigger differences between feedback conditions. The higher cognitive impairment user had, the more important feedback was in order to perform correctly the therapy. Results confirmed our hypothesis that in case of disabilities the mirror feedback mechanisms facilitated the interaction in the vision-based systems for rehabilitation.

We evaluated our interactive system during 24 weeks. The results suggest that it is feasible to use interactive serious games as an intervention for adults with cerebral palsy to address motivation in rehabilitation therapy and thereby enhance balance and gait motor performance.

From our work implementing vision-based serious games for motor rehabilitation, we have presented implementations guidelines for developing vision-based motor-rehabilitation serious games, in order to help others researchers in this field.

During the development of the experimental system, we noticed that users sometimes were distracted by what was happening around them. To keep the user centred on the game, background subtraction was applied, so we could change the real background and put another one adapted to the user needs. For that reason, we developed a new scene modelling approach, GSM, that uses both depth and color information in a unified way. We constructed a background model for each pixel of the scene and estimates the probability that a newly observed pixel value belongs to that model. These probabilities are estimated independently for each new frame.

We constructed our model using a KDE process, with a Gaussian Kernel. In order to construct only one model, we used a three dimensional kernel, one dimension to model depth information and two for normalised chromaticity coordinates. We modelled sensor *absent depth observations* using a probabilistic strategy in order to distinguish which belongs to the background model and which are provoked by foreground objects in order to detected those ones that are induced by foreground objects. That pixels cannot be classified as background or foreground so we used a third classification class, we called undefined, in order to classify that pixels. In addition, we developed an algorithm to detect changing background objects in the same frame they move based on the cdf of the pixel model. Two strategies are described in order to adapt the update phase to the different nature of the color and depth information, considering color information as short-term model and depth as long-term one. Results show that the presented algorithm is the most regular one, having good results in a wide range of situations, solving the issues of the depth data sensors. That means, it can handle with many different situations. We can conclude that by the combination of two kind of information in a 3D kernel helps to achieve better modelling algorithms.

Next step in our research is the development of vision-based applications for rehabilitation for large population groups, such as elderly people. The European ageing population is in fast grow. According to the World Bank, 18 % of the population in Spain is over 65 years. Among the elderly people, falls can lead to a major social problem, it influence significantly in life expectancy and lead to subsequent clinical problems with high social and health costs. The loss of autonomy is the most prominent consequence. Different studies have demonstrated the importance of falls prevention, and its effectiveness when specific programs of physiotherapy are used to improve balance [35, 70, 86]. In order to generalize the use of these systems, the presented scene modelling algorithm, GSM, is used in order to obtain environmental independence [3]. These applications can be used at the patients' homes so that they do not have to displace to a hospital or a rehabilitation center, and can therefore devote more time to their rehabilitation.

In this sense, our research group has developed some experimental systems in order to validate the possibility of automatically evaluate the FRT [4, 3]. FRT is one of the most used tests for measuring balance clinically because it measures the limits of stability while standing.

Bibliography

[1]G. Alankus, A. Lazar, M. May, and C. Kelleher. „Towards customizable games for stroke rehabilitation“. In: *Proc. 28th Int l ACM Conf. Human factors in computing systems (CHI 10)*. Atlanta, Georgia, USA, 2010, pp. 2113–2122 (cit. on pp. 9, 12, 39).

[2]F. Anderson, M. Annett, and W.F. Bischof. „Lean on Wii: Physical rehabilitation with virtual reality and Wii peripherals“. In: *Annual Review of CyberTherapy and Telemedicine* 8 (2010), pp. 181–184 (cit. on pp. 9, 10, 12).

[3]Ines Ayed, Gabriel Moyà-Alcover, Pau Martínez-Bueso, et al. „RGBD-based Serious Games for Fall Prevention in Elderly People“. In: *Cognitive Area Networks , vol. 1, no 3*. 2016 (cit. on p. 91).

[4]Ines Ayed, Gabriel Moyà-Alcover, Pau Martínez-Bueso, et al. „Validación de dispositivos RGBD para medir terapéuticamente el equilibrio: el test de alcance funcional con Microsoft Kinect“. In: *Revista Iberoamericana de Automática e Informática Industrial* (2016). In press (cit. on p. 91).

[5]Olivier Barnich and Marc Van Droogenbroeck. „ViBe: a universal background subtraction algorithm for video sequences.“ In: *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society* 20.6 (June 2011), pp. 1709–24 (cit. on pp. 44, 79).

[6]N.A. Bartolomé, A.M. Zorrilla, and B.G. Zapirain. „Can game-based therapies be trusted? Is game-based education effective? A systematic review of the Serious Games for health and education“. In: *Proc. 16th Int l IEEE Conf. Computer Games (CGAMES 11)*. Louisville, KY, USA, 2011, pp. 275–282 (cit. on p. 10).

[7]Yannick Benezeth and PM Jodoin. „Comparative study of background subtraction algorithms“. In: *Journal of Electronic Imaging 19.3 (2010): 033003-033003* (2010) (cit. on p. 43).

[8]K. Berg and K.E. Norman. „Functional assessment of balance and gait“. In: *Clin. Geriatr. Med.* 12.4 (1996), pp. 705–723 (cit. on p. 32).

[9]A.L. Betker, A. Desai, C. Nett, N. Kapadia, and T. Szturm. „Game-based exercises for dynamic short-sitting balance rehabilitation of people with chronic spinal cord and traumatic brain injuries“. In: *Phys. Ther.* 87.10 (2007), pp. 1389–1398 (cit. on pp. 9, 10).

[10] Nigel Bevan. „Measuring usability as quality of use“. English. In: *Software Quality Journal* 4.2 (1995), pp. 115–130 (cit. on p. 25).

[11] Bruno Bonnechere, Bart Jansen, P Salvia, et al. „Validity and reliability of the Kinect within functional assessment activities: Comparison with standard stereophotogrammetry“. In: *Gait & posture* 39.1 (2014), pp. 593–598 (cit. on p. 10).

[12] Thierry Bouwmans. „Traditional and recent approaches in background modeling for foreground detection: An overview“. In: *Computer Science Review* 11-12 (May 2014), pp. 31–66 (cit. on pp. 41, 44).

[13] Thierry Bouwmans, Fatih Porikli, Benjamin Höferlin, and Antoine Vacavant. *Background Modeling and Foreground Detection for Video Surveillance*. CRC Press, 2014 (cit. on pp. 41, 43).

[14] G.R. Bradski. „Computer vision face tracking for use in a perceptual user interface“. In: *Intel Technology Journal* 2.2 (1998), pp. 1–15 (cit. on p. 39).

[15] Doris Broetz and Niels Birbaumer. „Behavioral physiotherapy in post stroke rehabilitation“. In: *NeuroRehabilitation* 33.3 (2013), pp. 377–384 (cit. on pp. 4, 11).

[16] J.W. Burke, M.D.J. McNeill, D.K. Charles, et al. „Optimising engagement for stroke rehabilitation using serious games“. In: *Vis. Comput.* 25.12 (2009), pp. 1085–1099 (cit. on pp. 2, 3, 9–12, 14–16, 38).

[17] M.S. Cameirao, I.B.S. Bermudez, E. Duarte Oller, and PF Verschure. „The rehabilitation gaming system: a review“. In: *Stud. Health Technol. Inform.* 145 (2009), pp. 65–83 (cit. on pp. 9, 12).

[18] Massimo Camplani and Luis Salgado. „Background foreground segmentation with RGB-D Kinect data: An efficient combination of classifiers“. In: *Journal of Visual Communication and Image Representation* 25.1 (Jan. 2013), pp. 122–136 (cit. on pp. 45, 47, 48, 64, 68).

[19] Massimo Camplani, Carlos R. del Blanco, Luis Salgado, Fernando Jaureguizar, and Narciso García. „Multi-sensor background subtraction by fusing multiple region-based probabilistic classifiers“. In: *Pattern Recognition Letters* 50 (Dec. 2014), pp. 23–33 (cit. on pp. 46, 47).

[20] Antoni Jaume-i Capó, Javier Varona, Manuel González-Hidalgo, and Francisco J Perales. „Adding image constraints to inverse kinematics for human motion capture“. In: *EURASIP Journal on Advances in Signal Processing* 2010.1 (2009), pp. 1–13 (cit. on p. 2).

[21] Brenda L. Chan, Richard Witt, Alexandra P. Charrow, et al. „Mirror Therapy for Phantom Limb Pain“. In: *New England Journal of Medicine* 357.21 (2007). PMID: 18032777, pp. 2206–2207. eprint: `http://www.nejm.org/doi/pdf/10.1056/NEJMc071927` (cit. on p. 11).

[22] Yao-Jen Chang, Shu-Fang Chen, and Jun-Da Huang. „A Kinect-based system for physical rehabilitation: A pilot study for young adults with motor disabilities“. In: *Research in Developmental Disabilities* 32.6 (2011), pp. 2566 –2570 (cit. on pp. 10, 11).

[23] Sangwoo Cho, Jeonghun Ku, Yun Kyung Cho, et al. „Development of virtual reality proprioceptive rehabilitation system for stroke patients“. In: *Computer methods and programs in biomedicine* 113.1 (2014), pp. 258–265 (cit. on pp. 4, 11).

[24] Albert Clapés, Miguel Reyes, and Sergio Escalera. „Multi-modal user identification and object recognition surveillance system“. In: *Pattern Recognition Letters* 34.7 (May 2013), pp. 799–808 (cit. on pp. 44, 47).

[25] Ross A Clark, Yong-Hao Pua, Karine Fortin, et al. „Validity of the Microsoft Kinect for assessment of postural control“. In: *Gait & Posture* 36.3 (2012), pp. 372–377 (cit. on p. 10).

[26] Ross A Clark, Yong-Hao Pua, Adam L Bryant, and Michael A Hunt. „Validity of the Microsoft Kinect for providing lateral trunk lean feedback during gait retraining“. In: *Gait & posture* (2013) (cit. on p. 10).

[27] José M Cogollor, Charmayne Hughes, Manuel Ferre, et al. „Handmade Task Tracking Applied to Cognitive Rehabilitation“. In: *Sensors* 12.10 (2012), pp. 14214–14231 (cit. on p. 10).

[28] Roberto Colombo, Fabrizio Pisano, Alessandra Mazzone, et al. „Design strategies to improve patient motivation during robot-aided rehabilitation“. In: *J. Neuroeng. Rehabil.* 4.3 (2007), pp. 1–12 (cit. on pp. 9, 12).

[29] Marco Cristani, Michela Farenzena, Domenico Bloisi, and Vittorio Murino. „Background Subtraction for Automated Multisensor Surveillance: A Comprehensive Review“. In: *EURASIP Journal on Advances in Signal Processing* 2010 (2010), pp. 1–24 (cit. on pp. 42, 43, 56).

[30] A. Da Gama, T. Chaves, L. Figueiredo, and V. Teichrieb. „Guidance and Movement Correction Based on Therapeutics Movements for Motor Rehabilitation Support Systems“. In: *Virtual and Augmented Reality (SVR), 2012 14th Symposium on*. 2012, pp. 191–200 (cit. on pp. 10, 11).

[31] J.E. Deutsch, M. Borbely, J. Filler, K. Huhn, and P. Guarrera-Bowlby. „Use of a low-cost, commercially available gaming console (Wii) for rehabilitation of an adolescent with cerebral palsy“. In: *Physical Therapy* 88.10 (2008), p. 1196 (cit. on p. 10).

[32] RO Duda, PE Hart, and DG Stork. „Pattern classification. 2nd edn Wiley“. In: *New York* (2000), p. 153 (cit. on p. 52).

[33] P.W. Duncan, D.K. Weiner, J. Chandler, and S. Studenski. „Functional reach: a new clinical measure of balance“. In: *J. Gerontol.* 45.6 (1990), pp. M192–M197 (cit. on p. 32).

[34] J. Edmans, J. Gladman, M. Walker, et al. „Mixed reality environments in stroke rehabilitation: Development as rehabilitation tools“. In: *Int. J. Disabil. Hum. Dev.* 6.1 (2006), pp. 39–46 (cit. on p. 2).

[35] Steriani Elavsky, Edward McAuley, Robert W Motl, et al. „Physical activity enhances long-term quality of life in older adults: efficacy, esteem, and affective influences“. In: *Annals of Behavioral Medicine* 30.2 (2005), pp. 138–145 (cit. on p. 90).

[36] Ahmed Elgammal. „Background subtraction: Theory and practice“. In: *Synthesis Lectures on Computer Vision* 5.1 (2014), pp. 1–83 (cit. on pp. 41, 43, 80–83, 85, 86).

[37] Ahmed Elgammal, David Harwood, and Larry Davis. „Non-parametric model for background subtraction“. In: *Computer Vision—ECCV 2000* (2000) (cit. on pp. 42, 52, 53, 59, 79).

[38] K. Eng, E. Siekierka, P. Pyk, et al. „Interactive visuo-motor therapy system for stroke rehabilitation“. In: *Medical and Biological Engineering and Computing* 45.9 (2007), pp. 901–907 (cit. on pp. 4, 11).

[39] Péter Fankhauser, Michael Bloesch, Diego Rodriguez, et al. „Kinect v2 for Mobile Robot Navigation: Evaluation and Modeling“. In: *Advanced Robotics (ICAR), 2015 International Conference on* (2015) (cit. on p. 48).

[40] Enrique J Fernandez-Sanchez, Javier Diaz, and Eduardo Ros. „Background subtraction based on color and depth using active sensors.“ In: *Sensors (Basel, Switzerland)* 13.7 (Jan. 2013), pp. 8895–915 (cit. on pp. 44, 46, 47).

[41] E. Flores, G. Tobon, E. Cavallaro, et al. „Improving patient motivation in game development for motor deficit rehabilitation“. In: *Proc. Int l ACM Conf. Advances in Computer Entertainment Technology (ACE 08)*. New York, NY, USA, 2008, pp. 381–384 (cit. on pp. 3, 9).

[42] S. Flynn, P. Palma, and A. Bender. „Feasibility of using the Sony PlayStation 2 gaming platform for an individual poststroke: a case report“. In: *J. Neurol. Phys. Ther.* 31.4 (2007), pp. 180–189 (cit. on pp. 10, 11).

[43] J. Gallego. „Parametric Region-Based Foreground Segmentation in Planar and Multi-View Sequences“. phd. Universitat Politècnica de Catalunya (UPC), 2013 (cit. on p. 46).

[44] Jaime Gallego and Montse Pardàs. „Region based foreground segmentation combining color and depth sensors via logarithmic opinion pool decision“. In: *Journal of visual communication and image representation* 25.1 (2014), pp. 184–194 (cit. on pp. 45, 47).

[45] Daniela Giordano, Simone Palazzo, and Concetto Spampinato. „Kernel Density Estimation Using Joint Spatial-Color-Depth Data for Background Modeling“. In: *2014 22nd International Conference on Pattern Recognition (ICPR)*. IEEE. 2014, pp. 4388–4393 (cit. on pp. 45, 47).

[46] G Gordon and T Darrell. „Background estimation and removal based on range and color“. In: *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.. Vol. 2. IEEE* Vol. 2.June (1999) (cit. on pp. 44, 47, 68).

[47] D. Goude, S. Björk, M. Rydmark, et al. „Game design in virtual reality systems for stroke rehabilitation“. In: *Stud. Health Technol. Inform.* 125.2007 (2007), pp. 146–148 (cit. on pp. 2, 10, 11).

[48] N. Goyette, P. Jodoin, F. Porikli, J. Konrad, and P. Ishwar. „Changedetection.net: A new change detection benchmark dataset“. In: *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*. 2012, pp. 1–8 (cit. on pp. 42, 78).

[49] Klaus Greff, A Brandão, S Krauß, Didier Stricker, and Esteban Clua. „A Comparison Between Background Subtraction Algorithms using Depth Images with a Consumer Camera“. In: *VISAPP* (2012), pp. 431–436 (cit. on p. 55).

[50] Marco Iacoboni and John C Mazziotta. „Mirror neuron system: basic findings and clinical applications“. In: *Annals of neurology* 62.3 (2007), pp. 213–218 (cit. on pp. 4, 11).

[51] D. Jack, R. Boian, A.S. Merians, et al. „Virtual reality-enhanced stroke rehabilitation“. In: *IEEE Trans. Neural Syst. Rehabil. Eng.* 9.3 (2001), pp. 308–318 (cit. on pp. 9, 12).

[52] A. Jaume-i-Capó, P. Martinez-Bueso, B. Moya-Alcover, and J. Varona. „Interactive Rehabilitation System for Improvement of Balance Therapies in People With Cerebral Palsy“. In: *Neural Systems and Rehabilitation Engineering, IEEE Transactions on* 22.2 (2013), pp. 1–9 (cit. on pp. 3, 10, 22, 25).

[53] Antoni Jaume-i-Capó, Biel Moyà-Alcover, Javier Varona, Pau Martinez-Bueso, and Alejandro Mesejo Chiong. „MOTIVATIONAL REHABILITATION USING VISION-BASED SERIOUS GAMES“. In: *Biomedical Engineering* 764.1 (2013), pp. 62–70 (cit. on pp. 37, 38).

[54] Y. Jung, S.C. Yeh, and J. Stewart. „Tailoring virtual reality technology for stroke rehabilitation: a human factors design“. In: *CHI 06 extended abstracts on Human factors in computing systems*. ACM. 2006, pp. 929–934 (cit. on pp. 3, 10, 12, 38, 39).

[55] Kourosh Khoshelham and Sander Oude Elberink. „Accuracy and resolution of Kinect depth data for indoor mapping applications.“ In: *Sensors (Basel, Switzerland)* 12.2 (Jan. 2012), pp. 1437–54 (cit. on pp. 48, 63).

[56] K. Kim, Chalidabhongse, T. H., D. Harwood, and L Davis. „Background modeling and subtraction by codebook construction“. In: *Image Processing, ICIP 04. 2004 International Conference on. Vol. 5. IEEE* (2004), pp. 2–5 (cit. on p. 44).

[57] L.A. King and F.B. Horak. „Delaying mobility disability in people with Parkinson disease using a sensorimotor agility exercise program“. In: *Phys. Ther.* 89.4 (2009), pp. 384–393 (cit. on p. 13).

[58] V. Kolmogorov, A. Criminisi, A. Blake, G. Cross, and C. Rother. „Bi-Layer Segmentation of Binocular Stereo Video“. In: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 05)* 2 (2005), pp. 407–414 (cit. on p. 44).

[59] AN Krichevets, EB Sirotkina, IV Yevsevicheva, and LM Zeldin. „Computer games as a means of movement rehabilitation“. In: *Disabil. Rehabil.* 17.2 (1995), pp. 100–105 (cit. on pp. 3, 9).

[60] K.W. Krigger. „Cerebral palsy: an overview“. In: *Am. Fam. Physician* 73.1 (2006), pp. 91–100 (cit. on p. 1).

[61] Belinda Lange, Sebastian Koenig, Eric McConnell, et al. „Interactive game-based rehabilitation using the Microsoft Kinect“. In: *Proceedings of the 2012 IEEE Virtual Reality*. VR ’12. Washington, DC, USA: IEEE Computer Society, 2012, pp. 171–172 (cit. on pp. 10, 11).

[62] Benjamin Langmann, Seyed E Ghobadi, Klaus Hartmann, and Otmar Loffeld. „Multi-modal background subtraction using gaussian mixture models“. In: *ISPRS technical commission III symposium on photogrammetry computer vision and image analysis (PCV 2010)*. 2010, pp. 61–66 (cit. on pp. 44, 47).

[63] Kate Laver, Stacey George, Susie Thomas, Judith E Deutsch, and Maria Crotty. „Virtual reality for stroke rehabilitation“. In: *Stroke* 43.2 (2012), e20–e21 (cit. on p. 10).

[64] Jérôme Leens, Sébastien Piérard, Olivier Barnich, Marc Van Droogenbroeck, and Jean-Marc Wagner. *Combining color, depth, and motion for video segmentation*. 2009, pp. 104–113 (cit. on pp. 44, 47, 68, 80–86).

[65] Martin D Levine and Martin D Levine. *Vision in man and machine*. Vol. 574. McGraw-Hill New York, 1985 (cit. on p. 64).

[66] M. Ma and K. Bechkoum. „Serious games for movement therapy after stroke“. In: *Proc. IEEE Int l Conf. Systems, Man and Cybernetics (SMC 08)*. Derby, UK, 2008, pp. 1872–1877 (cit. on pp. 9, 12).

[67] Niall Maclean, Pandora Pound, Charles Wolfe, and Anthony Rudd. „The concept of patient motivation a qualitative analysis of stroke professionals’ attitudes“. In: *Stroke* 33.2 (2002), pp. 444–448 (cit. on p. 2).

[68] L. Maddalena and A. Petrosino. „A Self-Organizing Approach to Background Subtraction for Visual Surveillance Applications“. In: *IEEE Transactions on Image Processing* 17.7 (2008), pp. 1168–1177 (cit. on p. 69).

[69] A Matthew. „A Scoping Review of Exergaming for Adults with Systemic Disabling Conditions“. In: *Journal of Bioengineering & Biomedical Science* (2011) (cit. on p. 10).

[70] David H Metz. „Mobility of older people and their quality of life“. In: *Transport policy* 7.2 (2000), pp. 149–152 (cit. on p. 91).

[71] D.R. Michael and S.L. Chen. *Serious games: Games that educate, train, and inform*. Mason, OH, USA: Cenage Learning, 2005 (cit. on p. 3).

[72] B. Moyà-Alcover, A. Jaume-i-Capó, J. Varona, P. Martinez-Bueso, and A. Mesejo Chiong. „Use of serious games for motivational balance rehabilitation of cerebral palsy patients“. In: *The proceedings of the 13th international ACM SIGACCESS conference on Computers and Accessibility*. ASSETS ’11. ACM. Dundee, Scotland, UK, 2011, pp. 297–298 (cit. on pp. 9, 37).

[73] Eric B Nash, Gregory W Edwards, Jennifer A Thompson, and Woodrow Barfield. „A review of presence and performance in virtual environments“. In: *International Journal of human-computer Interaction* 12.1 (2000), pp. 1–41 (cit. on p. 11).

[74] L. del Nogal, A. González, and P. Lloro. „Evaluación del riesgo de caídas. Protocolos de valoración clínica“. In: *Rev. Esp. Geriatr. Gerontol.* 40.1 (2005), pp. 54–63 (cit. on p. 33).

[75] D.A. Norman. *The invisible computer: why good products can fail, the personal computer is so complex, and information appliances are the solution*. MIT press, 1999 (cit. on p. 37).

[76] Donald A. Norman. *The invisible computer*. Cambridge, MA, USA: MIT Press, 1998 (cit. on pp. 3, 9).

[77] Stepan Obdrzalek, Gregorij Kurillo, Ferda Ofli, et al. „Accuracy and robustness of Kinect pose estimation in the context of coaching of elderly population“. In: *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*. IEEE. 2012, pp. 1188–1193 (cit. on p. 10).

[78] Simona Ottonelli, Paolo Spagnolo, Pier Luigi Mazzeo, and Marco Leo. „Foreground Segmentation by Combining Color and Depth Images“. In: *Pattern Recognition and Image Analysis*. Springer, 2013, pp. 699–706 (cit. on pp. 45, 47).

[79] Simona Ottonelli, Paolo Spagnolo, Pier Luigi Mazzeo, and Marco Leo. „Improved video segmentation with color and depth using a stereo camera“. In: *Industrial Technology (ICIT), 2013 IEEE International Conference on*. IEEE. 2013, pp. 1134–1139 (cit. on pp. 45, 47).

[80] S.P. Overmyer. „Revolutionary vs. evolutionary rapid prototyping: Balancing software productivity and hci design concerns“. In: *Proceedings of the Fourth International Conference on Human-Computer Interaction*. 2002, pp. 303–307 (cit. on p. 37).

[81] Cristina Palmero, Albert Clapés, Chris Bahnsen, et al. „Multi-modal RGB–Depth–Thermal Human Body Segmentation“. In: *International Journal of Computer Vision* (2016), pp. 1–23 (cit. on p. 48).

[82] L. Pareto, J. Broeren, D. Goude, and M. Rydmark. „Virtual reality, haptics and post-stroke rehabilitation in practical therapy". In: *Proc. 7th Int l Conf. Disability, Virtual Reality and Associated Technologies (ICDVRAT 08)*. Porto, Portugal, 2008, pp. 245–252 (cit. on pp. 2, 10).

[83] D. Rand, R. Kizony, and PL Weiss. „Virtual reality rehabilitation for all: Vivid GX versus Sony PlayStation II EyeToy". In: *Proc. 5th Int l Conf. Disability, Virtual Reality, and Associated Technologies (ICDVRAT 04)*. Citeseer. Oxford, UK, 2004, pp. 87–94 (cit. on p. 10).

[84] P. Rego, P.M. Moreira, and L.P. Reis. „Serious games for rehabilitation: A survey and a classification towards a taxonomy". In: *Proc. 5th IEEE Iberian Conf. Information Systems and Technologies (CISTI 10)*. Viana do Castelo, Portugal, 2010, pp. 1–6 (cit. on pp. 3, 10, 12, 38, 39).

[85] A. Reyes-Amaro, Y. Fadraga-González, O.L. Vera-Pérez, et al. „Rehabilitation of patients with motor disabilities using computer vision based techniques". In: *JACCES: Journal of accessibility and design for all* (2012), pp. 62–70 (cit. on p. 37).

[86] Laurence Z Rubenstein and Karen R Josephson. „Falls and their prevention in elderly people: what does the evidence show?" In: *Medical Clinics of North America* 90.5 (2006), pp. 807–824 (cit. on p. 91).

[87] Y. Salem and E.M. Godwin. „Effects of task-oriented training on mobility function in children with cerebral palsy". In: *NeuroRehabilitation* 24.4 (2009), pp. 307–313 (cit. on p. 2).

[88] M. Sandlund, S. McDonough, and C. Häger-Ross. „Interactive computer play in rehabilitation of children with sensorimotor disorders: a systematic review". In: *Dev. Med. Child Neurol.* 51.3 (2009), pp. 173–179 (cit. on pp. 9, 10, 12).

[89] M. Sandlund, E. Lindh Waterworth, and C. Hager. „Using motion interactive games to promote physical activity and enhance motor performance in children with cerebral palsy". In: *Developmental Neurorehabilitation* 14.1 (2011), pp. 15–21 (cit. on p. 10).

[90] D.W Scott. „Multivariate Density Estimation". In: *Multivariate Density Estimation, Wiley, New York, 1992* 1 (1992) (cit. on pp. 52, 53).

[91] Roland Sigrist, Georg Rauter, Robert Riener, and Peter Wolf. „Augmented visual, auditory, haptic, and multimodal feedback in motor learning: A review". English. In: *Psychonomic Bulletin & Review* 20.1 (2013), pp. 21–53 (cit. on pp. 11, 22).

[92] J. Smisek, M. Jancosek, and T. Pajdla. „3D with Kinect". In: *Proc. IEEE Int l Conf. Computer Vision Workshops (ICCV Workshops 11), 2011*. Prague, Czech Republic, 2011, pp. 1154–1160 (cit. on p. 19).

[93] Andrews Sobral and Antoine Vacavant. „A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos“. In: *Computer Vision and Image Understanding* 122 (May 2014), pp. 4–21 (cit. on pp. 42, 43).

[94] Young-min Song, SeungJong Noh, Jongmin Yu, Cheon-wi Park, and Byung-Geun Lee. „Background subtraction based on Gaussian mixture models using color and depth information“. In: *Control, Automation and Information Sciences (ICCAIS), 2014 International Conference on*. IEEE. 2014, pp. 132–135 (cit. on pp. 45, 47).

[95] Andre Störmer, Martin Hofmann, and Gerhard Rigoll. „Depth gradient based segmentation of overlapping foreground objects in range images“. In: *Information Fusion (FUSION), 2010 13th Conference on*. IEEE. 2010, pp. 1–4 (cit. on pp. 44, 47, 68).

[96] Jörg Stückler and Sven Behnke. „Efficient Dense 3D Rigid-Body Motion Segmentation in RGB-D Video.“ In: *BMVC*. 2013 (cit. on p. 46).

[97] Kelvin Sung. „Recent videogame console technologies“. In: *Computer* 44.2 (2011), pp. 91–93 (cit. on p. 10).

[98] S. Swink. *Game feel: a game designer s guide to virtual sensation*. Morgan Kaufmann, 2008 (cit. on p. 28).

[99] Kazumoto Tanaka, JR Parker, Graham Baradoy, et al. „A Comparison of Exergaming Interfaces for Use in Rehabilitation Programs and Research“. In: *Loading...* 6.9 (2012) (cit. on p. 10).

[100] Alexandru Telea. „An image inpainting technique based on the fast marching method“. In: *Journal of graphics tools* 9.1 (2004), pp. 23–34 (cit. on p. 55).

[101] M.E. Tinetti. „Performance-oriented assessment of mobility problems in elderly patients“. In: *J. Am. Geriatr. Soc.* 34.2 (1986), pp. 119–126 (cit. on p. 33).

[102] Tom N Tombaugh and Nancy J McIntyre. „The mini-mental state examination: a comprehensive review.“ In: *Journal of the American Geriatrics Society* (1992) (cit. on p. 22).

[103] Kentaro Toyama, John Krumm, Barry Brumitt, and Brian Meyers. „Wallflower: Principles and Practice of Background Maintenance“. In: *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on* Vol. 1.September (1999) (cit. on pp. 41, 64).

[104] M. Turk. „Computer vision in the interface“. In: *Commun. ACM* 47.1 (2004), pp. 60–67 (cit. on pp. 2, 10).

[105] Javier Varona, Antoni Jaume-i Capó, Jordi González, and Francisco J Perales. „Toward natural interaction through visual recognition of body gestures in real-time“. In: *Interacting with computers* 21.1-2 (2009), pp. 3–10 (cit. on pp. 2, 37).

[106] Paulette M. van Vliet and Gabriele Wulf. „Extrinsic feedback for motor learning after stroke: What is the evidence?“ In: *Disability and Rehabilitation* 28.13-14 (2006), pp. 831–840. eprint: http://dx.doi.org/10.1080/09638280500534937 (cit. on pp. 11, 22).

[107] Matt P Wand and M Chris Jones. „Comparison of Smoothing Parameterizations in Bivariate Kernel Density Estimation“. In: *Journal of the American Statistical Association* 88.422 (1993), pp. 520–528 (cit. on p. 52).

[108] Matt P Wand and M Chris Jones. *Kernel smoothing*. Vol. 60. Crc Press, 1994 (cit. on p. 52).

[109] Gabriele Wulf, Suzete Chiviacowsky, and Rebecca Lewthwaite. „Normative feedback effects on learning a timing task“. In: *Research quarterly for exercise and sport* 81.4 (2010), pp. 425–431 (cit. on pp. 4, 11).

[110] Zoran Zivkovic. „Improved adaptive Gaussian mixture model for background subtraction“. In: *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*. Vol. 2. IEEE. 2004, pp. 28–31 (cit. on pp. 69, 79–86).

[111] M. Zyda. „From visual simulation to virtual reality to games“. In: *Computer* 38.9 (2005), pp. 25–32 (cit. on p. 8).

List of Figures

List of Tables