

SISTEMA DE VISIÓ ARTIFICIAL BASAT EN UNA CÀMERA ESTEREOSCÒPICA PER A LA MESURA DE VELOCITAT EN ENTRENAMENT DE PES

Eloy Sanchis López

Tutor: José Manuel Mossi García

Treball final de màster presentat en l'Escola Tècnica Superior d'Enginyeria de Telecomunicació de la Universitat Politècnica de València per a l'obtenció del títol de Màster Universitari en Enginyeria de Telecomunicació

Curs 2021/2022

València, 7 de març de 2022

Resum

El desplaçament vertical i repetit de pes és un exercici prescrit per a l'entrenament físic d'esportistes d'una gran varietat de disciplines esportives. En algunes metodologies d'entrenament, la mesura de la velocitat instantània de la barra amb els pesos és important perquè permet optimitzar la càrrega prescrita a l'esportista i reduir el nombre de dies de recuperació, de manera que s'aconsegueixen més sessions d'entrenament efectiu per setmana.

Este projecte proposa realitzar la mesura sobre una màquina de Smith mitjançant l'ús d'una càmera estereoscòpica i un sistema de visió artificial basat en el tractament del núvol de punts generat i de la informació bidimensional de manera complementària per a millorar la precisió i arribar-ne als nivells requerits. Este sistema permet la determinació automàtica de la geometria de la màquina, el seguiment de la barra i la detecció de les fases concèntrica i excèntrica de l'exercici, i genera els valors de mesura de velocitat mitjana i màxima, força i potència en cadascuna de les fases, que són necessaris per a la monitoració i planificació adequades dels programes d'entrenament.

Resumen

El desplazamiento vertical y repetido de peso es un ejercicio prescrito para el entrenamiento físico de deportistas de una gran variedad de disciplinas deportivas. En algunas metodologías de entrenamiento, la medida de la velocidad instantánea de la barra con las pesas es importante porque permite optimizar la carga prescrita al deportista y reducir el número de días de recuperación, consiguiendo así más sesiones de entrenamiento efectivo por semana.

Este proyecto propone realizar esa medida sobre una máquina de Smith mediante el uso de una cámara estereoscópica y un sistema de visión artificial basado en el tratamiento de la nube de puntos generada y de la información bidimensional de forma complementaria para mejorar la precisión y llegar a los niveles requeridos de la misma. Este sistema permite la determinación automática de la geometría de la máquina, el seguimiento de la barra y la detección de las fases concéntrica y excéntrica del ejercicio, y genera los valores de medida de velocidad media y máxima, fuerza y potencia en cada una de las fases, que son necesarios para la monitorización y planificación adecuadas de los programas de entrenamiento.

Abstract

A repeated displacement of weight in the vertical direction is a prescribed exercise in the physical training of athletes from a great variety of sports disciplines. Barbell velocity measurement is important in some training methodologies because it allows optimization of the load prescribed to the athlete and a reduction of recovery days, with a result of more effective training sessions each week.

This project proposes to carry out that measurement on a Smith machine by using a stereoscopic camera and a computer vision system based on the treatment of the generated point cloud and bidimensional information in a complementary fashion in order to improve precision and achieve its required levels. This system allows automatic determination of the machine's geometry, barbell tracking, and detection of the concentric and eccentric phases of the exercise, and it generates

measurement values for average and maximum velocity, force, and power in each phase. Those values are necessary for an adequate monitoring and planning of training programmes.

Índex

1. Introducció	1
2. Objectius del treball	3
3. Fonaments teòrics	5
3.1. Visió artificial	5
3.2. Visió estereoscòpica	6
3.2.1. Model de càmera	7
3.2.2. Geometria estereoscòpica	10
3.2.3. Característiques, tipus i models de càmeres estèreo	13
3.3. Núvols de punts i fitxers associats	15
3.4. Estat actual de la tècnica	16
4. Desenvolupament del sistema	19
4.1. Eines i entorn de treball	19
4.2. Càmera estèreo utilitzada	21
4.2.1. Característiques i especificacions	21
4.2.2. Fluxos de dades disponibles	24
4.2.3. Sistemes de coordenades	25
4.2.4. Qualitat del núvol de punts	27
4.2.4.1. Soroll i deformacions	27
4.2.4.2. Projecció d'infrarojos i il·luminació de l'escena	28
4.2.4.3. Alineació de les imatges dels diferents sensors	28
4.2.4.4. Orientació de la càmera	28
4.2.4.5. Calibració	29
4.2.4.6. Comprovacions	30
4.2.5. Qualitat de la imatge bidimensional de color	32
4.3. Tècniques i funcions de processament	34
4.3.1. Funcions de segmentació	35
4.3.2. Funcions d'ajust	37
4.4. Integració del processament 2D i 3D	39
4.4.1. Decisió de seguir la barra sobre una imatge plana	39
4.4.2. Mètode de detecció i seguiment de la barra transversal	40
4.4.3. Correspondència entre la imatge bidimensional i el núvol de punts	41
4.4.3.1. Projecció	42
4.4.3.2. Transformacions geomètriques	42

4.4.3.3.	Homografia plana	44
4.4.3.4.	Una situació semblant, però en 2D	45
4.4.3.5.	Aplicació de les relacions geomètriques al nostre treball	46
4.5.	Magnituds d'interés	53
4.5.1.	Posició, velocitat i acceleració	53
4.5.2.	Càlcul de forces	54
4.5.3.	Potència	55
4.5.4.	Fases de l'exercici	56
4.6.	Descripció de l'algoritme	56
4.6.1.	Lectura de paràmetres i importació de dades	56
4.6.2.	Transformació de l'escena	57
4.6.3.	Aïllament de la màquina de Smith	58
4.6.4.	Detecció i modelatge de les parts de la màquina	58
4.6.5.	Determinació del pla de moviment i del seu sistema de coordenades	60
4.6.6.	Detecció de la marca de la barra en les imatges 2D	61
4.6.7.	Determinació de la posició de la marca en l'espai i en el pla de moviment	64
4.6.8.	Càlcul de les magnituds d'interés	65
5.	Validació i resultats	67
5.1.	Preparació de les gravacions	67
5.2.	Gravacions en condicions reals	70
5.3.	Validació	71
5.4.	Resultats	73
6.	Conclusions i proposta de treball futur	79
	Bibliografia	81

Índex de figures

1.1. Màquina de Smith	2
3.1. Model de càmera <i>pinhole</i>	7
3.2. Model equivalent de càmera	8
3.3. Eixos de coordenades usuals en el model equivalent	9
3.4. Correspondència de punts entre dos imatges	10
3.5. Geometria de la restricció epipolar	11
3.6. Geometria del càlcul de la profunditat	12
3.7. Relació entre la disparitat i la distància	13
3.8. Zones de visió estereoscòpica i monocular	13
3.9. Dos vistes d'un núvol de punts bàsic	15
3.10. Detecció de punts singulars en cotxes en moviment	17
4.1. Patró de punts infrarojos	21
4.2. Franja sense informació al lateral d'una imatge de profunditat	22
4.3. Franja sense informació a diferents distàncies	22
4.4. Comparació dels camps de visió dels diferents sensors en imatges 2D	23
4.5. Comparació dels camps de visió dels diferents sensors en núvols de punts	23
4.6. Efecte dels punts "sense color" en obrir el núvol en un visor extern	23
4.7. Alguns dels <i>topics</i> d'un <i>bag</i> que conté una gravació de la imatge de profunditat (en groc, les imatges d'una gravació)	25
4.8. Dades dels primers missatges del <i>topics</i> de les dades de profunditat	26
4.9. Núvol generat per la càmera estèreo i disposició del sistema de coordenades	26
4.10. Qualitat de dos núvols de punts d'una mateixa cadira davant d'una paret	27
4.11. Efecte de la projecció d'un patró infraroig en la detecció de la profunditat	28
4.12. Alineació correcta de la informació de profunditat i la imatge de color	29
4.13. Efecte teòric de l'autocalibració	29
4.14. Patró utilitzat en la calibració dinàmica	30
4.15. Detecció de la barra transversal de la maqueta segons l'orientació de la càmera	31
4.16. Marques usades en les mesures de l'error de la càmera	32
4.17. Detall de la nitidesa de la barra amb diferents valors d'exposició en interiors	33
4.18. Detall de la nitidesa de la barra amb diferents valors d'exposició junt a una finestra	34
4.19. Detall de la nitidesa de la barra en exteriors	34
4.20. Imatge de la màquina de Smith del gimnàs de l'UPV	35
4.21. Segmentació de la zona de la màquina de Smith, on cada grup es mostra d'un color	36
4.22. Ajust d'un pla a un núvol format pels dos pilars de la màquina	37
4.23. Ajust d'un pla al terra de l'escena de la màquina de Smith	38
4.24. Ajust de rectes als pilars i a la barra transversal de la màquina	39

4.25. Comparació d'una part de la màquina de Smith en les captures 3D i 2D	40
4.26. Detecció de la marca en un fotograma de la gravació	42
4.27. Jerarquia de les transformacions geomètriques	43
4.28. Efecte de diverses transformacions geomètriques	44
4.29. Plans i sistemes de coordenades que intervenen en una homografia plana	46
4.30. Punts i plans que intervenen en la detecció d'un marcador utilitzant una imatge 2D	47
4.31. Projectió inversa que determina la posició espacial de la marca	47
4.32. Tria del sistema de coordenades del pla de moviment de la barra transversal	48
4.33. Obtenció de les coordenades d'un punt en un altre sistema de referència	49
4.34. Sistemes de coordenades que intervenen en la integració del processament 2D i el 3D	50
4.35. Forces que actuen sobre la barra (vista de perfil)	55
4.36. Núvol de punts original de l'escena amb un model de pla ajustat al terra	57
4.37. Núvol de punts de l'escena després de fer coincidir el terra amb $y = 0$	58
4.38. La màquina de Smith i el seu entorn en segmentar-los per connectivitat	59
4.39. Màquina de Smith aïllada	59
4.40. Modelatge de la barra i dels pilars per elements geomètrics	60
4.41. Ajust d'un pla als pilars de la màquina de Smith	61
4.42. Plans de l'escena que configuren el sistema del pla de la barra	62
4.43. Exemple de zona on es cerca la marca de la barra dins d'un fotograma	63
4.44. Detecció de la marca en un fotograma de la gravació	63
4.45. Cent primeres deteccions de la marca de la barra (punts de colors) sobre el núvol de l'escena	64
5.1. Imatges bidimensionals de la màquina de Smith del gimnàs de l'UPV	69
5.2. Núvol de punts de la màquina de Smith del gimnàs de l'UPV amb textura de color	69
5.3. Gravació en condicions reals amb un esportista	70
5.4. Mètode de validació format per una cinta mètrica i un fil d'aram	71
5.5. <i>Jitter</i> relatiu dels fotogrames de la gravació de la càmera estèreo	72
5.6. Posició de la barra transversal al llarg del temps	73
5.7. Error en la detecció de la barra	74
5.8. Velocitat de la barra transversal al llarg del temps	74
5.9. Error en l'estimació de la velocitat de la barra	75
5.10. Acceleració de la barra transversal al llarg del temps	75
5.11. Força exercida per l'esportista al llarg del temps	76
5.12. Error en l'estimació de la força exercida	76
5.13. Potència desenvolupada per l'esportista al llarg del temps	77
5.14. Error en l'estimació de la potència	77

Índex de taules

4.1. Especificacions dels sensors monocroms i del sensor RGB	21
4.2. Especificacions de la imatge de profunditat	22
4.3. Combinacions de resolucions i freqüències de camp per a cada flux de vídeo	24
4.4. Qualitat subjectiva del núvol de punts segons la distància i l'orientació relativa de la màquina i la càmera	31
4.5. Error en les mesures de la càmera a diferents distàncies	32
5.1. Error en la projecció de punts a un pla de l'escena	76

Capítol 1

Introducció

Un dels exercicis que sovint forma part de les rutines d'entrenament físic és el desplaçament vertical de pes. Es tracta d'una activitat molt versàtil en el sentit que està indicada per a una gran varietat de persones i situacions: és adequada, amb la intensitat i les variacions pertinents, per a l'entrenament d'esportistes professionals, per a pacients de fisioteràpia o rehabilitació, o simplement per a persones que vulguen mantindre una bona forma física general.

Els esportistes professionals són els qui tenen la necessitat de controlar i registrar el rendiment per tal d'optimitzar els entrenaments de pes i determinar quant i com s'ha d'entrenar de cara a aconseguir els objectius que es proposen. En general, l'objectiu és conèixer quin és l'estat de forma de l'esportista i millorar-lo. La manera clàssica de fer açò és a través del paràmetre 1RM, que és la càrrega màxima que l'esportista pot arribar a alçar però només una vegada (ja que els músculs s'esforcen massa com per a poder dur a terme dos repeticions). Els entrenaments posteriors es configuren com a un cert nombre de repeticions i sèries alçant una certa fracció de la càrrega 1RM.

Tanmateix, la manera en què se sol determinar 1RM no és massa beneficiosa per a la forma física dels esportistes: alçar tant de pes que els músculs no poden repetir-ho una segona vegada els sotmet a més estrès i més esgotament del que és recomanable. Com a conseqüència, eixa gran càrrega allarga el temps de recuperació que necessiten els músculs per a poder realitzar més esforços i, per tant, perquè l'esportista pugui entrenar de nou de manera eficaç. A més, els mètodes d'entrenament basats en 1RM pequen de no tindre en compte la variació en l'estat físic de l'esportista segons el dia.

Els mètodes d'entrenament alternatius als basats en 1RM són els que tenen en compte principalment la velocitat a la qual es realitzen cada repetició dels exercicis. Mitjançant estos mètodes, es pot determinar de manera precisa la quantitat de repeticions de cada sèrie i la càrrega de cada repetició per a optimitzar el rendiment, minimitzar el temps de recuperació entre sessions d'entrenament consecutives, i en general valorar la forma física de l'esportista sense tècniques perjudicials. L'avantatge principal dels mètodes de velocitat respecte als d'1RM és precisament l'evitació d'eixes sobreesforç (i, addicionalment, la consideració d'eixes variacions puntuals en la forma física de l'esportista). A més, en cas de voler conèixer de tota manera el valor de la càrrega 1RM, es pot determinar a partir de la velocitat i la massa segons taules de conversió.

El sistema que s'ha dissenyat i implementat en este treball s'emmarca dins d'eixes tècniques de quantificació dels entrenaments físics basades en la velocitat, i tracta de facilitar i apropar al públic en general este tipus de mètodes, que fins ara han tingut certs inconvenients pel que fa a

les tècniques de mesura. Els transductors lineals permeten obtenir les millors mesures però el seu preu és elevadíssim, els sistemes 3D multicamerals són molt aparatosos i requereixen de molta preparació per a muntar-los, i els dispositius amb unitats de mesura inercial (IMU) resulten poc fiables depenent del model.

En canvi, el nostre sistema no presenta eixos inconvenients. Es tracta d'un algoritme que calcula la velocitat a la qual s'executa l'exercici de *back squat* en una màquina de Smith (una estructura metàl·lica que permet alçar pes verticalment gràcies a unes guies fixades a eixa estructura, tal com s'observa en la figura 1.1) basant-se simplement en les imatges captades per una càmera estereoscòpica. En un núvol tridimensional de punts es localitza la màquina en l'espai mitjançant les tècniques de processament apropiades, i se'n distingixen les diferents parts. D'una altra banda, es detecta en una gravació 2D de l'exercici la posició d'una xicoteta marca de color fixada a la barra que suporta el pes. En integrar tots estos elements, podem saber la posició espacial de la barra amb precisió durant tot l'exercici, i extraure'n la velocitat i altres magnituds útils per als esportistes.

El sistema ha sigut desenvolupat amb la vista posada en una futura implementació en telèfons mòbils intel·ligents, que comencen a incorporar capacitat de visió estèreo. Això avança en la democratització de l'accés a estos mètodes d'entrenament físic, a més de proporcionar avantatges com la possibilitat de conèixer la informació útil instantàniament o el fet de tractar-se d'un mètode de mesura molt poc invasiu en comparació amb altres. Es tracta, en definitiva, d'una millora potencial en la manera d'entrenar de molts esportistes.



Figura 1.1: Màquina de Smith [1]

Capítol 2

Objectius del treball

El desenvolupament d'este algoritme ha comportat una quantitat notable de treball previ i ha estat format per diferents tasques, totes les quals han sigut necessàries per a completar el sistema. Els principals objectius han sigut els següents:

- *Familiaritzar-se amb el funcionament de la càmera estèreo.* Aprendre a utilitzar els programes de la càmera estèreo i realitzar proves de captura de núvols de punts per a adquirir una certa intuïció de com les condicions ambientals influïxen en els núvols obtinguts i quins objectes es pot esperar captar i amb quina qualitat en diferents contextos
- *Avaluar i familiaritzar-se amb les tècniques de processament de núvols.* Realitzar proves amb distintes biblioteques de processament de núvols de punts per tal de determinar quines poden ser útils en el nostre projecte, així com per tal d'adquirir experiència i intuïció en el seu maneig i en la tria dels paràmetres adequats per a un funcionament correcte en cada situació
- *Reconèixer automàticament les diferents parts de l'escena i de la màquina de Smith.* Utilitzar les funcions seleccionades abans combinant-les de la manera adequada per tal d'aconseguir detectar correctament els objectes de l'escena i eliminar els que no són la màquina de Smith, i una vegada aïllada la màquina, destriar-ne les parts principals (pilars i barra transversal) i modelar-les per entitats geomètriques
- *Integrar visió 2D i 3D per a conèixer la posició de la barra amb precisió.* Detectar la barra transversal de la màquina durant tot l'exercici amb tècniques de visió artificial bidimensional, i combinar esta informació mitjançant operacions geomètriques amb les dades sobre eixa barra procedents de la visió estèreo per tal d'obtindre'n la posició en l'espai amb una precisió major
- *Extraure la informació d'interés.* Calcular, a partir de la posició espacial de la barra transversal en cada instant, les magnituds que resulten d'interés per als esportistes (especialment la velocitat)

Capítol 3

Fonaments teòrics

Este projecte fa ús de diverses tecnologies relacionades amb l'adquisició i el processament d'imatges, i amb el tractament d'escenes tridimensionals per part d'ordinadors. Per la part esportiva, també hem tingut en compte la investigació existent sobre els mètodes d'entrenament i com la tecnologia pot optimitzar-los. Cal comprendre bé tots estos aspectes per tal de desenvolupar i explicar l'algoritme del sistema.

3.1. Visió artificial

Podríem dir que la visió artificial és l'extracció d'informació útil i més o menys complexa a partir d'una imatge o vídeo per part d'una màquina o ordinador anàlogament a com ho fem les persones [2, 3]. És a dir, les persones tenim òrgans (els ulls) que s'encarreguen de captar la realitat en forma de llum, però eixa llum procedent de les diferents zones de l'espai no significa res per ella mateixa, sinó que no és fins després de ser enviada en forma d'impulsos nerviosos al cervell, on es processa, quan adquirix un significat realment. De la mateixa manera, el vídeo que capta una càmera no proporciona cap informació per ell mateix si no es processa per a extraure'n les característiques i obtindre'n informació útil, tal com les persones aprenem a fer sense ser-ne conscients.

Seguint en la mateixa línia d'imitar les persones, la visió artificial va en molts casos –tot i que no en el nostre– acompanyada de la intel·ligència artificial (IA), fins al punt que moltes persones consideren els dos camps indissociables. La idea en què es basa això és que les persones aprenem al llarg dels primers anys de vida a processar i extraure informació útil de la llum a base de pràctica, ja que el fet d'utilitzar la vista durant una gran part del temps fa que disposem de molts estímuls que entrenen la nostra capacitat de reconeixement de formes i altres operacions de processament. El fonament de la IA és justament eixe: fer que la màquina aprenga sola (i en molt menys de temps que les persones) gràcies a conjunts de dades preclassificades d'entrenament que modelen les seues capacitats de processament. Les eines que possibiliten això són les xarxes neuronals i l'aprenentatge profund (*deep learning*). La relació entre la visió artificial i la IA es pot resumir dient que «si la IA fa que els ordinadors puguem pensar, la visió artificial fa que puguem veure, observar i entendre» [4].

Amb IA o sense, la visió artificial es val de certes tasques [2, 4] per a dur a terme els seus

objectius. Entre altres hi trobem:

- Classificació d'imatges, o reconeixement d'objectes: decidir a quina classe, d'entre un grup de classes prèviament definides, pertany una imatge
- Identificació d'objectes: reconèixer la presència d'una instància específica d'un objecte
- Detecció d'objectes: detectar si un objecte o característica està present en la imatge o no
- Seguiment d'objectes: detectar l'evolució de la posició i/o velocitat d'un objecte en imatges consecutives en el temps

Estes tasques es poden dur a terme gràcies a una sèrie de mètodes o tècniques [2]. Entre altres hi trobem:

- Adquisició d'imatges, que en el nostre cas es du a terme amb càmeres estèreo que registren la llum visible, però poden utilitzar-se altres classes de sensors capaços de generar una imatge, o fins i tot poden dependre de magnituds físiques diferents a la llum
- Extracció de característiques, consistent en detectar en les imatges característiques simples com vores, cantons, punts, etc. En este treball, està present en l'algoritme que detecta la posició de la barra en la gravació bidimensional, i alguns aspectes de la detecció de les parts de la màquina de Smith podrien considerar-se una forma simple d'extracció de característiques.
- Segmentació, consistent en separar una imatge o objecte en parts, i/o seleccionar-ne només les zones susceptible de processament. En este treball s'utilitza, per exemple, per a aïllar la màquina de Smith del seu entorn (tret que es fa en 3D, no en 2D).

Tot açò possibilita aplicacions de la visió artificial com poden ser: navegació automàtica (per exemple, per part de vehicles autònoms), control de qualitat, reconeixement facial, control d'aforament, reconeixement de caràcters, etc. [2, 4]

3.2. Visió estereoscòpica

La visió estereoscòpica (o, per abreviar, visió estèreo), és la captació d'imatges que inclouen informació de profunditat, normalment gràcies a l'ús de diversos sensors separats per una certa distància i el processament de les diferències entre les imatges que generen [5].

Igual que abans, es pot establir clarament una analogia o una voluntat d'imitació respecte al sistema visual humà. Les persones tenim la capacitat de veure en tres dimensions, és a dir, de percebre la profunditat, ja que cada un dels nostres dos ulls capten imatges lleugerament diferents, i el nostre cervell processa eixes diferències per a construir la percepció de profunditat. Està clar que el cervell i els ordinadors no processen la informació de la mateixa manera, però tot i així, la visió estèreo per part de màquines imita l'esquema de diversos sensors i l'explotació de les diferències entre les imatges que capten. De fet, en moltes ocasions es du a terme mitjançant dos sensors separats per una distància comparable a la distància interocular.

Per tal d'entendre millor com i per què funciona la visió estèreo duta a terme per part de màquines, revisem a continuació els fonaments geomètrics i de processament sobre els quals es basa.

3.2.1. Model de càmera

Una càmera normal i corrent capta la llum de l'escena i la projecta sobre una superfície fotosensible de tal manera que s'hi forma una imatge. És a dir, d'alguna manera una càmera transforma la realitat tridimensional en una imatge bidimensional mitjançant una projecció (en el sentit matemàtic i en el geomètric).

Els models de càmera [6] no són intents de descriure com capten la imatge les càmeres de manera físicament exacta, sinó més bé conceptualitzacions geomètriques que no necessàriament es corresponen amb el fenomen físic però sí que modelen bé els seus resultats i ens ajuden a comprendre'n el funcionament. Així, en els models se sol parlar de punts, línies, plans, etc. alguns dels quals poden ser més o menys abstractes.

Entenent el funcionament d'una càmera com a una projecció, en els models se solen definir els anomenats punt objecte i punt imatge. El punt objecte és un punt qualsevol de l'escena (és a dir, tridimensional). Podem imaginar que representa la posició d'un objecte d'interès. El punt imatge, per la seua part, és el punt on es projecta el punt objecte. El resultat de projectar tots els punts objecte possibles és el conjunt de tots els punts imatge, que formen un pla en l'espai anomenat pla imatge. Per tant, són punts bidimensionals.

El model més simple de càmera és l'anomenat model de càmera estenopecica (més conegut amb el seu nom anglés, *pinhole camera model*). En este model, la superfície fotosensible es troba dins d'un espai tancat quasi per complet, l'única obertura del qual és un forat molt xicotet (infinitesimal, en teoria). Això permet que la llum entre dins de l'espai de la càmera, però la mida reduïda de l'obertura fa que a cada punt de la superfície fotosensible (és a dir, del pla imatge) arribe un raig de llum procedent d'una única direcció, de tal manera que la imatge que es projecta sobre el pla imatge és clara i nítida i té tots els elements enfocats.

La relació matemàtica entre el punt objecte i el punt imatge es pot establir fàcilment mitjançant la semblança de triangles. L'eix òptic és la recta que passa pel forat o obertura de la càmera i és perpendicular al pla imatge, i podem representar la posició del punt objecte i la del punt imatge mitjançant la seua distància a eixe eix òptic.

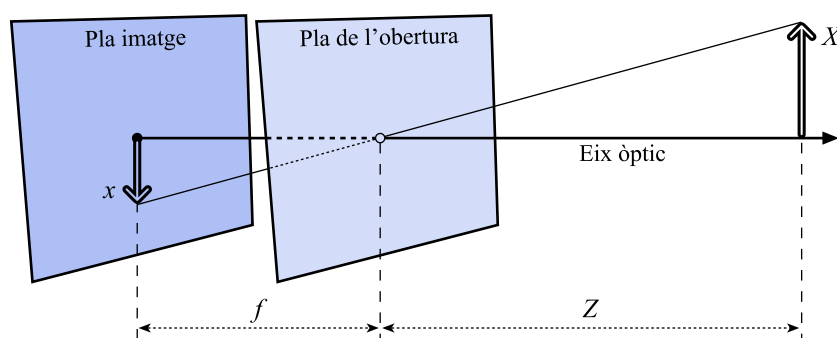


Figura 3.1: Model de càmera *pinhole* [6]

Així, i tal com podem veure en la figura 3.1, sent X la distància del punt objecte a l'eix òptic, sent x la distància del punt imatge a l'eix òptic, sent f la distància focal (és a dir, en este cas, la distància de l'obertura al pla imatge), i sent Z la distància de l'obertura al punt de projecció del

punt objecte sobre l'eix òptic, podem establir la relació de semblança de la manera següent:

$$\frac{-x}{f} = \frac{X}{Z} \quad (3.1)$$

Cal tindre en compte que la imatge apareix cap per avall per la geometria del model. Com que x té un valor positiu, apareix el signe negatiu en l'equació per tal de tindre en compte eixa inversió de la imatge. De totes maneres, la càmera redreça posteriorment la imatge, és a dir, du a terme una segona inversió per tal que aparega amb l'orientació correcta. Per descomptat, tot el que aplica per a x i X aplica també per a y i Y , que és l'altre eix de coordenades del pla imatge.

Com que el resultat final de l'orientació de la imatge és positiu, i per tal de fer el model més entenedor i que treballar amb ell siga més fàcil, el model de càmera estenopeica se sol canviar per un altre amb una disposició diferent dels elements, però que dona els mateixos resultats: es tracta de l'anomenat model equivalent.

Abans, l'obertura de la càmera se situava entre l'escena i el pla imatge; en canvi, en el model equivalent movem el pla imatge entre l'obertura i l'escena, tal com mostra la figura 3.2. En el pla imatge se seguix projectant l'escena amb la mateixa mida que abans, ja que la distància entre el pla imatge i l'obertura és la mateixa que abans, només que per l'altra banda. El fet de canviar la disposició també fa que ara la projecció ja no estiga cap per avall. A més, l'obertura es convertix ara en una espècie de punt de fuga cap al qual convergixen tots els rajos per a fer la projecció, i se li sol donar el nom de centre òptic o centre de projecció. De fet, amb esta nova situació podríem descriure la projecció definint el punt imatge com el punt d'intersecció entre el pla imatge i la recta que va des del punt objecte fins al centre òptic.

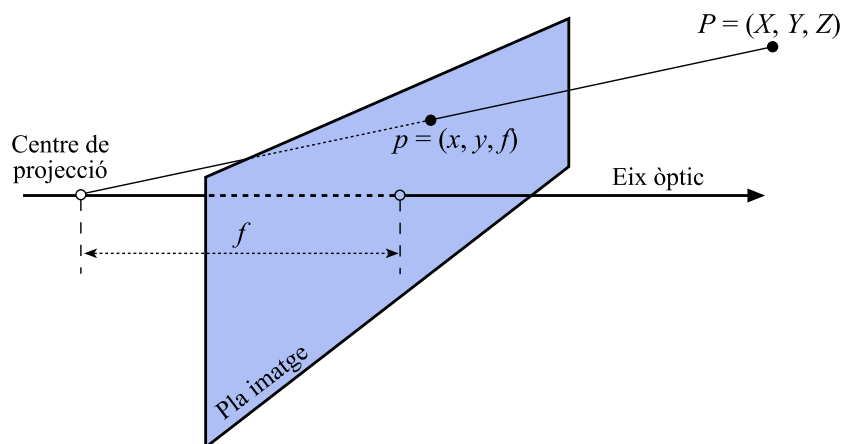


Figura 3.2: Model equivalent de càmera [6]

Tenint en compte tot això, el model equivalent es pot representar, de nou per semblança de triangles, amb esta equació:

$$\frac{x}{f} = \frac{X}{Z} \quad (3.2)$$

Cal distingir entre diferents sistemes de coordenades que apareixen en este context. L'eix de coordenades de l'escena (del qual hem parlat fins ara) se sol situar, tal com mostra la figura 3.3, amb l'eix X positiu cap a la dreta, l'eix Y positiu cap avall i l'eix Z cap a l'escena, i amb l'origen en el centre òptic, de tal manera que el pla imatge és paral·lel al pla XY . Així, el punt objecte queda representat com a $P = (X, Y, Z)$, mentre que el punt imatge queda representat, en el sistema de coordenades de l'escena, com a $p = (x, y, f)$, ja que el pla imatge es troba a una distància f de l'origen (tot amb unitats de distància). L'eix de coordenades de la imatge se sol situar amb l'eix x positiu cap a la dreta i l'eix y positiu cap avall, i amb l'origen en el cantó superior esquerre. Representarem el punt imatge en el sistema de coordenades (2D) de la imatge com a $p_{img} = (x_{img}, y_{img})$ (amb unitats de píxels).

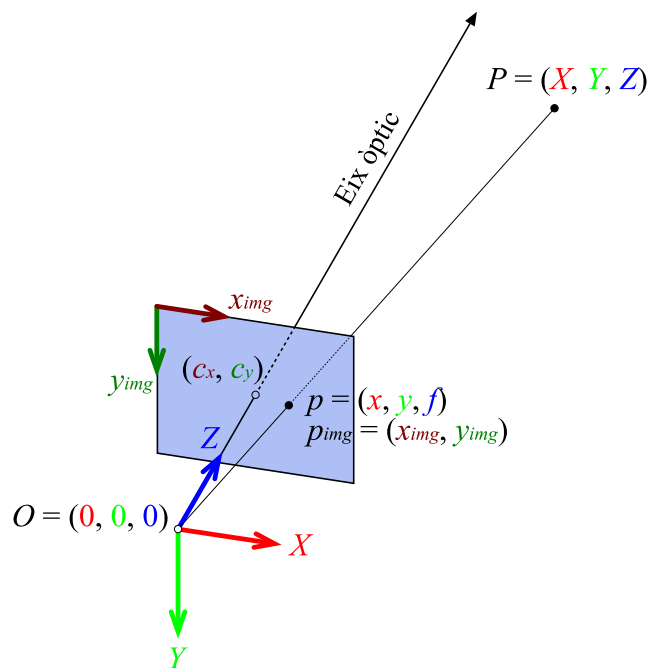


Figura 3.3: Eixos de coordenades usuals en el model equivalent [7]

L'equació 3.2, per tant, es referix sempre a distàncies físiques, però per a treballar amb les imatges s'ha de treballar en píxels. Per a poder adaptar-hi les equacions, és convenient fer un parell de puntualitzacions abans:

- El punt on l'eix òptic intersecciona el pla imatge s'anomena punt principal. Idealment, el punt principal coincideix amb el centre del sensor, però les imprecisions en la fabricació impedeixen que siga així, de manera que per a tindre-ho en compte se suma este paràmetre a les coordenades expressades en el sistema de referència de la imatge (i és també este paràmetre el que fa, de pas, que l'origen de coordenades siga el cantó superior esquerre, tal com s'ha dit abans).
- Les magnituds referents a la imatge passen a expressar-se en píxels, i f també ho fa, amb un matís. La distància focal expressada en píxels ja no és la distància focal físicament mesurable (f), sinó que és la distància focal físicament mesurable multiplicada pel factor de conversió a píxels. Eixe factor de conversió pot ser diferent en horitzontal (s_x) i en vertical (s_y) perquè els

píxels del sensor poden ser quadrats o rectangulars. Això fa que, parlant en píxels, tinguem dos distàncies focals diferents (f_x i f_y), una per a cada direcció (però s'anomenen distàncies focals igualment).

Tenint en compte tot això, l'adaptació del model equivalent per a passar de treballar amb distància a fer-ho amb píxels quedaria així:

$$x_{img} = f_x \cdot \frac{X}{Z} + c_x \quad (3.3)$$

$$y_{img} = f_y \cdot \frac{Y}{Z} + c_y \quad (3.4)$$

on $f_x = f \cdot s_x$ i $f_y = f \cdot s_y$

Existixen models de càmeres més complexos que permeten tindre en compte i corregir sensors mal orientats o no coplanars, la distorsió radial i la distorsió tangencial de les lents, etc., però no és necessari conèixer-los en profunditat en este treball, ja que les càmeres utilitzades proporcionen imatges amb una qualitat suficient.

3.2.2. Geometria estereoscòpica

La geometria de les càmeres estereoscòpiques (o càmeres estèreo, per abreujar) és clau en la seua capacitat de detectar la profunditat, de manera que expliquem els mecanismes que hi subjauen [7]. Considerem a partir d'ara una càmera estereoscòpica (o càmera estèreo) com les més habituals: amb dos sensors separats uns pocs centímetres i tan coplanars i alineats com és possible. Hem mencionat abans que el càlcul de la profunditat d'un punt, o el que és el mateix, la distància a la qual es troba respecte a la càmera, es basa en la diferència entre la posició en la qual apareix en una de les imatges i la posició en la qual apareix en l'altra. Això requerix relacionar cada punt d'una imatge amb un punt de l'altra (figura 3.4), cosa que no sempre és fàcil (i, de fet, a voltes no és possible) i que a més comporta un cost computacional gran. Tanmateix, la geometria de les càmeres estèreo facilita considerablement este procés.

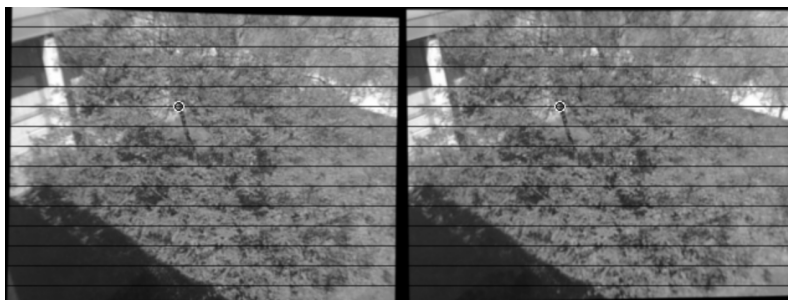


Figura 3.4: Correspondència de punts entre dos imatges [8]

La projecció de l'escena a una de les imatges és senzilla: tenint en compte el model equivalent, la recta del punt objecte al centre òptic intersectarà el pla imatge en algun lloc, i eixe lloc és el punt imatge, és a dir, el punt on veiem l'objecte en la imatge. En canvi, recórrer el camí invers (obtindre el punt objecte en l'espai coneixent el punt imatge) no és possible: podem considerar una recta en

l'espai que passe pel punt imatge i pel centre òptic, i sabem que el punt objecte es troba en eixa recta, però eixa recta conté infinits punts i no disposem d'informació suficient per a saber si el punt objecte es troba més a prop o més lluny sobre la recta. (De fet, per això necessitem dos càmeres en comptes d'una per a obtindre imatges amb informació de profunditat.)

Així, una càmera ens pot proporcionar una recta en l'espai que conté, entre els seus infinits punts, el punt objecte del qual volem saber la posició en l'escena. Com que les dos càmeres no estan en la mateixa posició, la perspectiva que tenen de l'escena no és la mateixa. Per tant, eixa recta en l'espai que una càmera veu com a un punt es veu en l'altra càmera com a una recta (anomenada recta epipolar). Recordem que en l'espai el punt objecte està contingut en la recta; en conseqüència, el punt objecte es projecta sobre el pla imatge de la segona càmera en un punt imatge que forma part de la recta epipolar, tal com s'aprecia en la figura 3.5.

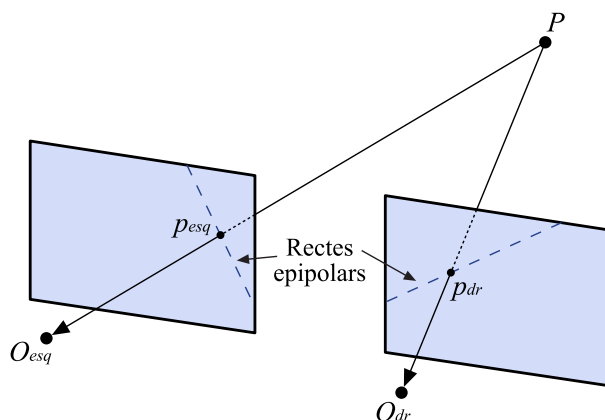


Figura 3.5: Geometria de la restricció epipolar

Açò implica que, a l'hora de buscar en una imatge la correspondència d'un punt de l'altra imatge, no cal fer-ho sobre tota la imatge com *a priori* es podria pensar, sinó que només cal buscar-la sobre la recta epipolar o en les seues proximitats, cosa que rep el nom de restricció epipolar. Això fa disminuir en gran mesura la càrrega computacional de calcular la profunditat dels milers de píxels d'una imatge, que tot i això és gran igualment.

Coneixent la recta en l'espai, calcular quina és la recta epipolar és molt més senzill i lleuger que calcular totes les possibles correspondències, i és possible si es coneixen les posicions i orientacions relatives dels dos plans imatge. És per això que és imprescindible dur a terme un procés de calibració de les càmeres que proporciona eixa informació.

Hem mencionat abans que cal que les dos càmeres estiguen ben alineades, que siguin coplanars, que la distància focal siga la mateixa en les dos càmeres, que no tinguin distorsions, etc. Cal puntualitzar que és pràcticament impossible que tot això es complisca de manera exacta, però existixen transformacions geomètriques i altres eines que permeten corregir eixes desviacions.

En qualsevol cas, una vegada s'ha relacionat un punt d'una imatge amb un punt de l'altra, el càlcul de la profunditat és senzill. Tenint en compte la geometria que es mostra en la figura 3.6, obtenim per semblança de triangles l'expressió 3.5:

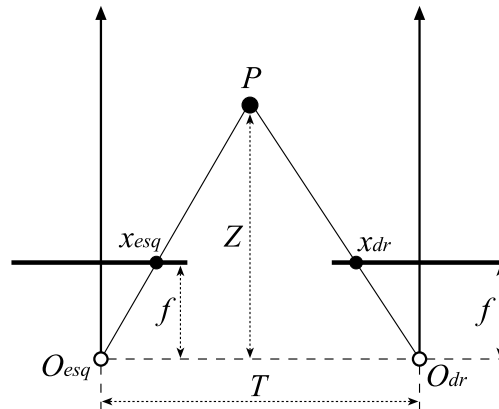


Figura 3.6: Geometria del càlcul de la profunditat [7]

$$\frac{T - (x_{esq} - x_{dr})}{Z - f} = \frac{T}{Z} \quad (3.5)$$

Cal tindre en compte que $T - (x_{esq} - x_{dr})$ és l'espai entre x_{esq} i x_{dr} . Reordenant i identificant arribem a la següent manera de calcular la profunditat de l'objecte:

$$Z = \frac{fT}{x_{esq} - x_{dr}} \quad (3.6)$$

En eixa expressió, $x_{esq} - x_{dr}$ s'anomena disparitat. Si la disparitat és més xicoteta, el punt ocupa quasi la mateixa posició en les dos imatges i l'objecte està més lluny; si és més gran, el punt està en una part diferent en cada imatge i l'objecte es troba més a prop de la càmera. Cal tindre en compte, doncs, que disparitat i distància (o profunditat) són inversament proporcionals, és a dir, la relació entre elles no és lineal, tal com s'aprecia en la figura 3.7. En termes pràctics, això implica que l'estimació de la profunditat funciona només en àrees properes a la càmera, on la paraula “properes” vol dir coses diferents segons els paràmetres de la càmera, sobretot la distància entre els sensors. La raó d'això és que les profunditats a partir d'un certa distància es corresponen amb disparitats massa xicotetes com per a ser detectades amb precisió.

La diferència de posició entre les dos càmeres provoca que hi haja zones dels laterals de la imatge que només una càmera pot veure. Això implica que el camp de visió estereoscòpica és la intersecció dels camps de visió de les dos càmeres, mentre que en les zones que només es troben dins del camp de visió d'una càmera la visió és monocular i no podem obtenir'n informació de profunditat (figura 3.8).

Respecte a les zones per a les quals es pot calcular la profunditat, podem assimilar el procés a un mètode de llançament de rajos. Imaginem que des del centre òptic es llancen rajos en totes les direccions, i estos rajos intersecten amb l'escena. Les zones de l'escena on ha arribat un raig són les que veu la càmera, però hi ha altres zones que queden ocultes de les quals la càmera no té informació. Es tracta de zones d'oclusió i poden constituir un problema ja que impedeixen conèixer la disposició de l'escena en la seua totalitat.

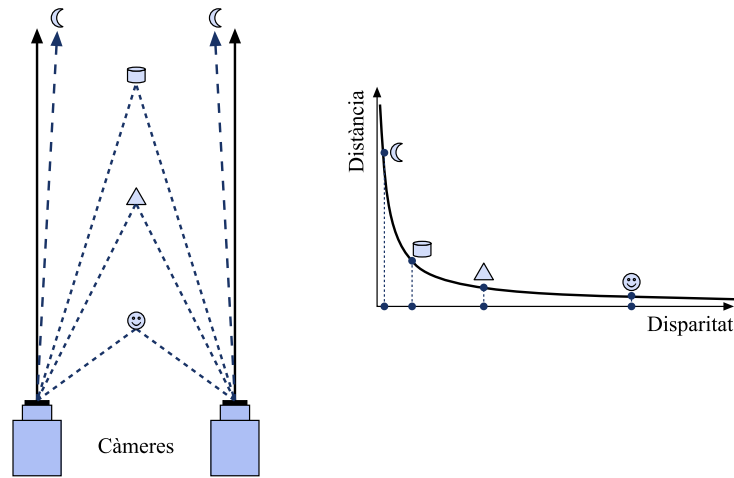


Figura 3.7: Relació entre la disparitat i la distància [7]

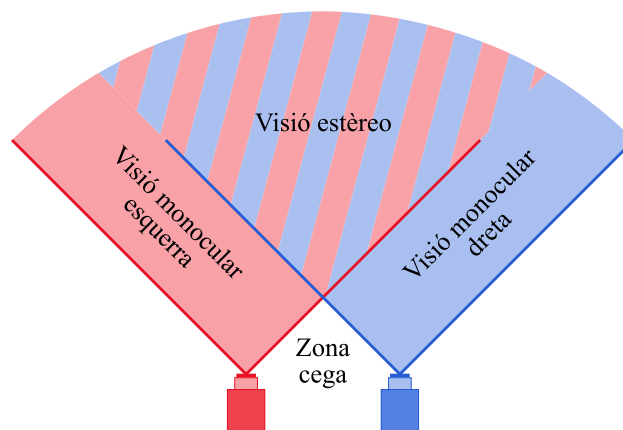


Figura 3.8: Zones de visió estereoscòpica i monocular

Una altra qüestió que pot provocar problemes en certes situacions és la dependència que té la disparitat respecte d'un bon mètode de cerca de correspondències entre les dos imatges, ja que requereix que l'escena tinga textura. Dit d'una altra manera, és fàcil que un algoritme de cerca de correspondències s'equivoque en superfícies llises d'un color uniforme.

3.2.3. Característiques, tipus i models de càmeres estèreo

Les càmeres estèreo estan caracteritzades per diversos paràmetres, però el més important dels que són específics per a estos sistemes és la distància de separació entre els dos sensors, també coneguda pel seu nom anglés *baseline*. Una separació major dona lloc a una disparitat major en relacionar els punts vistos per una càmera amb els vistos per l'altra, i conseqüentment a una major precisió en la detecció de la profunditat, que es pot dur a terme amb objectes més llunyans. Cal comentar també que l'error de profunditat augmenta quadràticament a mesura que augmenta la

profunditat mateixa.

En canvi, una major separació entre els sensors també provoca, com es pot deduir de la figura 3.8, un allunyament del camp de visió comú als dos sensors i per tant de la zona de visió estereoscòpica, a banda de més oclusions. Això du a la necessitat de trobar un equilibri entre una separació massa gran i massa xicoteta, o bé prioritzar la precisió o la zona de visió estèreo segons l'aplicació de què es tracte.

Un altre paràmetre dels sensors és la resolució (com en sensors de càmeres monoculars). En el cas de les càmeres estèreo, la resolució influeix en la precisió de la profunditat i en la distància més llunyana per a la qual se'n pot proporcionar informació, ja que una major resolució permet detectar disparitats més menudes. En canvi, també comporta un major cost computacional, ja que hi ha més píxels per als quals cal calcular la correspondència.

Pel que fa als tipus de càmeres, els sistemes passius són els bàsics que ja hem mencionat, que simplement busquen la correspondència de punts sense cap altra font externa d'informació i per tant patixen en zones sense textura. Per tal de solventar eixe problema, existixen sistemes de visió estèreo activa que contribueixen a l'escena amb il·luminació de certes característiques que millora la detecció de la profunditat [9].

Concretament, les càmeres que hem usat en este treball incorporen, a més dels dos sensors, un projector d'infrarojos que projecta un patró pseudoaleatori de punts sobre l'escena per a proporcionar una textura a zones de l'escena que potser no en tenien [9]. L'inconvenient és que en ambients amb altres fonts de llum infraroja, com una escena amb llum solar, la projecció no té efecte i el sistema és com un sistema d'estèreo passiu, només que amb el cost afegit del projector infraroig [10].

A més de càmeres amb dos sensors, existixen altres tècniques de visió tridimensional que utilitzen una sola càmera i una font de llum, com ara les mesures per temps de vol o interferometria (basades en quant tarda en tornar un puls lluminós, o amb quina fase ho fa) o la triangulació amb làser [9]. Tanmateix, malgrat tractar-se de tècniques de visió 3D, no ho són de visió estereoscòpica.

En el mercat podem trobar una àmplia gamma de càmeres estèreo. Com a exemples de models comercials llestos per a usar podem citar la sèrie RealSense d'Intel, la sèrie ZED de Stereolabs, les càmeres MYNT EYE o la càmera Tara, d'e-con Systems. A banda d'això, també existixen models sense carcassa destinats a ser integrats, tant a nivell d'usuari (com els models Arducam per a Arduino) com professional.

A més d'això, el fenomen més destacable pel que fa a càmeres estèreo comercials és la inclusió de diverses càmeres en els telèfons mòbils des de fa uns anys, ja que poden ser útils per a una certa varietat de propòsits. Alguns d'eixos propòsits són [11, 12]:

- Ajudar a obtindre una imatge ben enfocada
- Disposar de càmeres amb un camp de visió molt menor (per a fer *zoom* òptic) o molt major (gran angular)
- Aplicar l'efecte *bokeh* a les fotografies, mitjançant el qual el fons apareix desenfocat mentre que primer pla apareix nítid
- Millorar el detall de les imatges amb càmeres monocromàtiques que capten millor la llum que les càmeres RGB

- Conèixer la disposició espacial de l'escena per a aplicacions de realitat augmentada o altres, com ara filtres de privacitat per distància en videoconferències

Esta explosió en l'ús de diverses càmeres en mòbils fa que la captura tridimensional d'escenes s'haja convertit en potencialment portàtil: ja no calen muntatges aparatosos. És per això que la visió a llarg termini seria que este treball s'aplicara en última instància en un telèfon mòbil, tot i que per a això caldria que els fabricants permeteren accedir per separat a cadascun dels sensors en compte de fer el processament ells, cosa que fa uns pocs anys no passava [12] i que hui en dia encara depèn del fabricant del telèfon.

3.3. Núvols de punts i fitxers associats

Les càmeres estereoscòpiques, en calcular la profunditat de cada punt, poden situar-lo en l'espai amb tres coordenades. Un núvol de punts és el conjunt de tots els punts de la imatge (tots dels quals se'n sap la profunditat) situats en l'espai mitjançant eixes tres coordenades, i representats gràficament en un visor d'objectes tridimensionals com en la figura 3.9 poden donar una bona idea de com és l'escena captada.

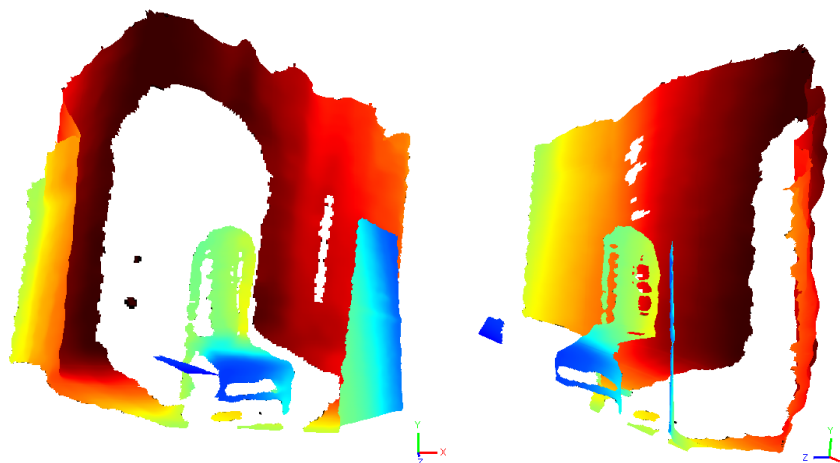


Figura 3.9: Dos vistes d'un núvol de punts bàsic

Així, podem considerar que el mostratge espacial de la realitat tridimensional dona com a producte un núvol de punts, tot i que també es poden generar núvols de punts d'altres maneres, com ara amb programes de modelatge 3D d'objectes.

Les dades que necessàriament conté un núvol de punts són les posicions dels punts en les tres coordenades espacials, però poden incorporar més informació, com per exemple el color que té cada un dels punts en la realitat o qualsevol altra propietat interessant per a l'aplicació.

Els núvols de punts, a més, poden ser organitzats o no organitzats. El fet que un núvol siga organitzat significa que els punts que conté estan ordenats d'alguna manera. Per exemple, si els punts es deriven d'alguna estructura de punts ordenats (com ara el sensor d'una càmera, on els píxels formen una graella), poden retindre eixa ordenació a nivell "lògic" i ser identificats i localitzats per les coordenades d'eixa estructura de la qual deriven. Es tracta, com diem, d'una ordenació

“lògica” perquè les coordenades de la seua posició en l’espai són independents de les coordenades de la seua estructura a nivell “lògic”. Com és natural, els núvols no organitzats no tenen eixa ordenació i consten simplement de les coordenades espacials dels punts.

Els núvols de punts s’emmagatzemen en fitxers, que són principalment de dos tipus: `.pcd` i `.ply`. Els dos formats tenen moltes semblances, ja que en essència són llistes de punts amb les seues coordenades i amb propietats predefinides o definides per l’usuari. A més, tant l’un com l’altre poden guardar les dades en text pla, perquè siga llegible, o en format binari, que ocupa menys espai d’emmagatzematge [13, 14, 15].

Tanmateix, hi ha diferències. Es nota que `.ply` sorgí del món de l’animació i el modelatge 3D en el fet que el seu concepte és el de representar sòlids com a malles de punts, i per tant pot emmagatzemar dades de les cares del sòlid de la mateixa manera que emmagatzema els vèrtexs de la malla [15]. El format `.pcd` es va desenvolupar per a la Point Cloud Library (PCL), una biblioteca per a treballar amb núvols procedents de mostrejar la realitat, i una de les grans millores que té respecte a `.ply`, que és més antic, és que ofereix suport per a núvols organitzats [13].

A més, existixen altres formats de fitxer com `.obj`, un precedent més limitat de `.ply`; `.stl`, orientat a la impressió 3D; o `.x3d`, un tipus de fitxer estàndard basat en XML [13, 14]. En este treball utilitzem el format `.ply`, ja que és el que proporciona la càmera estèreo de què hem disposat.

3.4. Estat actual de la tècnica

Tot i que la idea de la visió artificial sorgí a mitjans del segle XX i la de la visió estereoscòpica ho feu unes dècades abans, les seues possibilitats han anat ampliant-se a mesura que la capacitat de càlcul ha crescut al llarg del temps. Així, la investigació s’ha centrat en diferents aspectes al llarg de les dècades: extracció de vores, inferència de formes, reconstrucció d’escenes 3D, etc. Actualment, hi podem trobar tècniques de SLAM, aprenentatge profund o *structure from motion*, així com treball relacionat amb àrees com la imatgeria satel·lital o la biomedicina [2, 16].

SLAM (*Simultaneous Localization And Mapping*) és un camp que inclou tècniques automàtiques relacionades amb la detecció d’obstacles [17], el seguiment espacial d’objectes [18, 19], la determinació de la pròpia posició [20] o la conducció autònoma [21]. Pel que fa a la biomedicina, podem trobar tècniques de detecció de l’esquelet i les articulacions [22, 23] o de seguiment del cos en tres dimensions [24].

La investigació sobre tècniques generals relacionades amb visió artificial, visió estèreo i núvols de punts (en alguns casos per mitjà d’aprenentatge profund) se centra en aspectes com la millora dels algorismes i de la qualitat dels resultats [25, 26, 27], la fusió d’imatges [28, 29], la detecció i eliminació d’occlusions [30, 31, 32], la classificació i segmentació amb criteris geomètrics [33] o l’extracció de característiques geomètriques a baix nivell [34, 35, 36, 37, 38].

En resum, les àrees relacionades amb este treball són molt àmplies i s’hi ha dut a terme molta investigació, fonamentalment dedicada a la millora en la detecció o reconeixement d’objectes i de la seua posició i velocitat.

Centrant-nos en literatura científica relacionada de manera més directa amb els continguts del treball, podem citar, per exemple, un article que proposa usar un detector de punts singulars i seguir-ne la posició al llarg del temps per a determinar on hi ha un objecte en moviment, com un cotxe en un carrer (figura 3.10) [39].

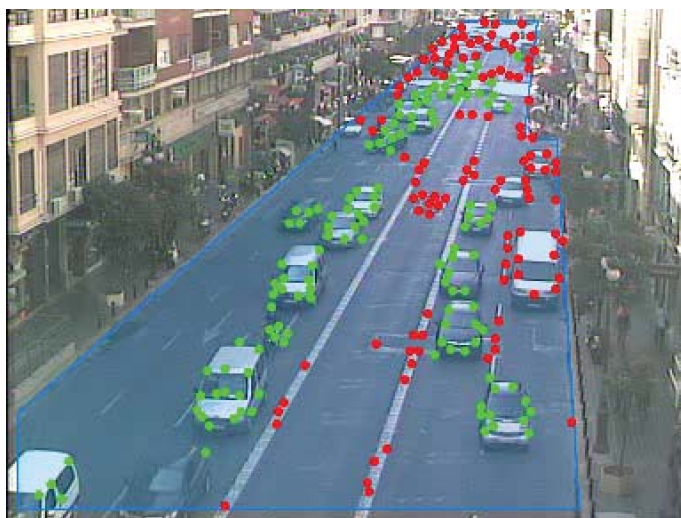


Figura 3.10: Detecció de punts singulars en cotxes en moviment [39]

Quant a la detecció i segmentació d'objectes com la barra o la màquina de Smith sencera, podem trobar paral·lelismes amb un parell d'articles que aborden l'extracció d'edificis de núvols de punts satel·litals [40] o la detecció de defectes en rodes metàl·liques [41] mitjançant refinaments de mètodes com detecció de vores, operacions morfològiques, ús d'informació espectral, segmentació per llindar adaptatiu o filtratge per característiques geomètriques com mida o relació d'aspecte.

Pel que fa al vessant esportiu del treball, les tècniques d'entrenament basades en la velocitat necessiten mesures precises de la velocitat amb què es duen a terme els exercicis per tal d'optimitzar les sessions. La conseqüència natural és que existix molta literatura en eixa intersecció entre esport i tècniques de mesura.

Concretament, existixen bastants articles que avaluen o comparen maneres o sistemes de mesurar la velocitat dels exercicis: transductors, unitats de mesura inercials (IMU), sistemes 3D multicamerals d'anàlisi de moviments o aplicacions per a mòbils mitjançant la càmera (2D). Pràcticament tots ells arriben de manera sistemàtica a les mateixes conclusions: la manera més fiable de realitzar les mesures és amb transductors lineals, després, amb els sistemes 3D multicamerals i les aplicacions, i finalment, les IMU solen ser poc fiables [42, 43, 44]. En el cas de les aplicacions, també tenen una validesa variable segons l'exercici, però són vàlides per al *back squat*, que és el que tractem en este treball [45].

El problema dels transductors és que el preu és massa elevat com per a ser permisible fora d'ambients professionals; el dels sistemes 3D, la gran aparatositat del seu muntatge; i el de les IMU, la seua no fiabilitat. Així, les proves de fiabilitat i repetibilitat de les aplicacions servirien per a facilitar i generalitzar l'accés a mesures fiables de velocitat per a poder dur a terme l'entrenament mitjançant les tècniques associades. Esta classe d'aplicacions són cada vegada més capaces de proporcionar bons resultats per les millores en els telèfons. Algunes depenen de marcadors externs, com ara Velowin [46] o un interessant sistema que descrivim a continuació en el qual ens basem per a dur a terme este treball.

Tot i que no es tracta estrictament d'una aplicació per a mòbils, en [47] trobem un sistema capaç de mesurar la velocitat de la barra transversal de la màquina de Smith a partir de vídeo. Per a aconseguir-ho, l'article proposa utilitzar marques d'un color determinat sobre la barra per a detectar

la seua posició en imatges consecutives, i unes altres marques sobre parts fixes de la màquina de Smith que es detecten mitjançant tècniques clàssiques de processament d'imatge i que permeten situar la màquina i la barra en l'espai. Açò, com s'explica més avant en el subapartat 4.4.3.4, es fa a terme mitjançant transformacions geomètriques com homografies, i requereix mesures prèvies precises per a poder funcionar, de manera que l'automatització de les mesures no és completa.

No coneixem cap investigació publicada que faça ús dels sistemes de visió estèreo (ja presents hui en dia en mòbils) i els aplique esta classe d'algoritmes. Així, la idea del treball present és millorar l'esquema de funcionament de [47] utilitzant la visió estèreo per a eliminar eixa necessitat de preparació, automatitzar el procés tant com siga possible i, per descomptat, millorar la precisió del sistema aprofitant la informació espacial que proporciona la càmera estèreo.

Capítol 4

Desenvolupament del sistema

El treball realitzat, que ha consistit en la creació i validació d'un algoritme com a part nuclear d'un sistema de visió estèreo, s'ha dut a terme per fases segons ha calgut a causa de l'ordre en què eixe algoritme fa el processament de les dades, i s'han provat sobre una maqueta que simula una màquina de Smith abans de validar el sistema com es descriu al capítol 5. Les diferents fases han anat tocant diversos aspectes de visió artificial, visió estèreo, processament d'imatges, aspectes més matemàtics o geomètrics i fins i tot temes colaterals que entrarien en el domini de la informàtica. De fet, molt sovint s'han presentat diversos d'eixos factors de manera simultània o com a vessants complementàries d'una mateixa tasca. Així doncs, en este capítol presentem la faena realitzada i els aspectes que s'han hagut de treballar i tindre en compte agrupats per temàtica per motius de claredat en l'exposició.

4.1. Eines i entorn de treball

Una de les primeres decisions fou quines eines de programari s'utilitzarien per a efectuar el processament dels núvols de punts i extraure'n la informació desitjada. Hui en dia existixen recursos com biblioteques que implementen la part rutinària i de baix nivell d'eixe processament; per tant, programar tot des de zero i a tots els nivells es descartà per complet. D'eixa manera ens vam poder concentrar més en quin processament volíem aplicar que no en implementar-lo partint del no-res.

La biblioteca estrella per a fer processament amb núvols de punts és la Point Cloud Library (PCL), ja que està específicament pensada per a este tipus de dades. PCL està escrita en C++ i es pot usar programant en eixe llenguatge. En ser un llenguatge més proper al baix nivell que altres és molt ràpid, cosa que el fa molt adequat per a processar grans quantitats de dades com és el cas dels núvols de punts. Tanmateix, C++ és molt sensible a errors en la sintaxi, en els tipus de dades, en l'ús de variables i punters, etc., de manera que es va pensar que dedicar temps a eixos aspectes secundaris podria alentir el procés de desenvolupament.

D'una altra banda, hi havia interès en utilitzar Jupyter Notebook, una plataforma de desenvolupament que inclou funcions anomenades de *notebook* o de codi en viu, és a dir, que permet mostrar gràfics interactius programats en el codi en el mateix document que el codi i simultàniament a la seua execució. Jupyter Notebook, però, està pensat per a programar en Python.

Així, es va decidir utilitzar Python per a poder treballar amb Jupyter Notebook i per a evitar els problemes de C++ mantenint l'avantatge de la rapidesa d'execució. Això va implicar buscar un *wrap* de PCL que permetera treballar amb la biblioteca des de Python. Se'n van trobar dos: Python-PCL i PCLPy. La primera té més recorregut i està basada en Cython, mentre que la segona va sorgir després i intenta millorar el rendiment de la primera treballant directament sobre les funcions de PCL en C++.

La instal·lació del programari necessari va topiar amb bastants entrebancs, i de fet no va arribar a ser exitosa, ja que requerix tindre instal·lats molts components amb una gran quantitat de dependències. L'embolic no es va resoldre tampoc amb Conda, un gestor d'entorns aïllats del sistema operatiu. Es van dur a terme moltes proves sense èxit, i finalment vam descobrir que el *wrap* Python-PCL no està mantingut des de fa anys i per això no funciona bé.

Arribats a eixe punt, es van valorar alternatives per a dur a terme el treball:

- La primera opció va ser, lògicament, PCLPy, l'altre *wrap* de PCL per a Python, que a més sí que semblava més activament mantingut. Tanmateix, a la seua pàgina avisaven que es trobava en desenvolupament actiu i que no se'n garantia el funcionament. Com que, a més, semblava estar molt poc documentada i de manera poc pedagògica, es va descartar.
- Es van trobar altres biblioteques que permeteren fer processament de núvols sobre Python sense dependre de PCL, com ara PyntCloud, PyVista, PDAL, Open3D o PPTK, però moltes permeten només visualitzar els núvols o com a molt tasques molt bàsiques de processament que no apleguen a les funcionalitats que necessitem. Només Open3D proporciona un poc més de funcions, i ni no és ni de lluny tan completa com PCL.
- Vam trobar a la web de MathWorks un paquet de visió artificial disponible per a MATLAB que inclou moltes funcions com registre i correspondència de núvols, preparació de núvols per a aprenentatge profund, extracció de característiques, correspondència i ajust a formes geomètriques, segmentació i etiquetatge automàtic de núvols, o seguiment i estimació de moviment.

Finalment, vam triar l'opció de treballar amb el paquet de visió artificial de MATLAB per ser la més completa i coneguda de les alternatives i tindre suport fiable darrere. El preu a pagar ha sigut la impossibilitat d'implementar l'algoritme en temps real, ja que MATLAB és un llenguatge molt més pesat i lent que C++. Així, tot el desenvolupament de l'algoritme ha tingut lloc en MATLAB.

Pel que fa a l'adquisició de les dades, s'ha optat per utilitzar el propi programari d'Intel per tal de facilitar les operacions i minimitzar hipotètics problemes de compatibilitat amb la càmera. Entre els programes que ofereix Intel es troba Intel RealSense Viewer, la funció principal del qual és mostrar en directe les imatges que proporciona el dispositiu. Pot mostrar qualsevol dels senyals generats per la càmera (i que descriurem més avant en l'apartat 4.2.2) com a imatges planes o en un visor 3D que permet girar, desplaçar i fer *zoom* en l'escena, i a més permet guardar les imatges i els núvols de punts respectivament, que és la funcionalitat que ens interessa. A més, s'ha utilitzat el programa Intel RealSense Dynamic Calibrator, que té un nom autoexplicatiu: s'utilitza per a calibrar la càmera estèreo amb un patró imprès o mostrat al mòbil. També s'ha fet ús de les eines de l'SDK d'Intel RealSense, executades en la línia d'ordres, per tal d'obtenir de la càmera els paràmetres de calibració que necessita saber l'algoritme, o siga, la posició i orientació relatives entre els diferents sensors.

Per últim, la visualització dels núvols de punts ja adquirits s'ha dut a terme a través del visor de CloudCompare, ja que permet un maneig més intuïtiu que el visor de MATLAB.

4.2. Càmera estèreo utilitzada

La càmera estèreo de què hem fet ús per a dur a terme este treball és la Intel RealSense D435I. Es tracta d'un model de la sèrie D400 de càmeres estèreo d'Intel que en el moment de començar el treball es venia com a una càmera de gamma bastant alta.

4.2.1. Característiques i especificacions

Esta càmera implementa visió estèreo activa mitjançant la projecció d'un patró de punts en la banda infraroja (figura 4.1) per tal de dotar de textura a possibles objectes que no en tinguen. Per això, consta d'un mòdul de profunditat format per un projector d'infrarojos i dos sensors monocroms separats 50 mm, i a banda, un sensor RGB i una unitat de mesura inercial (IMU). És de suposar que el sensor RGB porta incorporat un filtre que elimina la banda d'infrarojos, ja que el patró de punts no apareix en les imatges de color.

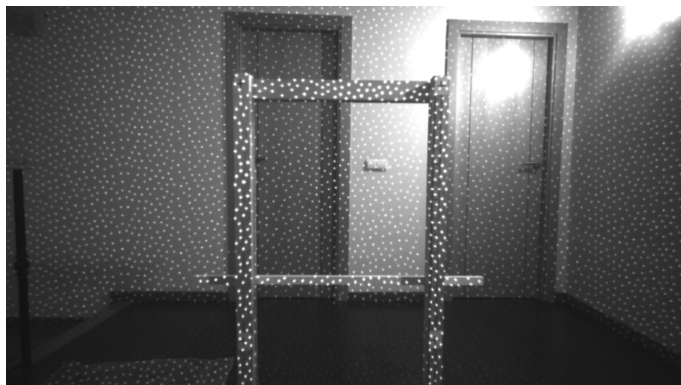


Figura 4.1: Patró de punts infrarojos

En la taula 4.1 trobem les especificacions dels sensors monocroms i del sensor RGB, i en la taula 4.2 podem veure algunes característiques de la imatge de profunditat que resulta de processar les imatges crues dels sensors monocroms.

	Sensors monocroms	Sensor RGB
Resolució	1280 px × 800 px	1920 px × 1080 px
Relació d'aspecte	8 : 5	16 : 9
Distància focal	1,93 mm	1,88 mm
Camp de visió horitzontal	91,2°	69,4°
Camp de visió vertical	65,5°	42,5°
Camp de visió diagonal	100,6°	77°
Distorsió	≤1,5 %	≤1,5 %

Taula 4.1: Especificacions dels sensors monocroms i del sensor RGB [48]

Baseline (distància entre sensors)	50 mm
Camp de visió horitzontal (HD)	$(87 \pm 3)^\circ$
Camp de visió vertical (HD)	$(58 \pm 1)^\circ$
Camp de visió diagonal (HD)	$(95 \pm 3)^\circ$
Camp de visió horitzontal (VGA)	$(75 \pm 3)^\circ$
Camp de visió vertical (VGA)	$(62 \pm 1)^\circ$
Camp de visió diagonal (VGA)	$(89 \pm 3)^\circ$

Taula 4.2: Especificacions de la imatge de profunditat [48]

De totes maneres, el camp de visió de la imatge de profunditat varia amb la profunditat mateixa (i, de fet, en la taula 4.2 ve donat per a una distància de 2 m). Açò ocorre perquè, com que els dos sensors estan separats horitzontalment i tenen diferents perspectives, al lateral exterior de cada un d'ells hi ha punts que no són captats per l'altre sensor, tal com s'ha explicat abans en la figura 3.8. Això genera una franja per a la qual no es disposa d'informació, com es mostra en la figura 4.2. Tanmateix, l'amplària d'esta franja representa una fracció més xicoteta del total de la imatge a mesura que ens allunyem de la càmera (figura 4.3).

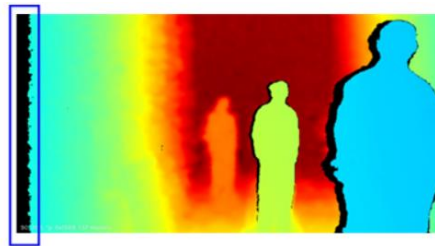


Figura 4.2: Franja sense informació al lateral d'una imatge de profunditat [48]

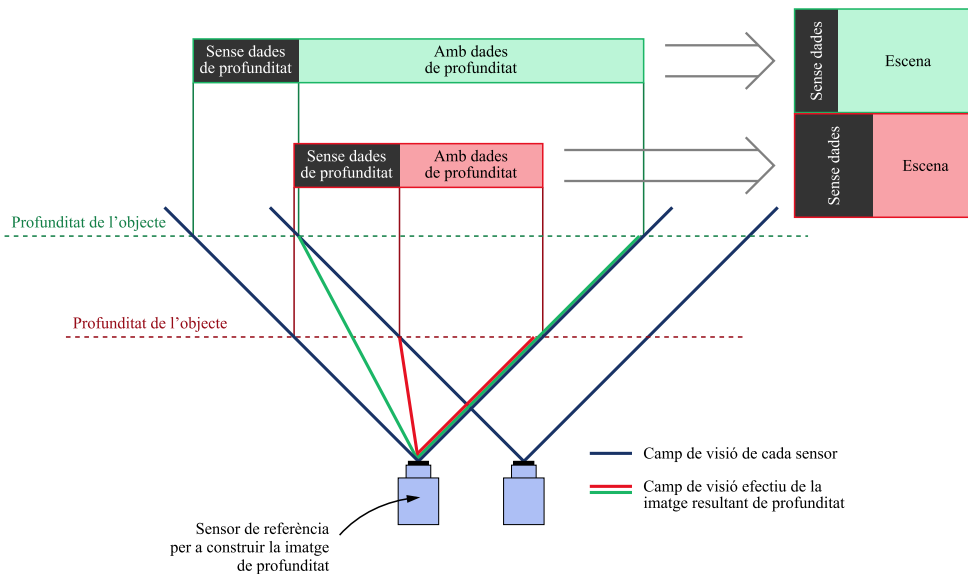


Figura 4.3: Franja sense informació a diferents distàncies

És curiosa, o estranya, la diferència notable entre el camp de visió dels sensors monocroms (i, per tant, de la imatge de profunditat) i el del sensor RGB, que és més menut, tal com s'aprecia en la figura 4.4. Quan es genera i e visualitza un núvol de punts amb propietats de color procedents del sensor RGB, este fenomen hi pot tindre uns efectes o uns altres depenent de l'entorn i la situació. En el visor de l'aplicació d'Intel, els punts que queden dins de la imatge de profunditat però fora de la imatge RGB simplement no apareixen, com s'aprecia en la figura 4.5, on trobem el mateix punt de vista sense i amb el color del sensor RGB. En canvi, si el núvol es guarda i s'obri posteriorment amb algun visor extern, com el de MATLAB o el de CloudCompare, els punts "sense color" prenen el color dels punts més propers que sí que en tenen, excepte als cantons, creant una espècie de ratlles a les vores del núvol (fig. 4.6). La informació de profunditat sí que està present de manera normal.

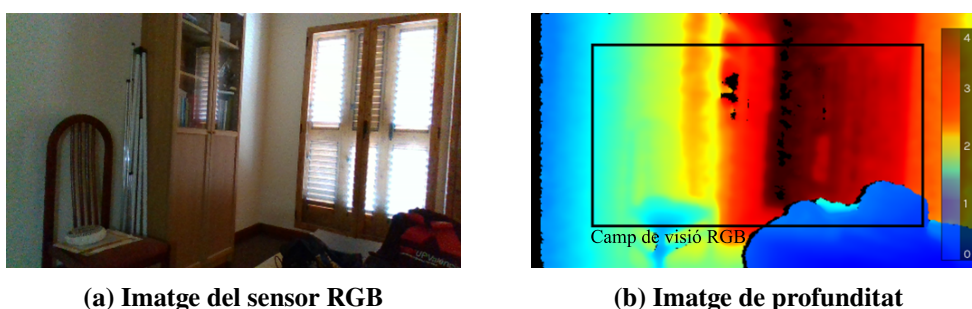


Figura 4.4: Comparació dels camps de visió dels diferents sensors en imatges 2D

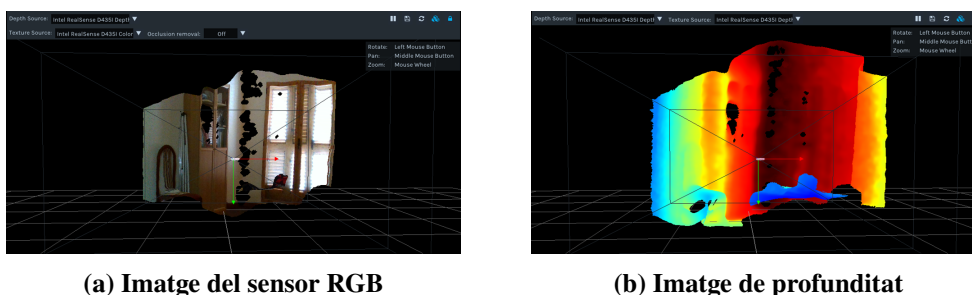


Figura 4.5: Comparació dels camps de visió dels diferents sensors en núvols de punts

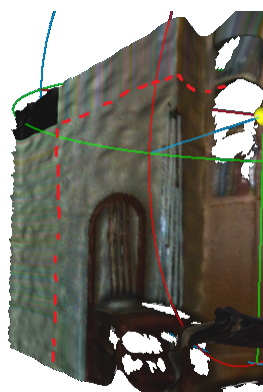


Figura 4.6: Efecte dels punts "sense color" en obrir el núvol en un visor extern

En un principi, es va pensar que açò podria constituir un problema, ja que hi havia la intenció

d'abordar la relació entre la imatge RGB i l'escena a través d'una correspondència 1 a 1 o d'alguna manera semblant, i es va plantejar que potser caldria retallar el núvol de punts o la imatge de profunditat per tal d'igualar els camps de visió de les imatges. Tanmateix, finalment això no s'ha dut a terme així, tal com explicarem més avant: com que en el moment en què processem el núvol de punts el color d'eixos punts no és important, no importa que existisquen punts sense color.

4.2.2. Fluxos de dades disponibles

Quan es connecta a un ordinador, la càmera pot proporcionar cinc senyals diferents: la imatge de color del sensor RGB, la imatge de luminància del sensor monocrom esquerre i la del dret, la imatge de profunditat calculada a partir de les de luminància, i dades del giroscopi i l'acceleròmetre. La imatge de profunditat és una imatge 2D on cada píxel té un valor de profunditat; és diferent a un núvol de punts en tres dimensions, que en este cas és calculat per l'ordinador.

A més, cada un dels fluxos de vídeo està disponible en diferents resolucions i freqüències de quadre, tal com mostra la taula 4.3. Tanmateix, cal tindre present que no és possible rebre tots els fluxos de manera simultània, a la màxima resolució i a la màxima freqüència de quadre, ja que això ve restringit per la capacitat de la connexió USB 3.0 i per la potència de processament gràfic de l'ordinador.

Flux de vídeo	Resolució (px)	Freqüència de quadre (fotogrames/s)	Format
Profunditat	1280 × 720	6, 15 i 30	Z (16 bits)
	848 × 480	6, 15, 30, 60 i 90	
	640 × 480		
	640 × 360		
	480 × 270		
	424 × 240		
Luminància (senyors monocroms esquerre i dret)	1280 × 720	6, 15 i 30	Y (8 bits)
	848 × 480	6, 15, 30, 60 i 90	
	640 × 480		
	640 × 360		
	480 × 270		
	424 × 240		
Color (sensor RGB)	1920 × 1080	6, 15 i 30	YUV 4:2:2 (16 bits)
	1280 × 720	6, 15, 30 i 60	
	960 × 540		
	848 × 480		
	640 × 480		
	640 × 360		
	424 × 240		
	320 × 240		
	320 × 180		

Taula 4.3: Combinacions de resolucions i freqüències de camp per a cada flux de vídeo [48]

Pel que fa a la captura d'imatges, les estàtiques es guarden en format PNG (i acompanyades de fitxers de dades en el cas de les imatges de profunditat), mentre que les gravacions de "vídeo" es

guarden en format ROS.

ROS és una espècie de sistema operatiu enfocat a robots. El seu funcionament es basa en nodes connectats que s'envien missatges entre si, bé siga directament o a través d'una espècie de bústies (*topics*). Dins d'estos missatges es pot enviar qualsevol tipus de dades, tot i que per l'àmbit d'ús de ROS s'utilitzen sobretot per a enviar dades de sensors. Segons de quin sensor es tracte, o en general, de què duga el missatge, el missatge té un *topic* o un altre. Els *bags* són fitxers que s'utilitzen per a emmagatzemar una sèrie de missatges transmesos.

Quan el visor d'Intel fa una gravació, l'emmagatzema en forma de *bag*. En llegir eixe fitxer amb MATLAB, s'ha vist que emmagatzema missatges que contenen la configuració completa del visor i de la càmera, els seus paràmetres i, per descomptat, les dades en si, tal com mostra la figura 4.7.

	NumMessages	MessageType	
/device_0/info	13	diagnostic_msgs/KeyValue	{'cl
/device_0/sensor_0/Depth_0/image/data	292	sensor_msgs/Image	{'
/device_0/sensor_0/Depth_0/image/metadata	7008	diagnostic_msgs/KeyValue	{'cl
/device_0/sensor_0/Depth_0/info	1	realSense_msgs/StreamInfo	{'cl
/device_0/sensor_0/Depth_0/info/camera_info	1	sensor_msgs/CameraInfo	{'
/device_0/sensor_0/Depth_0/tf/0	1	geometry_msgs/Transform	{'
/device_0/sensor_0/info	2	diagnostic_msgs/KeyValue	{'cl
/device_0/sensor_0/option/Asic_Temperature/description	1	std_msgs/String	{0XI
/device_0/sensor_0/option/Asic_Temperature/value	1	std_msgs/Float32	{0XI
/device_0/sensor_0/option/Depth_Units/description	1	std_msgs/String	{0XI
/device_0/sensor_0/option/Depth_Units/value	1	std_msgs/Float32	{0XI
/device_0/sensor_0/option/Emitter_Always_On/description	1	std_msgs/String	{0XI
/device_0/sensor_0/option/Emitter_Always_On/value	1	std_msgs/Float32	{0XI
/device_0/sensor_0/option/Emitter_Enabled/description	1	std_msgs/String	{0XI
/device_0/sensor_0/option/Emitter_Enabled/value	1	std_msgs/Float32	{0XI
/device_0/sensor_0/option/Emitter_On_Off/description	1	std_msgs/String	{0XI
/device_0/sensor_0/option/Emitter_On_Off/value	1	std_msgs/Float32	{0XI
/device_0/sensor_0/option/Enable_Auto_Exposure/description	1	std_msgs/String	{0XI
/device_0/sensor_0/option/Enable_Auto_Exposure/value	1	std_msgs/Float32	{0XI
/device_0/sensor_0/option/Error_Polling_Enabled/description	1	std_msgs/String	{0XI
/device_0/sensor_0/option/Error_Polling_Enabled/value	1	std_msgs/Float32	{0XI
/device_0/sensor_0/option/Exposure/description	1	std_msgs/String	{0XI
/device_0/sensor_0/option/Exposure/value	1	std_msgs/Float32	{0XI
/device_0/sensor_0/option/Frames_Queue_Size/description	1	std_msgs/String	{0XI
/device_0/sensor_0/option/Frames_Queue_Size/value	1	std_msgs/Float32	{0XI
/device_0/sensor_0/option/Gain/description	1	std_msgs/String	{0XI
/device_0/sensor_0/option/Gain/value	1	std_msgs/Float32	{0XI

Figura 4.7: Alguns dels *topics* d'un *bag* que conté una gravació de la imatge de profunditat (en groc, les imatges d'una gravació)

Examinant la llista de missatges del *topic* de les dades, és a dir, de la gravació pròpiament dita, observem que tots els missatges venen amb una marca de temps que indica quan s'han generat (figura 4.8), i que cada missatge correspon a un fotograma de la gravació.

També cal tindre en compte que el que s'emmagatzema, concretament en el cas de la profunditat, en fer una gravació no són núvols de punts successius (punts amb coordenades 3D), sinó imatges de profunditat successives (imatges 2D on el valor és la profunditat). Açò ha limitat les possibilitats del sistema perquè ha eliminat l'opció de fer el seguiment directament sobre el núvol.

4.2.3. Sistemes de coordenades

Tal com hem comprovat experimentalment, els punts dels núvols que genera esta càmera estan referenciats segons un sistema de coordenades que no és el que mencionàvem en la pàgina 9, sinó que, tal com mostra la figura 4.9, els eixos són, des del punt de vista de la càmera: x positiu cap a la dreta, y positiu cap a dalt i z positiu cap a arrere de la càmera (és a dir, tots els punts captats per la càmera tenen una coordenada z negativa). Les coordenades dels punts estan expressades en

Time	Topic	MessageType	FileOffset
8e-07	/device_0/sensor_0/Depth_0/image/data	sensor_msgs/Image	39963
0.032164	/device_0/sensor_0/Depth_0/image/data	sensor_msgs/Image	1892619
0.066271	/device_0/sensor_0/Depth_0/image/data	sensor_msgs/Image	3738209
0.10181	/device_0/sensor_0/Depth_0/image/data	sensor_msgs/Image	5583732
0.1331	/device_0/sensor_0/Depth_0/image/data	sensor_msgs/Image	7429255
0.16706	/device_0/sensor_0/Depth_0/image/data	sensor_msgs/Image	9274778
0.20114	/device_0/sensor_0/Depth_0/image/data	sensor_msgs/Image	11120301
0.23407	/device_0/sensor_0/Depth_0/image/data	sensor_msgs/Image	12965824
0.26647	/device_0/sensor_0/Depth_0/image/data	sensor_msgs/Image	14811347
0.30072	/device_0/sensor_0/Depth_0/image/data	sensor_msgs/Image	16656870
0.33401	/device_0/sensor_0/Depth_0/image/data	sensor_msgs/Image	18502393
0.36724	/device_0/sensor_0/Depth_0/image/data	sensor_msgs/Image	20347916
0.40092	/device_0/sensor_0/Depth_0/image/data	sensor_msgs/Image	22193439
0.43225	/device_0/sensor_0/Depth_0/image/data	sensor_msgs/Image	24038962
0.46598	/device_0/sensor_0/Depth_0/image/data	sensor_msgs/Image	25884485
0.49992	/device_0/sensor_0/Depth_0/image/data	sensor_msgs/Image	27730008
0.53215	/device_0/sensor_0/Depth_0/image/data	sensor_msgs/Image	29575531
0.56555	/device_0/sensor_0/Depth_0/image/data	sensor_msgs/Image	31421054
0.59849	/device_0/sensor_0/Depth_0/image/data	sensor_msgs/Image	33266577
0.63193	/device_0/sensor_0/Depth_0/image/data	sensor_msgs/Image	35112100
0.66646	/device_0/sensor_0/Depth_0/image/data	sensor_msgs/Image	36957623
0.69864	/device_0/sensor_0/Depth_0/image/data	sensor_msgs/Image	38803146
0.73208	/device_0/sensor_0/Depth_0/image/data	sensor_msgs/Image	40648669
0.76511	/device_0/sensor_0/Depth_0/image/data	sensor_msgs/Image	42494192
0.79995	/device_0/sensor_0/Depth_0/image/data	sensor_msgs/Image	44339715

Figura 4.8: Dades dels primers missatges del *topics* de les dades de profunditat

metres.

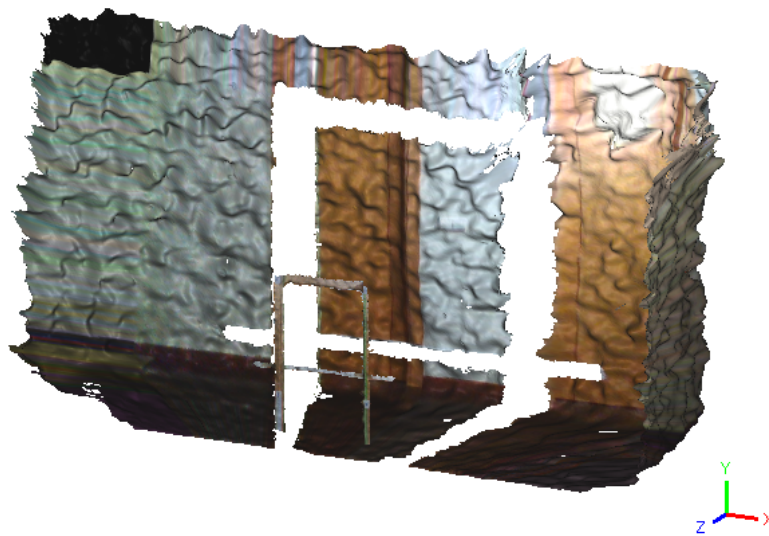


Figura 4.9: Núvol generat per la càmera estèreo i disposició del sistema de coordenades

Com que la relació entre les imatges dels dos sensors s'establix referenciant la imatge del dret respecte de la imatge de l'esquerre, l'origen de les coordenades x i y se situa al centre del sensor esquerre. Segons [48], per al model Intel RealSense D435I l'origen de la coordenada z se situa 4,2 mm cap a dins de la càmera.

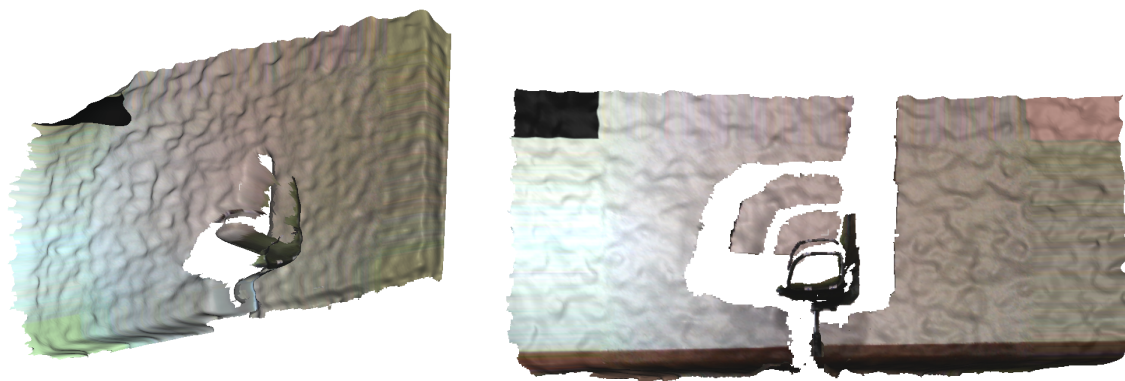
Tots estos aspectes s'han hagut de tindre en compte en programar les parts del processament geomètric de l'escena, especialment les referent als eixos dels sistemes de coordenades.

4.2.4. Qualitat del núvol de punts

Els núvols de punts que proporciona la càmera tenen una qualitat determinada (entesa com a grau de fidelitat respecte a la realitat) que respon a les característiques de la càmera mateixa, però existixen uns quants factors que l'afecten de manera notable, com les condicions ambientals lumíniques de l'escena, la seua geometria o alguns paràmetres modificables de la càmera.

4.2.4.1. Soroll i deformacions

La principal causa d'una no fidelitat entre el núvol i la realitat és l'aparició d'unions entre elements de l'escena que estan físicament separats però més o menys propers. En la nostra experiència, açò ha ocorregut sobretot amb objectes que en el núvol apareixen units al fons de l'escena. Este fenomen es va mitigar en part assegurant una distància suficient entre l'objecte i el fons, que segons es va comprovar experimentalment és d'uns 60 cm quan l'objecte es troba a uns 2 m de la càmera. En la figura 4.10 podem observar la mateixa escena amb l'objecte (una cadira) a diferents distàncies de la paret de fons.



(a) Cadira unida a la paret per estar-hi massa a prop

(b) Cadira ben definida i separada de la paret

Figura 4.10: Qualitat de dos núvols de punts d'una mateixa cadira davant d'una paret

Esta espècie d'“unions” o “rampes” apareixen en general en les vores dels objectes en les condicions mencionades i s'ha parat atenció per a evitar-les tant com ha sigut possible, ja que en certs casos poden causar problemes als mètodes de processament implementats (sobretot a la segmentació) fins al punt de convertir-los en inservibles en els casos més extrems.

La causa d'este comportament és que la càmera fa un filtratge passa baixa de l'escena com a part de la cadena de processament. Açò és *a priori* una bona idea, ja que els núvols de punts tenen tendència a ser sorollosos, i així se suavitzen i, de pas, s'omplien buits en l'escena creats per les oclusions pròpies d'una captura 3D. Tanmateix, el filtratge també crea estos inconvenients, i no és possible desactivar-lo.

Un altre fenomen que degrada la qualitat dels núvols és el soroll, que afecta a l'escena de manera pseudo-ondulatòria i és especialment apreciable en superfícies que en la realitat són planes, com ara els fons de les figures 4.9 i 4.10. Esta classe de soroll s'ha reduït lleugerament recalibrant la càmera. També trobem una altra classe de soroll més aleatori en zones com la intersecció d'objectes irregulars amb superfícies planes.

4.2.4.2. Projectió d'infrarojos i il·luminació de l'escena

Tal com s'ha explicat, en esta càmera la detecció correcta de la profunditat depèn de la projecció d'un patró d'infrarojos que aporta textura. Per tal de comprovar la magnitud d'este efecte, s'ha comparat la detecció de la profunditat en la mateixa escena amb i sense l'ús de la projecció infraroja.

La desactivació de la projecció dona com a resultat el desastre que s'observa en la figura 4.11, on es pot apreciar que les zones que més patixen són les parets, que tenen molt poca textura. Així, es comprova que el patró infraroig és irrenunciable.

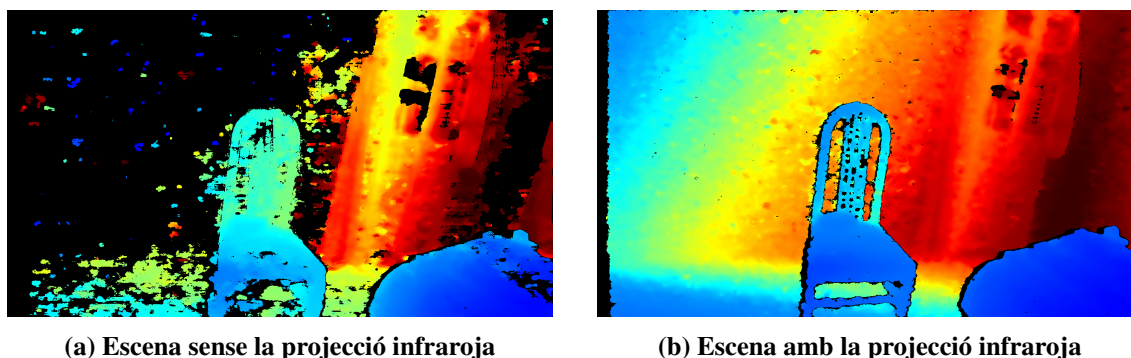


Figura 4.11: Efecte de la projecció d'un patró infraroig en la detecció de la profunditat

Precisament per això la il·luminació de l'escena ha suposat en ocasions un problema per a la captura adequada de la profunditat de l'escena: si la il·luminació és intensa, el patró infraroig destaca menys en la imatge i és més difícil de detectar. La llum artificial només té un efecte lleuger sobre la detecció de la profunditat, però la llum solar té més contingut espectral en la banda d'infrarojos i causa distorsions apreciables en el núvol de punts. Estes condicions s'han tingut en compte en les gravacions per tal d'obtindre núvols òptims.

4.2.4.3. Alineació de les imatges dels diferents sensors

El sensor de color i el sensor monocrom esquerre (la referència de la profunditat) estan separats 15 mm, cosa que suposa un canvi de perspectiva important. Per tant, va sorgir la qüestió de si la informació de profunditat i la imatge de color estarien desalineades en el núvol de punts, però s'ha comprovat que la càmera realitza les correccions oportunes com a part del processament perquè el núvol de punts proporcionat estiga tot alineat correctament. En la figura 4.12 podem veure com, malgrat el soroll, les dades de color de la maqueta que simula la màquina de Smith i les dades de color del fons estan correctament assignades als volums adequats del núvol de manera quasi exacta (és a dir, sobre el volum de la maqueta no apareix pintat el fons, ni viceversa).

4.2.4.4. Orientació de la càmera

En la càmera estèreo, el càlcul de la profunditat es basa en la disparitat. Com que els sensors estan disposats horitzontalment, la disparitat de les seues imatges també es calcula horitzontalment. Això implica que l'error és major en eixa direcció que verticalment i que els objectes més "horizontals" són més propensos a soroll i distorsions que altres.



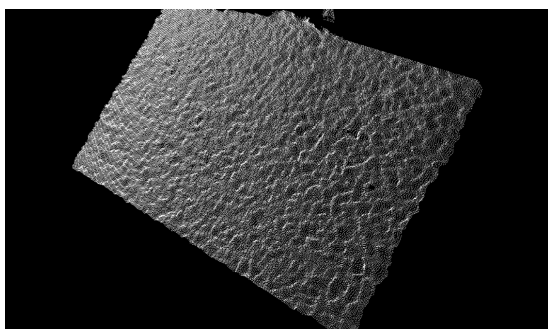
Figura 4.12: Alineació correcta de la informació de profunditat i la imatge de color

En este treball, el que necessitem del núvol de punts és, principalment, poder detectar la barra transversal de la màquina de Smith, que és horitzontal. Per tal de reduir l'error a què es veu sotmesa, es va provar (amb resultats exitosos) a prendre les captures amb la càmera girada 90° per tal que la barra aparega verticalment en la imatge de la càmera. Açò augmenta l'error en els pilars de la màquina, però no és tan greu perquè són més grossos i es detecten millor, i a més no són tan importants per a l'algoritme.

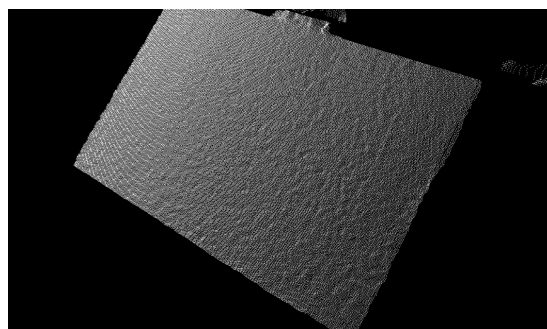
4.2.4.5. Calibració

Per tal de reduir eixe soroll i eixes distorsions al mínim possible, s'han provat dos tipus de calibració que ofereix Intel per a la càmera. Al llarg del temps és fàcil que varien algunes característiques de la càmera com les posicions d'alguns elements a nivell submil·limètric. Per tant, si fem que la càmera tinga un millor autoconeixement d'estos aspectes, pot calcular la profunditat d'una manera més fiable.

Una d'eixes dos tècniques és l'autocalibració [49], que permet fer una avaluació de la descalibració dels components de la càmera que té lloc de manera natural pel pas del temps, i corregir-la de manera més o menys senzilla i automàtica, tal com mostra la figura 4.13. Açò ha suposat una certa millora en el soroll ondulatori de les superfícies planes.



(a) Abans de l'autocalibració



(b) Després de l'autocalibració

Figura 4.13: Efecte teòric de l'autocalibració [49]

L'altra tècnica és la calibració dinàmica, que requereix un patró especial de mesures conegudes

(figura 4.14) sobre una superfície plana externa, com pot ser un tros de cartolina o la pantalla del mòbil mitjançant una aplicació d'Intel. L'eina de calibració, que s'executa en l'ordinador al qual està connectada la càmera, dona instruccions per a moure el patró per diferents zones del camp de visió i amb diferents orientacions.

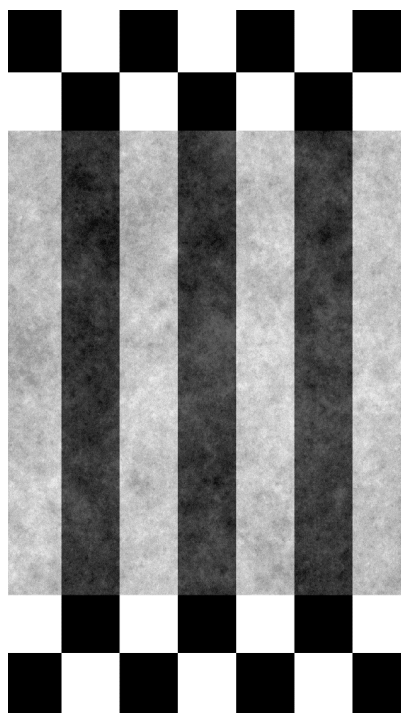


Figura 4.14: Patró utilitzat en la calibració dinàmica [50]

D'això podem deduir que es tracta d'una versió del clàssic algoritme de calibració de càmeres [51] en què es prenen diferents fotografies d'un patró, que sol ser com el d'un tauler d'escacs, i se n'extrau informació suficient per a conèixer els paràmetres intrínsecs de la càmera (el punt principal, les distàncies focals i els paràmetres de distorsió) i, en el cas de les càmeres estèreo, els paràmetres extrínsecs (posició i orientació relatives) entre els dos sensors.

En el nostre cas, la calibració dinàmica ha suposat una gran milloria en la qualitat del núvol de punts, tal com explicarem després i es mostra en la taula.

4.2.4.6. Comprovacions

Com que la qualitat del núvol de punts és important per a poder aplicar-hi les tècniques de processament, s'ha volgut estudiar bé la seua variació en funció de diversos paràmetres. Entre les proves que hem dut a terme, hi ha un xicotet estudi qualitatiu en què avaluem el soroll aleatori i la deformació de la connectivitat entre elements segons la distància i l'orientació relativa entre l'objecte (en este cas, una maqueta que simula la màquina de Smith) i la càmera. Els núvols s'han classificat:

- Com a bons si tenen ondulacions xicotetes, una connectivitat perfectament fidel a la real i interferències lleus o ocasionals amb altres objectes

- Com a regulars si les ondulacions són un poc més grans, presenten alguna discontinuïtat o patixen interferències notables
- Com a roïns si l'objecte apareix greument desfigurat

Distància i orientació càmera-objecte	0°		30°	45°	60°	
	Horitz.	Vert.	Horitz.	Horitz.	Horitz.	Vert.
1 m	Bo	Bo	Bo	Bo	Bo	Bo
1,8 m	Regular	Regular	Regular	Regular	Regular	Regular
2,5 m	Roïn	Regular	Roïn	Roïn	Roïn	Regular

Taula 4.4: Qualitat subjectiva del núvol de punts segons la distància i l'orientació relativa de la màquina i la càmera

Dels resultats, que es presenten en la taula 4.4, es dedueix que la distància és el principal factor que afecta a la qualitat del núvol de punts, i que l'orientació de la màquina no hi té efectes apreciables. Per la seua banda, la recalibració ha fet aparèixer la paret del fons en les captures de 2,5 m de distància. A més, tal com es pretenia i s'observa en la figura 4.15, col·locar la càmera en posició vertical en compte d'horitzontal millora la detecció de la barra en el núvol (i, per tant, facilita el processament posterior). Això s'ha plasmat en el fet que la distància a partir de la qual el núvol comença a empitjorar notablement s'ha vist endarrerida des de vora 1,5 m fins a vora 1,7 m.

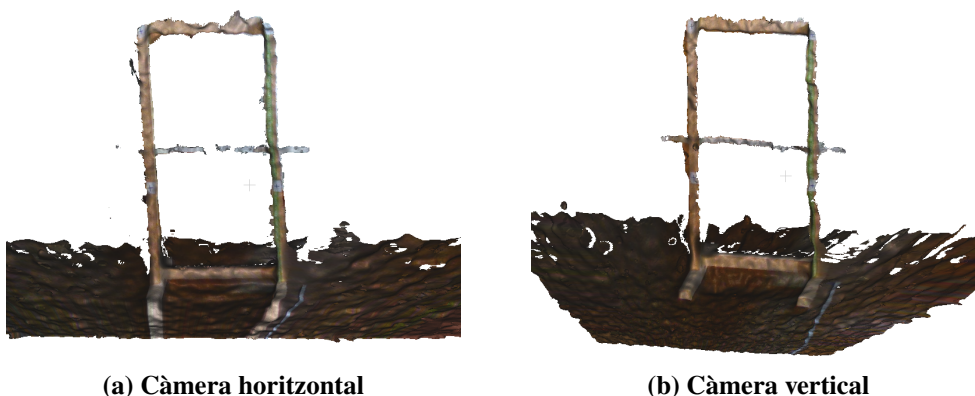


Figura 4.15: Detecció de la barra transversal de la maqueta segons l'orientació de la càmera

En qualsevol cas, hem volgut anar més enllà i quantificar esta avaluació de la qualitat del núvol de punts. Per a fer-ho, s'han disposat quatre creuetes (figura 4.16) formant un rectangle sobre l'armassó de la maqueta que simula una màquina de Smith, se n'han capturat núvols i s'ha comparat la distància entre les creuetes –i d'extrem a extrem de la barra transversal– sobre els núvols i en la realitat.

Concretament, s'han pres captures des de tres distàncies diferents (1 m, 1,8 m i 2,5 m aproximadament) amb la maqueta més o menys paral·lela a la càmera. Al llarg del desenvolupament del treball s'ha repetit esta avaluació diverses vegades a mesura que s'ha anat augmentant la resolució, recalibrant la càmera, girant-la 90°, etc. Els resultats finals, que corresponen a la configuració utilitzada en les gravacions reals, es mostren en la taula 4.5.

Els nivells d'error, que arriben al voltant d'un 2 % o un 3 %, es consideren acceptables per a la nostra aplicació, i podem dir el mateix del xicotet error d'escala que s'aprecia (ja que tots els



Figura 4.16: Marques usades en les mesures de l'error de la càmera

Situacions i distàncies	Real	Càmera a 1 m		Càmera a 1,8 m		Càmera a 2,5 m	
		Núvol	Error	Núvol	Error	Núvol	Error
Horitz. sup.	55,5 cm	56,52 cm	+1,84 %	57,59 cm	+3,76 %	57,58 cm	+3,74 %
Horitz. inf.	55,5 cm	55,94 cm	+0,80 %	56,07 cm	+1,03 %	56,73 cm	+2,22 %
Vert. esq.	72,8 cm	73,95 cm	+1,58 %	74,72 cm	+2,63 %	75,25 cm	+3,36 %
Vert. dr.	72,8 cm	73,97 cm	+1,61 %	74,92 cm	+2,92 %	75,40 cm	+3,57 %
Barra	88,2 cm	89,88 cm	+1,91 %	88,58 cm	+0,44 %	89,57 cm	+1,55 %
Mitjana			+1,54 %		+2,15 %		+2,89 %

Taula 4.5: Error en les mesures de la càmera a diferents distàncies

errors són positius). Com és lògic, són en general més grans a mesura que la distància a l'objecte augmenta.

4.2.5. Qualitat de la imatge bidimensional de color

Com que la imatge de color també afecta a l'èxit dels algorismes de processament i, al final, al funcionament del sistema, s'ha volgut determinar com li afecten una sèrie de factors per tal d'optimitzar-ne la qualitat. La imatge de color és proporcionada per un únic sensor RGB, i els seus ajustos són els d'una càmera normal (temps d'exposició, guany, etc.). Els objectius que ens hem posat pel que fa a la imatge de color són:

1. Que en cada fotograma es mostre la barra transversal de la màquina de Smith en una posició clara, és a dir, sense desenfocament de moviment (*motion blur*)
2. Que els colors es captin de manera tan fidel a la realitat com siga possible
3. Que la imatge tinga tan poc de soroll com siga possible per a optimitzar la detecció

Els objectius són eixos perquè part de l'algoritme es basa en detectar la posició de la barra transversal a partir del color d'una marca.

La naturalesa dels paràmetres de la càmera fa que el compliment total dels punts anteriors siga complicat d'aconseguir, i ens obliga a jugar buscant un equilibri entre ells. L'estratègia que hem ideat ha sigut determinar en primer lloc el major temps d'exposició que permet una detecció nítida de la barra. Com que eixe temps d'exposició és tan xicotet que enfosquix molt els colors, després s'han buscat els nivells de guany o d'il·luminació externa que proporcionen uns colors de nou semblants a la realitat.

Començant per l'ajust de l'exposició, un dels entrebancs és que ni el programa d'Intel ni cap manual ni full d'especificacions n'indica les unitats. Un valor de 10000 pareix correspondre a un temps d'exposició d'1 s, de manera que podem pensar que està en dècimes de mil·lisegon, i valors menors es corresponen amb exposicions menors. Així doncs, com que les gravacions es fan a 30 fotogrames/s, el màxim valor d'exposició per a nosaltres és 333,333. El temps d'exposició es pot ajustar en valors discrets: 312, 156, 78, 39 i 1.

S'han fet en primer lloc gravacions de la barra transversal de la maqueta en moviment en una habitació interior amb un nivell mitjà d'il·luminació, però suficient. Els resultats es mostren en la figura 4.17.

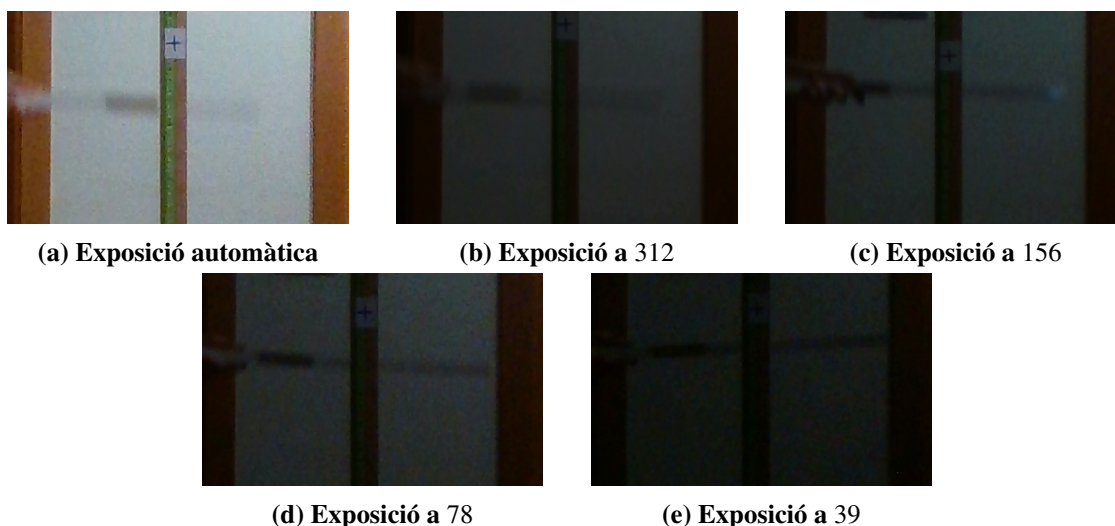


Figura 4.17: Detall de la nitidesa de la barra amb diferents valors d'exposició en interiors

Es pot comprovar que la barra ix massa borrosa amb l'exposició en automàtic, a 312 i 156, borrosa però acceptable a 78 i nítida a 39. No s'ha gravat amb l'exposició a 1 perquè fins i tot amb el guany al màxim la imatge era massa fosca com per a apreciar res. Així, els valors adequats d'exposició són 78 i 39.

A continuació, com que les imatges són molt fosques fins i tot amb el guany, s'ha gravat una altra escena al costat d'una finestra, amb resultats que es veuen en la figura 4.18.

En els dos casos la barra es pot veure amb una nitidesa i una il·luminació suficients. Evidentment, amb l'exposició a 78 es veu un poc borrosa, però pot ser suficient.

S'ha provat també en l'exterior, i amb estos valors d'exposició la imatge eixia cremada fins i tot

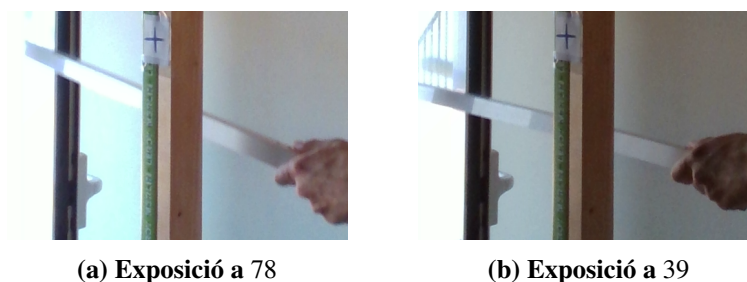


Figura 4.18: Detall de la nitidesa de la barra amb diferents valors d'exposició junt a una finestra

amb el guany mínim, de manera que ha calgut posar l'exposició al valor mínim (1). La il·luminació és acceptable i la nitidesa és molt bona (figura 4.19).



(a) Exposició a 1

Figura 4.19: Detall de la nitidesa de la barra en exteriors

De tot açò s'ha conclòs que la quantitat de llum té un efecte molt gran en la imatge i que, perquè la imatge estiga ben il·luminada amb el temps d'exposició que assegura una barra nítida, cal que hi haja llum natural, o bé bona llum artificial.

Tots estos coneixements s'han tingut en compte quan s'han fet les gravacions reals. Al gimnàs de l'UPV, el lloc on es troba la màquina de Smith és un racó molt fosc, de manera que amb el temps d'exposició a 78 i el guany al màxim la imatge quedava molt enfosquida (figura 4.20a). Això ha implicat la necessitat d'afegir un focus bastant potent al muntatge per a poder captar correctament els colors (figura 4.20b).

Tanmateix, recordem que la il·luminació addicional no és massa bona per a la detecció correcta del núvol de punts per part de la càmera estèreo (i gens bona en el cas de llum natural). Així, el que resulta necessari per a la detecció del color en la imatge 2D pot ser contraproductiu per al núvol de punts.

4.3. Tècniques i funcions de processament

L'extracció d'informació útil de l'escena per tal de reconèixer-hi la màquina de Smith implica l'ús de certes tècniques de processament sobre el núvol de punts capturat per la càmera estèreo. MATLAB n'oferix moltes, i s'han revisat i provat per a trobar les adequades per al nostre cas i adquirir experiència i intuïció en el seu ús. Concretament, s'han investigat funcions de segmentació i funcions d'ajust.



(a) Només amb la il·luminació del gimnàs i pujant molt el guany

(b) Amb il·luminació addicional i ajustos moderats

Figura 4.20: Imatge de la màquina de Smith del gimnàs de l'UPV

Les funcions de segmentació implementen la divisió del núvol en parts segons la geometria dels elements que hi apareixen, o l'extracció d'elements del núvol amb certes característiques geomètriques, que és un altre vessant del mateix. Entre altres, podem trobar una funció genèrica de segmentació per distància [52], o una d'específica per a núvols de punts procedents de lidar i que incorpora elements d'aprenentatge profund [53].

Pel que fa a l'ajust, MATLAB implementa funcions que ajusten un cos geomètric definit per paràmetres a un núvol de punts o a una zona concreta d'un núvol de punts. Els cossos disponibles per a ajustar són cilindres [54], plans [55] i esferes [56], a més d'altres definits per l'usuari mitjançant una expressió matemàtica [57]. Es pot controlar la distància màxima dels punts a l'objecte geomètric i, en els casos on té trellat, es poden posar restriccions geomètriques. Estes funcions utilitzen l'algoritme MSAC (*M-estimator sample consensus*), que és una variant del conegut RANSAC. Encara que es pot fer també amb ajust a un pla, existixen funcions per a eliminar el terra d'una escena (pensades per a lidar, això sí) [58].

4.3.1. Funcions de segmentació

Des del moment de concebre l'algoritme s'ha pretès que les tècniques de segmentació juguen el paper de detectar quins punts de l'escena corresponen a la màquina de Smith i també separar les diferents parts de la màquina (els pilars i la barra, sobretot). En efecte, això s'ha aconseguit amb una única funció de MATLAB (`pcsegdist` [52], que és simple però versàtil) utilitzada de diverses maneres i amb diferents propòsits.

La funció `pcsegdist` és una funció de segmentació per distància, és a dir, separa els punts d'un núvol per grups (o *clusters*) basant-se en la distància amb altres punts. Així, si a menys d'una certa distància d'un punt n'hi ha un altre, llavors són del mateix grup, i el grup va creixent. Si, en canvi, el punt més proper a un grup de punts es troba a una distància major que la especificada, pertany a un grup diferent.

Es tracta d'una implementació tan senzilla que l'únic paràmetre que es pot ajustar és eixa distància a partir de la qual es considera un grup diferent, i la funció retorna simplement a quin grup pertany cada punt. Així, el ventall de possibilitats s'obri quan pensem en les diferents maneres en què podem utilitzar eixa informació.

Una d'elles és reconèixer objectes filtrant per mida. Si volem detectar un objecte del qual coneixem les dimensions, podem comparar-les amb les de cadascun dels grups en què s'ha dividit el núvol i així saber quin o quins poden ser l'objecte buscat. Açò funciona perquè és difícil que aparega en l'escena un altre objecte de les mateixes mides. Tanmateix, requereix que l'objecte no estiga en contacte amb cap altra part de l'escena (és a dir, cal haver eliminat el terra en el núvol) perquè s'agrupe independentment de la resta, i les distàncies entre els punts que el formen han de ser molt diferents a les distàncies entre l'objecte i altres elements.

En el nostre treball, este ús de la segmentació s'ha aplicat en un pas clau de l'algoritme com és la detecció de la màquina de Smith en l'escena (figura 4.21). Esta ha resultat ser l'operació més sensible de l'algoritme, ja que les condicions de l'escena no complien l'últim dels requisits definits anteriorment, i això ha fet que variacions xicotetes en el paràmetre de la distància resulten en segmentacions diferents. Addicionalment s'ha utilitzat esta tècnica per a verificar que la barra transversal o els pilars detectats en la màquina ho eren realment.

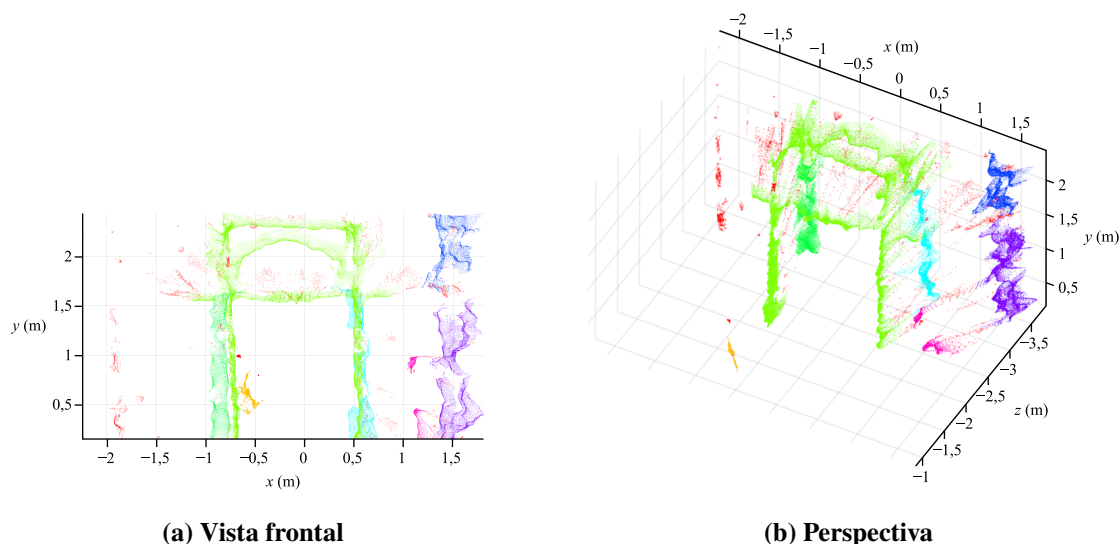


Figura 4.21: Segmentació de la zona de la màquina de Smith, on cada grup es mostra d'un color

Una segona manera en què hem utilitzat la informació de segmentació en el nostre treball és com a comprovació de l'existència de continuïtat d'objectes en una zona de l'espai. És a dir, com que la distància a la qual es troben els punts determina quants grups diferents formen, s'ha utilitzat la quantitat de grups detectats per a saber si en una determinada zona de l'espai hi ha un, dos

o més objectes. Això ha resultat útil per a determinar, una vegada aïllada la màquina, si en una determinada zona en trobem només els pilars (dos objectes separats) o els pilars units per la barra transversal (un objecte, ja que hi ha continuïtat). D'aixa manera s'ha pogut verificar la ubicació de la barra transversal.

Com ja s'ha mencionat, existixen altres funcions de segmentació [53], algunes de les quals, com `segmentGroundSMRF` [58], fins i tot s'han arribat a provar, però s'ha determinat que en el nostre cas oferixen resultats pobres o no funcionen, i s'han descartat. Esta classe de segmentació es faria en imatges 2D mitjançant operacions morfològiques, però eixa classe d'operacions no estan definides per a núvols de punts.

4.3.2. Funcions d'ajust

Mitjançant les anomenades funcions d'ajust, podem modelar elements del núvol de punts com a cossos geomètrics senzills que s'hi ajusten. És a dir, podem representar un grapat de punts amb soroll i imprecisions per una figura tridimensional caracteritzada per uns pocs paràmetres, i això ens permet comprendre millor la disposició de l'escena i processar-la tenint en compte informació més valuosa.

Eixa facilitat per a operar sobre l'escena gràcies al valor afegit dels cossos geomètrics definits per paràmetres és ja una finalitat en si mateixa i un bon motiu per a fer ús de les funcions d'ajust, i l'hem aprofitada per a modelar geomètricament els pilars i la barra transversal de la màquina, així com per a determinar l'orientació del pla que formen els dos pilars (figura 4.22). A més, també es pot utilitzar este ajust com a mètode per a detectar elements de l'escena amb certes característiques geomètriques.

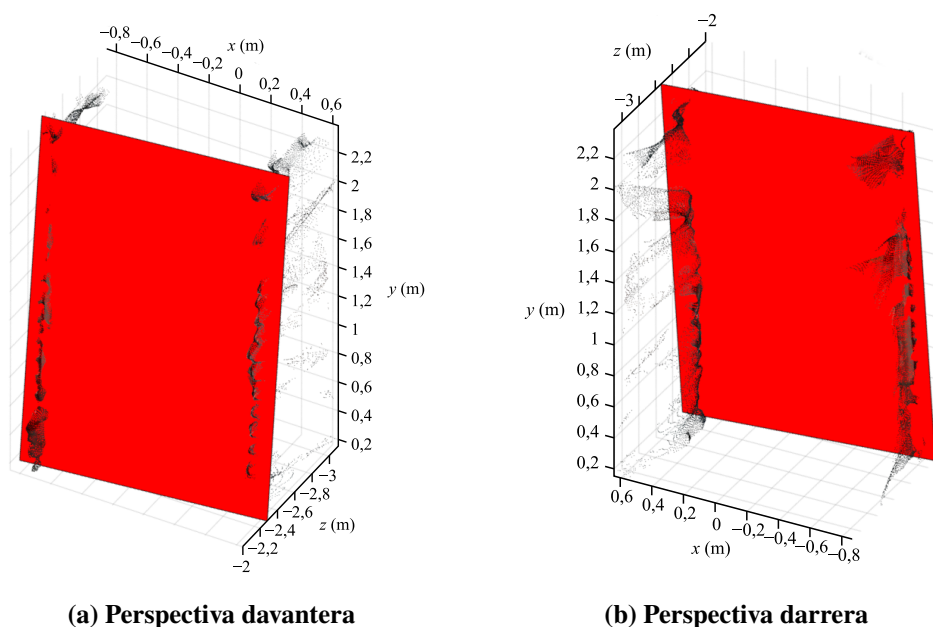


Figura 4.22: Ajust d'un pla a un núvol format pels dos pilars de la màquina

D'entre les funcions que oferix MATLAB d'ajust a una esfera, a un cilindre, a un pla i a una figura definida per l'usuari, la d'ajust a un pla (`pcfitplane`) és la que més útil ha resultat en el

nostre treball. Concretament, s'ha usat per a ajustar un pla al terra de l'escena (figura 4.23), cosa que n'ha possibilitat l'eliminació d'una manera còmoda i la rotació de l'escena per a alinear-la amb els eixos del sistema de coordenades. Així, l'ajust ha sigut important per a dur a terme un pas necessari per a la segmentació del qual parlàvem abans: eliminar la connexió entre els elements a segmentar.

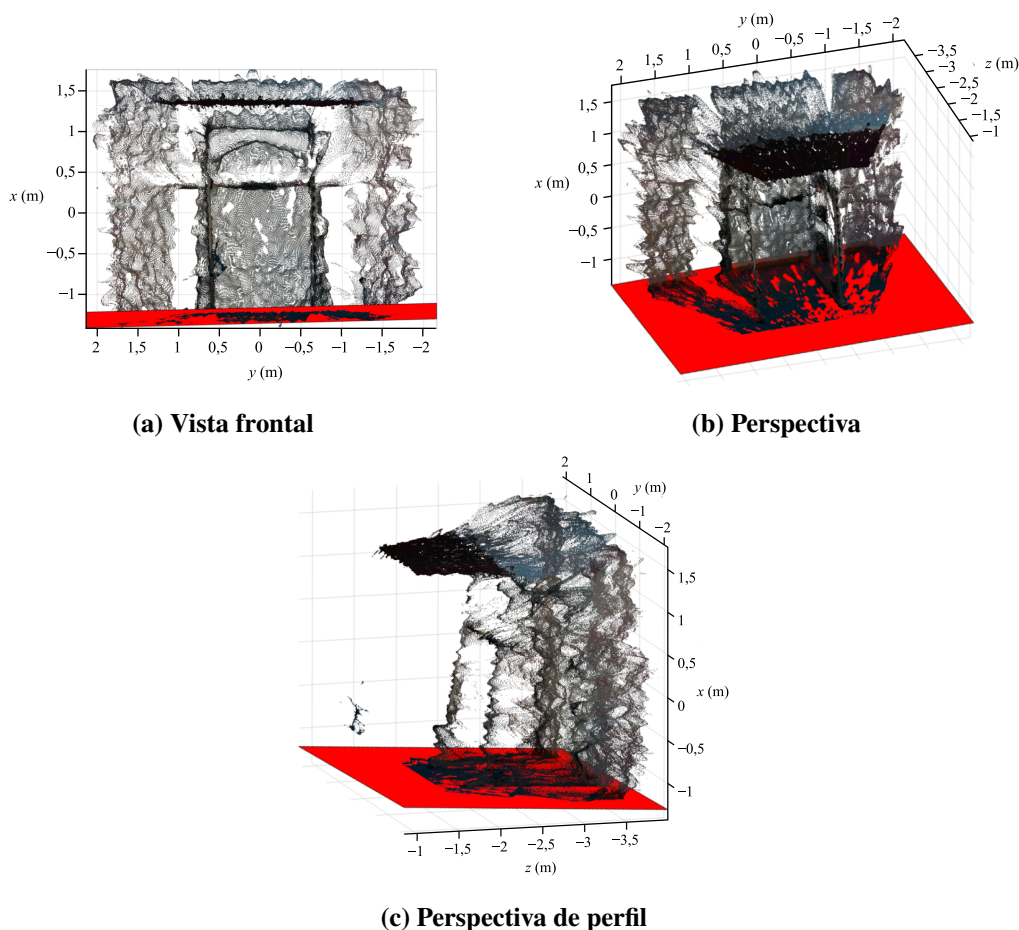


Figura 4.23: Ajust d'un pla al terra de l'escena de la màquina de Smith

Una funció que s'ha trobat a faltar entre les que ofereix MATLAB en la seua biblioteca de processament de núvols de punts és la d'ajustar punts a una recta en l'espai en comptes d'una figura amb volum. És per això que, prenent com a base el codi detallat en [59], s'ha creat una funció senzilla que determina el centroide i la direcció dominant d'un núvol de punts. En el cas d'elements molt allargats com la barra transversal i els pilars de la màquina (figura 4.24), eixa direcció dominant és molt clara i, a més, el fet de forçar que siga una recta permet eliminar les ondulacions que provoca el soroll i que dobleguen estos elements.

Per tal de trobar els extrems dels pilars i la barra, s'ha fet ús de la funció `findNearestNeighbors` de MATLAB [60], que determina quins són els punts més propers a una coordenada en el núvol, i a quina distància es troben. Eixa informació s'utilitza per a detectar el final de l'element en qüestió en el moment en què, en anar recorrent la recta que el modela, no hi ha cap punt proper a una distància menor que 10 cm.

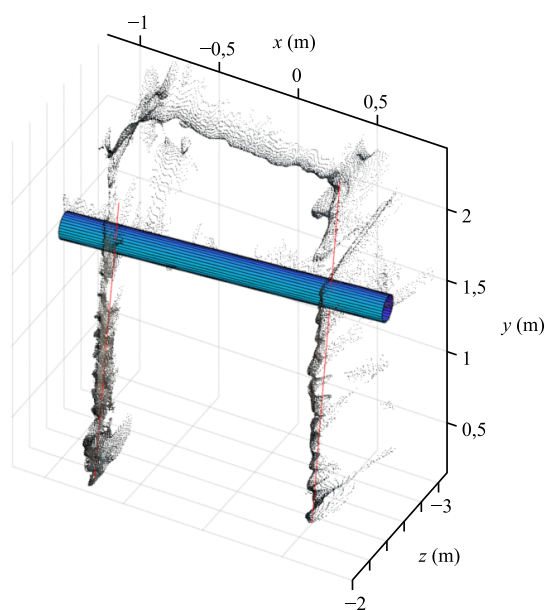


Figura 4.24: Ajust de rectes als pilars i a la barra transversal de la màquina

Finalment, i com a apunt secundari, cal mencionar que la barra transversal s'ha representat amb un model de cilindre (a diferència dels pilars de la màquina) per tal de mostrar gràficament el seu gruix. Tanmateix, este model de cilindre s'ha creat a partir dels paràmetres de la recta, no mitjançant l'ajust d'un cilindre mencionat abans.

4.4. Integració del processament 2D i 3D

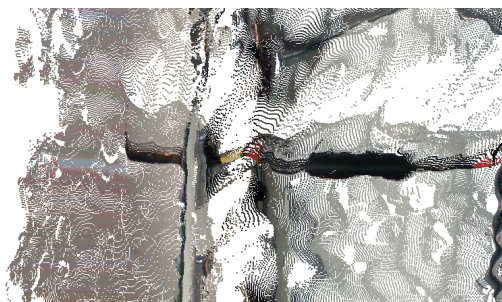
4.4.1. Decisió de seguir la barra sobre una imatge plana

Una de les parts de l'algoritme plantejat és el seguiment de la barra transversal de la màquina de Smith, que és menester per a mesurar la velocitat amb què l'esportista du a terme l'exercici. Després de reflexionar sobre les diferents maneres de dur-lo a terme, hem conclòs que fer eixe seguiment sobre una imatge plana pot donar resultats bastant millors que basar-lo en el càlcul de la posició de la barra al núvol de punts.

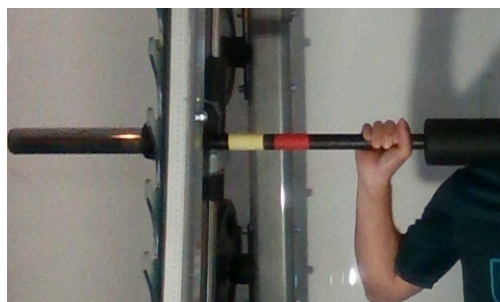
La raó principal d'això és la quantitat de soroll que conté el núvol de punts, així com les deformacions i efectes no desitjats que hem explicat abans, en el subapartat 4.2.4.1. A més, la resolució de les imatges bidimensionals que proporciona la càmera és major que la dels núvols de punts, i s'hi poden aplicar tècniques de processament que són d'ús habitual en el tractament d'imatges i que no es poden aplicar a núvols de punts per la seua naturalesa.

La conseqüència d'eixa decisió és que el treball consta de parts sobre les quals operem mitjançant processament 3D i altres on apliquem processament 2D. Això ens ha permés aprofitar les fortaleces de cada àmbit: el núvol de punts és molt millor que una imatge plana per a detectar la disposició espacial d'una escena, i a més permet fer-ho sense marques ni mesures de referència addicionals; per la seua banda, detectar formes o colors és més senzill i més precís en una imatge

plana que sobre el núvol, tal com s'aprecia en la figura 4.25.



(a) Núvol de punts (3D)



(b) Imatge bidimensional (2D)

Figura 4.25: Comparació d'una part de la màquina de Smith en les captures 3D i 2D

A l'hora de capturar les dades per a aplicar esta espècie d'algoritme mixt, ha sigut prou amb adquirir simplement les necessàries en cada àmbit. És a dir, quant a núvols de punts, només en cal un d'estàtic, i no cal que hi aparega l'esportista, ja que servix per a determinar la disposició espacial de l'escena. L'exercici en si mateix només cal gravar-lo a continuació igual que qualsevol vídeo habitual format per imatges planes, i és on se segueix la barra transversal de la màquina. La posició i l'orientació relatives entre la càmera i l'escena no han de canviar entre estes dos captures, perquè si canvien, la disposició espacial capturada al núvol ja no es correspon amb la gravada al vídeo. Això no passa quan, com en el nostre cas, s'utilitza un trípode.

Tot i dur a terme el seguiment de la barra en 2D, s'ha mantingut en l'algoritme la detecció i l'ajust de la barra transversal en el núvol (la determinació de la seua posició en l'espai, en definitiva). La raó d'això és que la posició de la barra en l'espai és clau a l'hora de determinar quin és el pla en què s'efectua el moviment de l'exercici, cosa que, com veurem un poc més avant, és menester justament per a saber amb precisió la posició de la barra durant el transcurs de l'exercici.

Una de les qüestions que ha sorgit és si la imatge 2D sobre la qual s'ha de fer el seguiment de la barra és la de color o la del sensor esquerre de profunditat. En principi, l'opció lògica és la segona, ja que el sensor esquerre de profunditat és on es troba l'origen del sistema de coordenades espacial del núvol. Tanmateix, la imatge monocroma no sempre permet detectar bé les vores dels elements, i menys encara en este cas, en què el seguiment es basa en la detecció d'una marca de color fixada a la barra. Així doncs, s'ha usat la imatge de color per a seguir la barra, i s'ha tingut en compte que el sensor RGB dista 1,5 cm del sensor esquerre de profunditat, ja que això suposa una diferència considerable en la perspectiva i fa necessari relacionar matemàticament els dos punts de vista. Per comparació, recordem que el núvol de punts recull l'escena en 3D gràcies als sensors, que estan separats 5 cm.

4.4.2. Mètode de detecció i seguiment de la barra transversal

El seguiment de la barra transversal de la màquina de Smith es basa en la detecció en cada fotograma d'una marca constituïda per una faixa de cartolina de color groc fixada a la barra a prop d'un dels pilars laterals. La detecció d'eixa marca es du a terme identificant el color dels píxels i, posteriorment, mitjançant operacions morfològiques que actuen de filtre per tal d'assegurar-se que l'element resultant correspon realment a la marca.

La funció, que s'ha inspirat en [47], pren una única imatge (en el nostre cas un fotograma de la gravació) i calcula el punt on es troba la marca. Per a tot això es té en compte informació sobre l'orientació de la càmera respecte a la màquina i la delimitació de la zona on s'ha de buscar la marca (que es determina en l'algoritme principal).

En primer lloc, es conceptualitza l'anomenada piràmide multiresolució d'imatges, on la imatge de cada nivell té la meitat de resolució que l'anterior, i s'especifica a quin nivell es farà el processament. La tria d'un nivell o un altre resulta de posar en la balança el temps de processament i la precisió de la detecció. En un cas com el nostre, en què l'algoritme està en fase de desenvolupament, prioritzem la precisió del punt detectat per a poder analitzar-lo i avaluar-ne el funcionament, de manera que s'aplica l'algoritme sobre la imatge amb la màxima resolució possible.

A continuació, es configuren paràmetres com ara la mida de la faixa de color o les seues característiques de color, es definixen la zona de la imatge on s'ha de buscar i la mida de la marca que assenyalarà el resultat, i se selecciona la zona de la imatge definida per l'àrea de cerca. Tots estos paràmetres s'adapten al nivell de la piràmide que s'haja triat i a l'orientació de la càmera en cas que no siga horitzontal.

Una vegada està tot preparat, es processa la imatge. Es descarten només començar els punts que no complixen les condicions de to, valor i saturació. En el cas del valor o brillantor, la funció és molt flexible per a poder acomodar-se a entorns amb il·luminacions diverses. A continuació, sobre les àrees restants s'apliquen operacions morfològiques per a eliminar zones amb mides que no corresponen (concretament, s'aplica un *top hat* i una obertura). Després d'això, sol quedar una àrea només, i si en queden més, s'intenta aglutinar-les.

El punt considerat la marca és el cantó superior esquerre de l'àrea, i de nou es modifica si cal per a tindre en compte que havíem reduït la mida de la imatge o que la càmera no estava en horitzontal. Opcionalment es genera una imatge amb la marca assenyalada (figura 4.26). A més, una vegada s'ha obtingut la seqüència de punts al llarg del temps, s'hi aplica un filtre de suavitzat per tal d'eliminar soroll i possibles deteccions espúries.

Tal com s'ha mencionat ja, és important que l'escena estiga ben il·luminada per a fer possible la detecció del color mantenint alhora un temps d'exposició baix, que és necessari per a captar un bon núvol 3D. Això pot implicar la necessitat d'aportar il·luminació artificial en cas que l'escena siga fosca. No és recomanable amplificar la brillantor de la imatge en l'algoritme, ja que això pot modificar-ne els colors a causa del soroll.

4.4.3. Correspondència entre la imatge bidimensional i el núvol de punts

Hem explicat abans que coneixem la disposició espacial de l'escena gràcies al núvol de punts tridimensional que obtenim de la càmera estèreo, i que el seguiment de la barra transversal de la màquina es du a terme buscant una marca de color en la imatge RGB (bidimensional), de manera que aprofitem els avantatges dels dos mons. La qüestió que sorgix, però, és com integrar estes dos parts per a poder determinar la posició de la barra en l'espai, que és el que volem conèixer en realitat: cal calcular correspondències entre la imatge RGB i el núvol de punts. La manera en què s'ha resolt això beu d'algunes de les idees de [47], així que les mostrarem abans d'explicar el nostre treball. Per tal de fer això, cal conèixer dos tipus de relacions matemàtiques entre entitats geomètriques: les projeccions i les transformacions geomètriques.

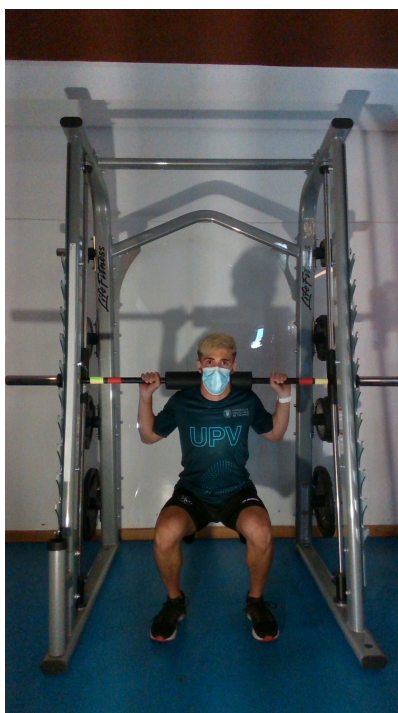


Figura 4.26: Detecció de la marca en un fotograma de la gravació

4.4.3.1. Projectió

La projecció representa la relació entre els punts de l'espai i els de la imatge. No es tracta ni més ni menys que del que s'ha explicat sobre els models de càmera i la projecció de l'escena en l'apartat 3.2.1 del capítol dels fonaments teòrics, que s'aplica ací directament i també a la inversa.

Cal tindre en compte que la projecció de punts mitjançant rajos implica que infinits punts de l'escena es projecten sobre un sol punt de la imatge, de manera que eixa projecció inversa que mencionàvem (és a dir, calcular el punt de l'escena a partir del de la imatge) no es pot dur a terme sense conèixer més dades. La informació addicional que permet fer-ho pot ser, per exemple, un pla de l'escena que sabem que conté el punt que busquem. Així, eixe punt quedaria determinat per la intersecció del pla i el raig.

4.4.3.2. Transformacions geomètriques

Les transformacions geomètriques entre dos imatges planes representen operacions com desplaçaments, rotacions, canvis de mida, deformacions, etc. que es poden aplicar sobre imatges, i matemàticament relacionen els punts d'una imatge amb els punts del resultat d'estes operacions. Una transformació geomètrica està definida per una matriu $\underline{\underline{H}}$ de 3×3 elements, tal que

$$P' = \underline{\underline{H}} \cdot P \quad (4.1)$$

on P i P' són vectors columna de 3 elements que representen, en coordenades homogènies, un punt de la imatge original i un punt de la imatge transformada respectivament. Les coordenades homogènies d'un punt tenen una dimensió més que la forma habitual de representar-lo, i tenen

la propietat que la divisió de totes les components excepte l'última entre eixa última resulta en la forma habitual (o coordenades inhomogènies) del punt [61]. Per exemple, els punts (9, 15, 3), (15, 25, 5) i (1,5; 2,5; 0,5) en coordenades homogènies representen tots el mateix punt: (3, 5). Esta és una manera molt estesa de representar punts en les transformacions geomètriques perquè permet treballar còmodament amb punts situats en l'infinit que sorgixen de vegades en estes transformacions simplement posant l'última component a 0. Així, l'equació 4.1 es desenvolupa com a:

$$\begin{bmatrix} x' \\ y' \\ w' \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (4.2)$$

i el punt P' seria, en coordenades inhomogènies (les habituals), $(x'/w', y'/w')$. La matriu $\underline{\underline{H}}$ té 8 graus de llibertat perquè està definida fins a un factor d'escala, o el que és el mateix, dos matrius proporcionals representen la mateixa transformació [61].

La matriu $\underline{\underline{H}}$, de la manera genèrica en què l'hem definida, representa una transformació de perspectiva qualsevol. Després hi tornarem. Si els elements d' $\underline{\underline{H}}$ prenen valors amb certes característiques, $\underline{\underline{H}}$ passa a representar els altres tipus de transformacions, que en realitat són casos particulars de la transformació de perspectiva [61] (figura 4.27). Estes altres transformacions són: la transformació afí o afinitat, on la imatge es pot rotar, emmirallar, moure, i escalar en diferent grau en cada direcció; la semblança, que permet rotar, emmirallar i moure la imatge, i escalar-la en totes les direccions per igual (és, per tant, un cas particular d'afinitat); la isometria, que només permet rotar, emmirallar i moure la imatge (un cas particular de semblança); i la rotació i la translació, que són casos particulars de la isometria. Els efectes d'algunes d'estes transformacions geomètriques es poden observar en la figura 4.28.

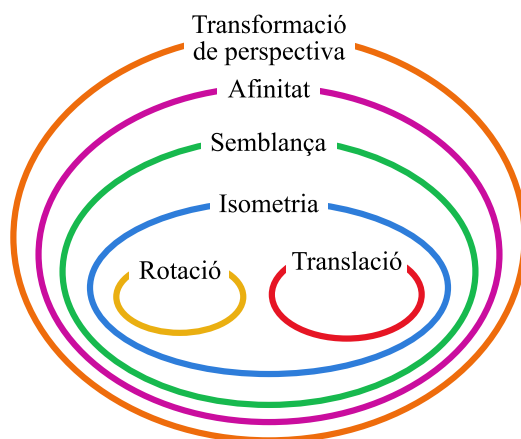


Figura 4.27: Jerarquia de les transformacions geomètriques

En qualsevol cas, ens centrem en la transformació de perspectiva genèrica, que representa una classe de distorsió que, per descriure-ho visualment, convertix els quadrats en trapezoides. Esta transformació és la que té lloc quan canviem la perspectiva que tenim sobre una escena, i el fenomen més destacat n'és que fa convergir línies paral·leles en un punt finit.

De vegades, volem saber com es veu des d'un cert lloc un pla que hi ha en una escena (o, equivalentment, punts referenciats a ell). En termes geomètrics, volem establir una relació de

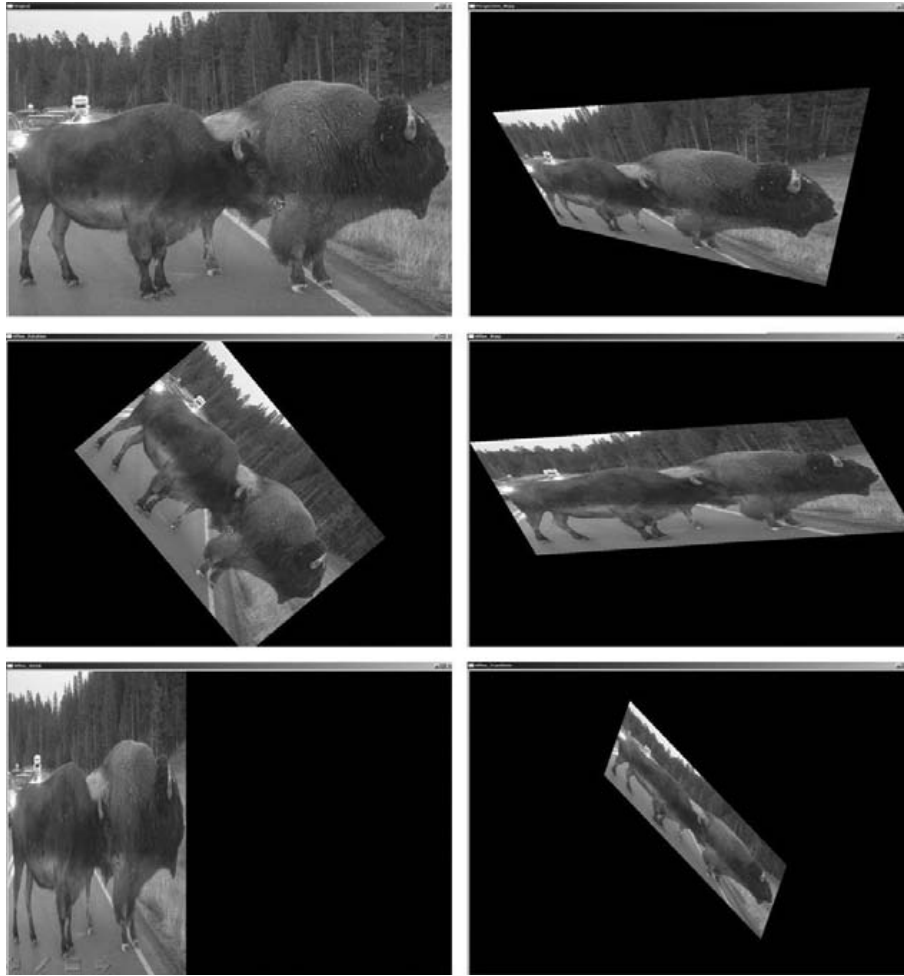


Figura 4.28: Efecte de diverses transformacions geomètriques [62]

transformació entre els punts del pla que està en l'escena i el sensor de la imatge, tenint en compte que tenen una posició i una orientació relatives qualssevol. Esta relació és una classe especial de transformació de perspectiva anomenada homografia plana, i és clau en la tasca duta a terme en [47].

4.4.3.3. Homografia plana

Podem expressar matemàticament una homografia plana d'una manera semblant a una transformació de perspectiva:

$$p = \underline{\underline{H}} \cdot P \quad (4.3)$$

Ací, $p = [x \ y \ 1]^T$ és el punt del pla imatge referit al sistema de coordenades 2D de la imatge, i $P = [X \ Y \ Z \ 1]^T$ és el punt contingut en un pla de l'espai i referit a un sistema de coordenades 3D situat en un punt arbitrari de l'escena (figura 4.29). Com que el sistema de

coordenades és 3D, el punt té coordenada Z , amb la qual sumen 4 coordenades en estar expressat en coordenades homogènies. Per tant, $\underline{\underline{H}}$ no té dimensions de 3×3 sinó de 3×4 .

La particularitat de l'homografia plana és que representa simultàniament una projecció (que té lloc en determinar com apareix en el pla imatge un punt de l'escena) i la rotació i la translació que relacionen els dos sistemes de coordenades. Açò també es reflecteix en la matemàtica, ja que la matriu $\underline{\underline{H}}$ es pot descompondre en una matriu $\underline{\underline{M}}$, que conté els paràmetres intrínsecs de la càmera i representa la primera part, una matriu $\underline{\underline{W}}$, que conté la matriu de rotació i el vector de translació que relacionen els dos sistemes de coordenades, i un factor d'escala s [63] (equacions 4.4 i 4.5).

$$p = s \cdot \underline{\underline{M}} \cdot \underline{\underline{W}} \cdot P \quad (4.4)$$

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = s \cdot \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (4.5)$$

La clau per a simplificar la matemàtica de les homografies planes, però, és col·locar el sistema de coordenades tridimensional de tal manera que el pla XY coincidisca amb el pla de l'espai que conté els punts d'interès, és a dir, que el pla que conté els punts siga $Z = 0$ en el sistema de coordenades 3D, tal com es mostra en la figura 4.29. Això, sense pèrdua de generalitat, permet eliminar una columna de la matriu $\underline{\underline{W}}$ que seria multiplicada per 0 i, consegüentment, fer que $\underline{\underline{H}}$ siga de dimensions 3×3 , tal com mostra l'equació 4.6:

$$\begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = s \cdot \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} r_{11} & r_{12} & t_1 \\ r_{21} & r_{22} & t_2 \\ r_{31} & r_{32} & t_3 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \quad (4.6)$$

Així, hem obtingut una expressió idèntica a la d'una transformació de perspectiva perquè, com que els punts d'interès pertanyen a un pla en l'espai, realment ho és. Col·locar el sistema de coordenades d'una manera convenient permet operar formalment com a tal.

4.4.3.4. Una situació semblant, però en 2D

En [47] es planteja un problema que guarda certes semblances amb el nostre. Concretament, els autors localitzen en l'espai, i segueixen, un marcador que es desplaça en un pla, però a diferència del nostre treball, disposen d'una càmera normal que genera una imatge 2D. A més, el pla en què es desplaça el marcador està parcialment ocult i és de difícil accés.

La manera en què ho han resolt és col·locant quatre marques en un altre pla paral·lel al pla en què es desplaça el marcador mòbil, i amb dos homografies i una projecció. També coneixen les posicions relatives d'estes marques fixes auxiliars.

És a dir, estableixen un sistema de coordenades 2D contingut en el pla en què es troben les marques fixes i amb origen en una d'elles, de manera que coneixen la posició de totes elles en eixe sistema de coordenades perquè l'han mesurada. Un algoritme de detecció troba eixes marques en la imatge (en el sistema de coordenades de la imatge), i així formen quatre parelles de punts

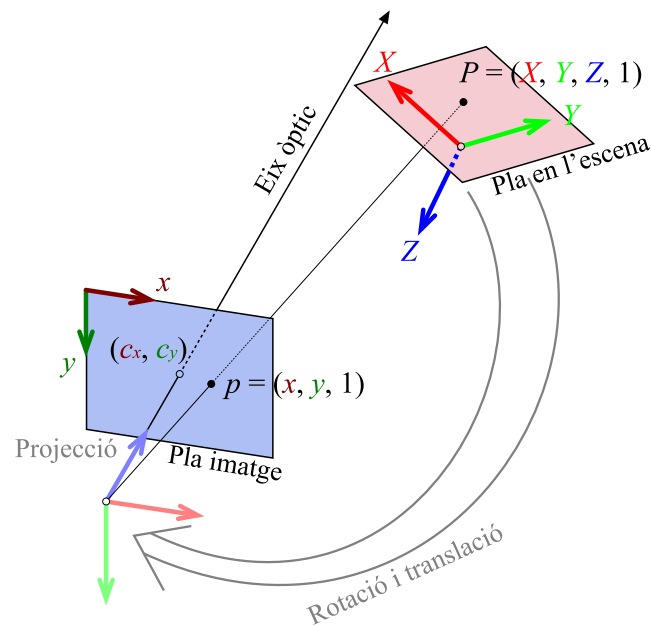


Figura 4.29: Plans i sistemes de coordenades que intervenen en una homografia plana

mitjançant els quals establixen una homografia plana entre el pla de la imatge i el que conté les marques fixes en l'escena. Però realment cal arribar al pla en què es desplaça el marcador mòbil.

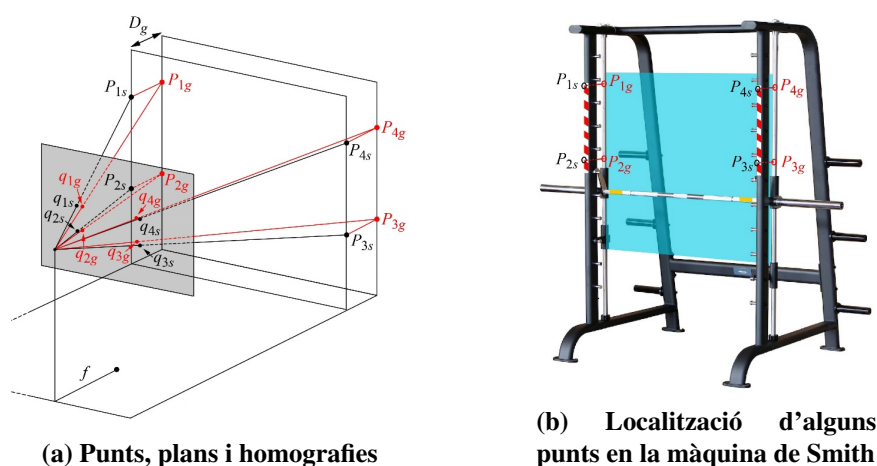
Per això, sabent els paràmetres intrínsecs de la càmera, extrauen de l'homografia plana la posició i orientació relatives dels plans, de manera que obtenen el pla de l'espai en què es troben les marques fixes, i amb tot això i una projecció inversa calculen la posició de les quatre marques fixes en el sistema de coordenades espacial de l'escena.

A continuació, determinen en quin pla de l'espai es mou el marcador ja que coneixen la distància entre este i el de les marques fixes, i prenen quatre punts en el pla del marcador mòbil (per exemple, els que queden darrere de les marques fixes). Mitjançant una projecció directa, troben en quins punts de la imatge apareixerien estos quatre punts "ficticis" de l'espai, i de nou disposen de quatre correspondències per a establir una homografia plana, esta volta sí, entre el pla de la imatge i el pla en què es desplaça el marcador. Així, esta homografia els permet calcular de manera ràpida i senzilla la posició del marcador en el sistema de coordenades del pla de moviment una volta s'haja detectat en la imatge.

La figura 4.30 mostra els punts i plans que intervenen en el procés.

4.4.3.5. Aplicació de les relacions geomètriques al nostre treball

L'element que diferencia este treball és l'ús de la càmera estereoscòpica, que suplix part de les operacions explicades en el subapartat anterior i elimina la necessitat de col·locar marques fixes i efectuar mesures per a determinar la posició de la màquina en l'espai. Les tècniques de segmentació i ajust que hem explicat en la secció 4.3 permeten determinar el pla que formen els pilars de la màquina de Smith en el sistema de coordenades de l'escena sense necessitat d'establir cap homografia plana. De [47] es pren la idea, doncs, de trobar el pla dels pilars com a pas previ



(a) Punts, plans i homografies

(b) Localització d'alguns punts en la màquina de Smith

Figura 4.30: Punts i plans que intervien en la detecció d'un marcador utilitzant una imatge 2D [47]

per a determinar el de la barra, ja que són paral·lels. De nou, la detecció de l'escena en 3D ens estalvia haver de mesurar la distància entre els dos plans, perquè gràcies a l'ajust i la segmentació detectem el centroide de la barra transversal i posem la restricció que el pla d'interés passe per eixe punt.

Una vegada coneguts el pla de la barra (o de moviment) i el punt de la imatge on s'ha detectat la marca groga de la barra, una projecció inversa (figura 4.31) ens permet saber el punt de l'espai on es troba la marca. Com que ens interessa el desplaçament vertical d'eixe punt sobre el pla de moviment, podem establir un sistema de coordenades sobre eixe mateix pla i amb un eix paral·lel a la direcció del moviment per a obtindre eixa informació, i tot seguit calcular les coordenades del punt en eixe sistema. D'eixa manera realment estem establint una homografia plana com la segona de [47], sols que desglossada per passos.

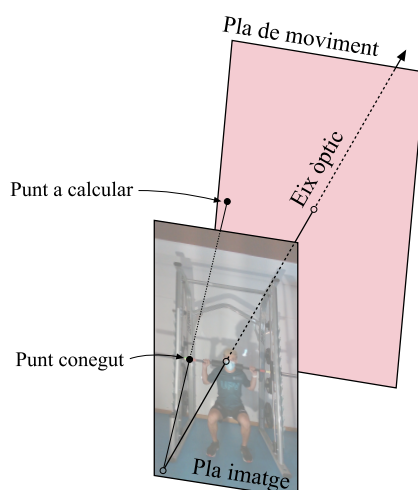


Figura 4.31: Projecció inversa que determina la posició espacial de la marca

Llavors, cal tindre en compte que en l'escena intervien tres sistemes de coordenades tridi-

mensionals (figura 4.34): el sistema que està centrat en el sensor RGB, el sistema que està centrat en el sensor esquerre de la imatge de profunditat (al qual estan referits els punts del núvol i sobre el qual es fa tot el processament relacionat amb la captura tridimensional de l'escena), i el sistema del pla de moviment de la barra (que és en el qual volem expressar els punts al final del procés). Tal com mostra la figura 4.32, este últim sistema de coordenades ve determinat per tres plans: el de moviment de la barra, el de terra i un pla auxiliar perpendicular als altres dos i que passa pel centroide del pilar esquerre. En concret, l'origen de coordenades es troba en la intersecció dels tres plans, l'eix x correspon a la intersecció entre el pla de la barra i el de terra i és positiu cap a la dreta, l'eix y correspon a la intersecció entre el pla de la barra i l'auxiliar i és positiu cap a dalt, i l'eix z s'obté com a producte vectorial dels altres dos aplicant la regla de la mà dreta.

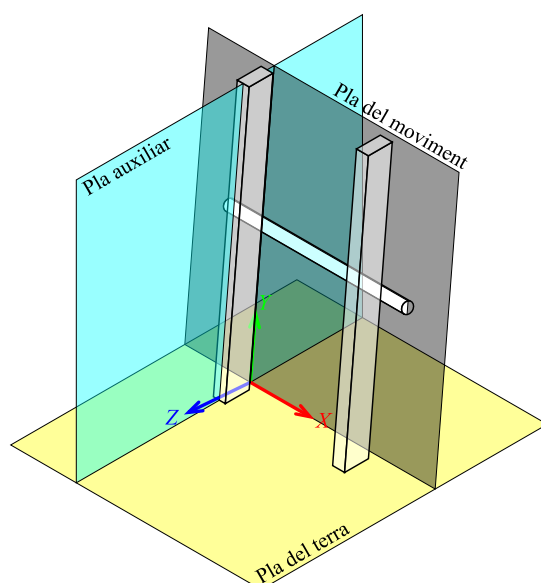


Figura 4.32: Tria del sistema de coordenades del pla de moviment de la barra transversal

Per sort, la càmera proporciona, mitjançant l'ordre `rs-enumerate-devices -c` executada en la carpeta d'eines de l'SDK d'Intel RealSense, una gran quantitat de paràmetres entre els quals es troben la matriu de rotació $\underline{R}_{RGB \rightarrow prof}$ i el vector de translació $\vec{T}_{RGB \rightarrow prof}$ que permeten “traduir” un punt de les coordenades del sensor de profunditat a les del sensor RGB.

Malgrat que pugui semblar que la nomenclatura siga la contrària, açò té una explicació. El sistema de coordenades del qual partim és el del sensor de profunditat i, per tant, és en l'únic en el qual podem saber les coordenades dels punts. Si sabem les coordenades d'un punt en eixe sistema i volem saber-les en el del sensor RGB, l'única manera de fer-ho és aplicar-li al punt una rotació i una translació que facen que tinga la mateixa posició respecte al sistema de coordenades de referència (el del sensor de profunditat) que la que tenia abans respecte al sistema de coordenades del sensor RGB. La figura 4.33 il·lustra esta idea.

Així, en primer lloc, cal expressar el pla en què es mou la barra (obtingut amb els algorismes de processament de l'escena) en el sistema de coordenades de la càmera RGB. Per a això, prenem la normal \vec{N}_{prof} i un punt qualsevol Q_{prof} del pla –que el determinen per complet– i els transformem al nou sistema de coordenades segons les equacions 4.7 i 4.8:

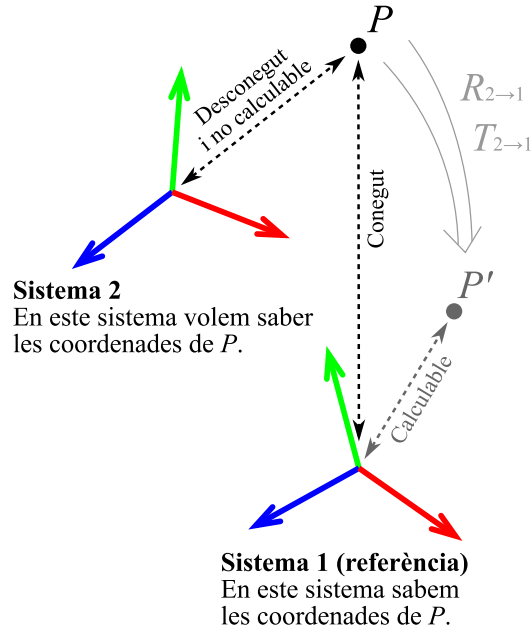


Figura 4.33: Obtenció de les coordenades d'un punt en un altre sistema de referència

$$\vec{N}_{RGB} = \vec{N}_{prof} \cdot \underline{\underline{R}}_{RGB \rightarrow prof} \quad (4.7)$$

$$\underline{\underline{Q}}_{RGB} = \left(\underline{\underline{Q}}_{prof} + \vec{T}_{RGB \rightarrow prof} \right) \cdot \underline{\underline{R}}_{RGB \rightarrow prof} \quad (4.8)$$

A partir de la nova normal $\vec{N}_{RGB} = [A \ B \ C]$ i el nou punt pertanyent al pla $\underline{\underline{Q}}_{RGB} = [x_0 \ y_0 \ z_0]^T$, es pot operar com mostra l'equació 4.9 per a arribar a l'equació general del pla (4.10) en les noves coordenades.

$$\vec{N}_{RGB} \cdot \left(\begin{bmatrix} x \\ y \\ z \end{bmatrix} - \underline{\underline{Q}}_{RGB} \right) = \vec{0} \quad (4.9)$$

$$Ax + By + Cz + D = 0 \quad (4.10)$$

L'altre ingredient és un punt de la imatge RGB, que està expressat en el sistema de coordenades de la imatge, no el de la càmera RGB. El sistema de la càmera RGB, en el qual estem expressant tots els elements rellevants, és tridimensional i, segons el criteri de la càmera utilitzada, amb x positiva cap a la dreta, y positiva cap a dalt i z positiva cap arrere, i té l'origen en el centre òptic del sensor. El sistema de la imatge RGB és bidimensional, centrat en el cantó superior esquerre de la imatge i amb x positiva cap a la dreta i y positiva cap avall. El pla de la imatge és paral·lel al pla XY del sistema de la càmera RGB. Els diferents sistemes de coordenades es mostren en la figura 4.34.

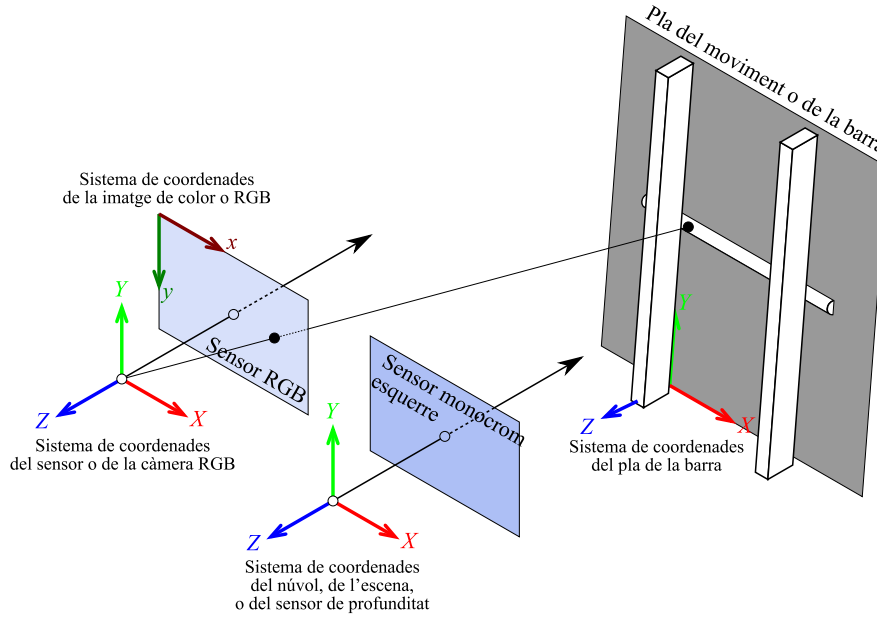


Figura 4.34: Sistemes de coordenades que intervenen en la integració del processament 2D i el 3D

Per tal d'expressar un punt de la imatge en les coordenades espacials de la càmera RGB, s'ha de tindre en compte l'origen del sistema de referència de la imatge aplicant un desplaçament per valor del punt principal $\mathbf{c} = [c_x \ c_y]$, s'ha d'invertir la coordenada y per a plasmar que l'eix y té un sentit diferent en cada sistema de coordenades, i cal afegir la dimensió addicional del sistema de la càmera RGB, que pren el valor de la coordenada z on se situa el pla de la imatge, és a dir, la distància focal f sobre la meitat negativa d'eix z . Així, un punt original de la imatge

$$\mathbf{p}_{img} = [x_{img} \ y_{img}] \quad (4.11)$$

es pot expressar en coordenades de la càmera RGB com a un punt

$$\mathbf{P}_{RGB} = [x_{RGB} \ y_{RGB} \ -f] \quad (4.12)$$

on

$$[x_{RGB} \ y_{RGB}] = [x_{img} - c_x \ -(y_{img} - c_y)] \quad (4.13)$$

Coneixent el punt de la imatge i sabent que tots els rajos passen pel centre òptic, que es troba en $(0, 0, 0)$, tracem el raig, que no és més que una recta. El seu vector director $\vec{\mathbf{v}}$, doncs, és

$$\vec{\mathbf{v}}_{dir} = \frac{[x_{RGB} \ y_{RGB} \ -f] - [0 \ 0 \ 0]}{\|[x_{RGB} \ y_{RGB} \ -f] - [0 \ 0 \ 0]\|} = \frac{[x_{RGB} \ y_{RGB} \ -f]}{\|[x_{RGB} \ y_{RGB} \ -f]\|} \quad (4.14)$$

de tal manera que l'equació vectorial de la recta que representa el raig és

$$[x \ y \ z] = \vec{\mathbf{v}}_{dir} \cdot t = \left[\frac{x_{RGB}}{\|[x_{RGB} \ y_{RGB} \ -f]\|} t \ \frac{y_{RGB}}{\|[x_{RGB} \ y_{RGB} \ -f]\|} t \ \frac{-f}{\|[x_{RGB} \ y_{RGB} \ -f]\|} t \right] \quad (4.15)$$

Si coneixem el pla on està el punt que busquem de l'escena (equació 4.10) i també el raig resultant de projectar-lo des de la imatge RGB, podem trobar el punt com a la intersecció d'estos dos elements:

$$A \frac{x_{RGB}}{\| [x_{RGB} \ y_{RGB} \ -f] \|} t + B \frac{y_{RGB}}{\| [x_{RGB} \ y_{RGB} \ -f] \|} t + C \frac{-f}{\| [x_{RGB} \ y_{RGB} \ -f] \|} t + D = 0 \quad (4.16)$$

Açò es pot resumir, fent ús de la notació vectorial, com a

$$\left(\vec{v}_{dir} \cdot \vec{N}_{RGB} \right) \cdot t + D = 0 \quad (4.17)$$

D'ací podem aïllar t , que representa la distància des del punt de referència (l'origen, en el nostre cas) a la qual es troba el punt:

$$t = \frac{-D}{\vec{v}_{dir} \cdot \vec{N}_{RGB}} \quad (4.18)$$

Així, el punt de l'escena P_{RGB} resultat de projectar la marca de la barra sobre el pla de moviment de la barra s'obté substituint l'equació 4.18 en l'equació 4.15:

$$P_{RGB} = \vec{v}_{dir} \cdot \frac{-D}{\vec{v}_{dir} \cdot \vec{N}_{RGB}} \quad (4.19)$$

El punt obtingut està expressat en el sistema de coordenades de la càmera RGB, i l'objectiu final és referir-lo al del pla de la barra. Com que es coneix la relació entre el sistema de la càmera RGB i el del sensor esquerre de profunditat, i es pot calcular la relació entre este últim i el sistema del pla de la barra, el canvi de sistemes de coordenades s'ha dut a terme per passos.

Per a "traduir" els punts del sistema de la càmera RGB al del sensor esquerre de profunditat, és prou amb dur a terme l'operació contrària a 4.8. De nou, cal tindre present que la matriu de rotació i el vector de translació es denoten amb subíndexs contraris a l'operació que aparentment es vol realitzar perquè, per a passar les coordenades dels punts del sistema del sensor RGB al sistema centrat en el sensor de profunditat, cal fer que tinguin les mateixes posicions respecte al sistema de referència (el del sensor RGB) que la que tenien abans respecte al sistema de coordenades del sensor de profunditat.

Sabem que la matriu de rotació s'obté de la manera següent:

$$\underline{\underline{R}}_{prof \rightarrow RGB} = \underline{\underline{R}}_{RGB \rightarrow prof}^{-1} \quad (4.20)$$

Per analogia amb l'expressió 4.8, els punts s'han d'obindre com a:

$$P_{prof} = \left(P_{RGB} + \vec{T}_{prof \rightarrow RGB} \right) \cdot \underline{\underline{R}}_{prof \rightarrow RGB} \quad (4.21)$$

però justament operant a partir de la mateixa expressió 4.8, arribem a:

$$\mathbf{P}_{prof} = \mathbf{P}_{RGB} \cdot \underset{prof \rightarrow RGB}{\mathbf{R}} - \vec{\mathbf{T}}_{RGB \rightarrow prof} \quad (4.22)$$

Per comparació de les expressions 4.21 i 4.22, i tenint en compte l'expressió 4.20, arribem a la conclusió que:

$$\vec{\mathbf{T}}_{prof \rightarrow RGB} = -\vec{\mathbf{T}}_{RGB \rightarrow prof} \cdot \underset{RGB \rightarrow prof}{\mathbf{R}} \quad (4.23)$$

Les expressions 4.20 i 4.23 ens permeten calcular l'expressió dels punts de l'espai en el sistema de coordenades centrat en el sensor de profunditat esquerre, i ho fem de la manera que es mostra en l'expressió 4.21.

En el següent pas, que és expressar els punts en el sistema de coordenades del pla de la barra, cal saber també la relació entre este sistema i el centrat en el sensor esquerre de profunditat, és a dir la matriu de rotació i el vector de translació que els relacionen.

La matriu de rotació es pot calcular resolent un sistema d'equacions de 9 incògnites que queda determinat amb tres correspondències entre els dos sistemes de coordenades després de fer-ne coincidir l'origen. Per a facilitar la tasca a nivell conceptual, s'han pres els vectors unitaris que formen la base ortonormal dels sistemes de coordenades. És a dir, per part del sistema de coordenades del sensor de profunditat (en el qual estan expressats fins ara els punts d'interés), prenem $\hat{\mathbf{x}}_{prof} = (1, 0, 0)$, $\hat{\mathbf{y}}_{prof} = (0, 1, 0)$ i $\hat{\mathbf{z}}_{prof} = (0, 0, 1)$, i per part del sistema de coordenades del pla de la barra, prenem els vectors unitaris dels eixos x , y i z però expressats segons les coordenades del sistema del sensor de profunditat. Açò no presenta cap problema, ja que eixos vectors s'han calculat, tal com s'ha explicat abans, com a intersecció de plans coneguts. Estos vectors els denotarem per $\hat{\mathbf{x}}_{barra} = (x_x, x_y, x_z)$, $\hat{\mathbf{y}}_{barra} = (y_x, y_y, y_z)$ i $\hat{\mathbf{z}}_{barra} = (z_x, z_y, z_z)$ respectivament. Així, les equacions queden de la següent manera:

$$\begin{bmatrix} \hat{\mathbf{x}}_{barra} \\ \hat{\mathbf{y}}_{barra} \\ \hat{\mathbf{z}}_{barra} \end{bmatrix} \cdot \underset{barra \rightarrow prof}{\mathbf{R}} = \begin{bmatrix} \hat{\mathbf{x}}_{prof} \\ \hat{\mathbf{y}}_{prof} \\ \hat{\mathbf{z}}_{prof} \end{bmatrix} \quad (4.24)$$

Desenvolupant:

$$\begin{bmatrix} x_x & x_y & x_z \\ y_x & y_y & y_z \\ z_x & z_y & z_z \end{bmatrix} \cdot \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (4.25)$$

Les incògnites serien, doncs, els 9 elements de la matriu $\underset{barra \rightarrow prof}{\mathbf{R}}$.

Per la seua banda, el vector de translació es calcula com a la diferència entre els orígens dels dos sistemes de coordenades ($\mathbf{O}_{barra} = (O_x, O_y, O_z)$ i $\mathbf{O}_{prof} = (0, 0, 0)$):

$$\vec{\mathbf{T}}_{barra \rightarrow prof} = \mathbf{O}_{prof} - \mathbf{O}_{barra} = [0 \ 0 \ 0] - [O_x \ O_y \ O_z] \quad (4.26)$$

Una vegada coneguts la matriu de rotació i el vector de translació, es procedix de la mateixa manera que en l'expressió 4.21 per tal d'obtindre les coordenades dels punts en el sistema del pla de la barra. De nou, cal recordar que la matriu de rotació i el vector de translació es denoten amb

subíndexs contraris a l'operació que aparentment es vol realitzar perquè, per a passar les coordenades dels punts del sistema del sensor de profunditat al sistema del pla de la barra, cal fer que tinguin les mateixes posicions respecte al sistema de referència (el del sensor de profunditat) que la que tenien abans respecte al sistema de coordenades del pla de la barra.

Així doncs, esta seria l'expressió:

$$P_{barra} = \left(P_{prof} + \vec{T}_{barra \rightarrow prof} \right) \cdot \underline{R}_{barra \rightarrow prof} \quad (4.27)$$

D'estos punts expressats en el sistema de coordenades de la barra és d'interés la component y (la vertical), que és la que determina l'altura de la barra i la que intervé en magnituds com la força o la potència desenvolupades per l'esportista.

4.5. Magnituds d'interés

L'objectiu final és proporcionar a l'esportista magnituds rellevants relacionades amb el seu rendiment esportiu de cara a optimitzar les rutines d'entrenament. Les magnituds que interessin de cara a l'aplicació (la velocitat i la força i la potència aplicades per l'esportista) es poden calcular a partir de la posició de la barra que hem obtingut abans amb les projeccions.

4.5.1. Posició, velocitat i acceleració

Mitjançant la projecció i l'homografia hem obtingut la posició al llarg del temps de la marca groga fixada a la barra en el sistema de coordenades que hem determinat i que té el pla XY coincident amb el pla de moviment de la barra. Este sistema de coordenades és el que utilitzem per als càlculs de les diferents magnituds, i s'ha triat de manera que el moviment té lloc, per les restriccions de la màquina de Smith, exclusivament al llarg de la dimensió y , amb components nul·les en les direccions x i z .

Extraiem, per tant, eixa segona component de la posició i , per tal d'unificar referències amb la validació, prenem com a posició 0 la inicial (abans de començar a baixar per a realitzar l'exercici), de manera que la posició es fa més negativa com més baixa l'esportista (equació 4.28). Per comoditat, el resultat d'això és el que a partir d'ara anomenem "posició" i denotem per y (y_i si fem referència al seu valor en l'instant i).

$$y_i = P_{barra,y,i} - P_{barra,y,0} \quad (4.28)$$

Com que a més del vector de posició al llarg del temps disposem també d'un vector amb els temps corresponents a eixos valors de posició, podem calcular la velocitat instantània. La velocitat és la derivada de la posició respecte al temps, però com que les dades de què disposem són en temps discret, l'hem calculada amb diferències finites centrades (equació 4.29). No és un vector sinó un escalar perquè, de nou, la velocitat només té una component no nul·la en la direcció y del sistema de referència (la direcció del moviment).

$$v_{y,i} = \frac{y_{i+1} - y_{i-1}}{t_{i+1} - t_{i-1}} \quad (4.29)$$

En altres treballs on s'utilitzen esta classe de tècniques de detecció és comú aplicar als resultats de les derivades algun tipus de filtratge per a reduir el soroll que hi sol aparéixer. Com que això també ocorre en el nostre cas, hem aplicat a la velocitat un filtre que en el nostre cas és de tipus Savitzky-Golay.

L'acceleració s'ha calculat a partir de la velocitat de la mateixa manera que la velocitat a partir de la posició, tal com mostra l'equació 4.30, i també se li ha aplicat un filtre Savitzky-Golay.

$$a_{y,i} = \frac{v_{y,i+1} - v_{y,i-1}}{t_{i+1} - t_{i-1}} \quad (4.30)$$

4.5.2. Càlcul de forces

Com que coneixem l'acceleració, la 2a llei de Newton ens permet calcular la força resultant $\vec{F}_{res,i} = [0 \quad F_{res,y,i} \quad 0]$ que actua sobre la barra en cada instant i que és responsable del seu moviment (equació 4.31). La massa m és una dada d'entrada, i és la suma de la massa de l'esportista, la barra i les peses. Això vol dir que, tal i com es fa habitualment en ciències de l'esport, modelem el problema com si l'esportista, la barra i les peses foren tot un conjunt que segueix el mateix desplaçament que realitza la barra i està sotmés a les mateixes forces que ella.

$$F_{res,y,i} = m \cdot a_{y,i} \quad (4.31)$$

Tanmateix, el que ens interessa és la força que realment exercix l'esportista, que no és l'única existent, ja que la força de la gravetat també té un efecte sobre la barra. Considerem que tant la força de la gravetat com la de l'esportista són perpendiculars a terra, tenen sentits oposats i els seus valors determinen el moviment de la barra. En algunes màquines de Smith, pot ser que el pla de moviment de la barra no siga perpendicular a terra, sinó que forme un angle amb ell per les restriccions de la geometria de la màquina. Això vol dir que, en un cas general, les dos forces mencionades poden descompondre's en una component paral·lela a la direcció de moviment (és a dir, en la direcció y del nostre sistema) i una component perpendicular al pla de moviment (és a dir, en la direcció z). No hi ha component lateral o x . Només les components y de les forces tenen algun efecte en el moviment de la barra, mentre que les components z es veuen compensades per forces iguals en magnitud i de sentit contrari que exercix la guia de la màquina d'acord amb la 3a llei de Newton. La força exercida realment per l'atleta, però, compta amb les dos components, i és el que volem calcular. Totes estes forces es mostren en el diagrama de la figura 4.35.

Per tal de realitzar els càlculs, en primer lloc calculem el vector unitari de la direcció perpendicular a terra expressat en el sistema de coordenades de la barra. Per a aconseguir això, simplement prenem el vector, que ja el coneixem, i el transformem amb una rotació de la mateixa manera que hem fet abans amb altres vectors en el procés d'establiment de l'homografia. La gravetat $\vec{F}_g = [0 \quad F_{g,y} \quad F_{g,z}]$ és eixe vector perpendicular a terra amb sentit cap a terra i magnitud $|\vec{F}_g| = g \cdot m$ on $g = 9,81 \frac{N}{m}$.

Prenem només la seua component y per tal de calcular la component y de la força exercida per l'esportista sabent que la suma de les dos ha de ser la força resultant:

$$F_{g,y} + F_{esportista,y,i} = F_{res,y,i} \rightarrow F_{esportista,y,i} = F_{res,y,i} - F_{g,y} \quad (4.32)$$

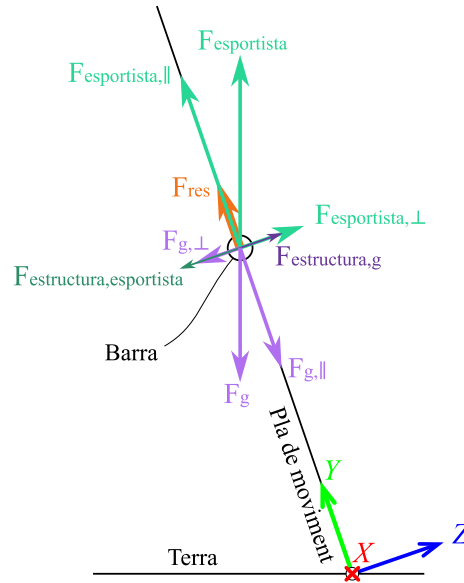


Figura 4.35: Forces que actuen sobre la barra (vista de perfil)

Una vegada conegut el valor de la component y , volem esbrinar quin vector múltiple de l'unitari perpendicular a terra té eixe valor com a component y . Eixe vector és la força exercida per l'esportista. Si denotem per $\hat{u} = [0 \ u_y \ u_z]$ eixe vector unitari i per l el factor multiplicatiu, podem calcular la força exercida per l'esportista de la manera següent:

$$l \cdot u_y = F_{esportista,y,i} \rightarrow l = \frac{F_{esportista,y,i}}{u_y} \quad (4.33)$$

$$\vec{F}_{esportista,i} = l \cdot \hat{u} \quad (4.34)$$

on $\vec{F}_{esportista,i} = [0 \ F_{esportista,y,i} \ F_{esportista,z,i}]$ és el vector que buscàvem, la força total exercida per l'esportista, de la qual prenem la norma com a magnitud d'interés.

4.5.3. Potència

Junt amb la força, una altra magnitud d'interés per a optimitzar els entrenaments és la potència exercida per l'esportista, que es calcula directament segons la seua definició:

$$Pot_i = \vec{F}_{esportista,i} \cdot \vec{v}_i \quad (4.35)$$

on $\vec{v}_i = [0 \ v_{y,i} \ 0]$.

4.5.4. Fases de l'exercici

Tot i no tractar-se d'una magnitud estrictament parlant, la detecció de les diferents fases de l'exercici també és un càlcul que cal realitzar amb posterioritat a la detecció de la posició de la barra en l'espai al llarg del temps. Una repetició de *back squat* (l'exercici per al qual hem desenvolupat l'algoritme) comprén dos fases: la primera, anomenada fase excèntrica, que correspon a la baixada de la barra, i la segona, anomenada fase concèntrica, que correspon a la pujada de la barra i en la qual es fa la major part de l'esforç. A més, tenint en compte que les gravacions poden contindre diverses repeticions, cal delimitar també el temps entre repeticions, que anomenem ací temps de descans.

El càlcul de les diferents fases és relativament senzill. En primer lloc, es determinen com a períodes de descans aquells en què la velocitat de la barra és quasi nul·la (per a poder tindre en compte soroll en les mesures o xicotetes variacions) i la posició de la barra és propera a la de partida. No es pot prendre un criteri estricte solament de posició perquè no necessàriament totes les repeticions comencen i acaben al mateix punt, i tampoc es pot tindre en compte només la velocitat perquè la velocitat també és nul·la en la part més baixa del recorregut, quan passa de la fase excèntrica a la concèntrica. Concretament, les condicions que s'han pres són una velocitat menor a $0,05 \frac{m}{s}$ i una posició a menys de 0,15 m del punt d'inici durant almenys 0,2 s (per a filtrar algun possible compliment espuri de les condicions anteriors).

Una vegada determinats els períodes de descans, la resta correspon a les diferents repeticions. Dins de cada una, es calcula el mínim de la posició (és a dir, la posició més baixa) i es considera que és el límit entre la fase excèntrica i la fase concèntrica de la repetició. A més de ser útil en si mateixa per a l'anàlisi de la realització de l'exercici, aprofitem esta informació per a calcular la velocitat mitjana de la barra durant la fase concèntrica de cada repetició, ja que es tracta d'una magnitud d'interès per a avaluar el rendiment dels esportistes. Açò es fa a terme simplement com a la mitjana aritmètica de les mostres de velocitat instantània que cauen dins de la fase concèntrica de cada repetició.

4.6. Descripció de l'algoritme

Tenint en compte tots els aspectes que s'han explicat fins ara, descrivim a continuació l'algoritme complet en el mateix ordre en què s'executa, i expliquem algunes decisions de disseny.

4.6.1. Lectura de paràmetres i importació de dades

Primer que res, es lligen el fitxer `.ply` del núvol de punts estàtic i el fitxer `.bag` que conté la gravació bidimensional de la imatge de color. Este darrer fitxer és un *bag* de ROS que conté diferents *topics*, dels quals ens interessa `/device_0/sensor_1/Color_0/image/data`, que conté les imatges captades en seqüència. Els missatges del *bag*, que són les imatges, les guardem per a més avant, i fem el mateix amb les marques de temps dels missatges, que usarem després com a els instants en què s'ha capturat cada fotograma de la gravació.

A continuació, carreguem també els paràmetres intrínsecs i extrínsecs de la càmera estèreo que hem obtingut a través del programari del fabricant. Els paràmetres intrínsecs són les distàncies focals de la càmera (f_x i f_y) i les coordenades del punt principal (c_x i c_y), i els paràmetres extrínsecs

són la matriu de rotació ($\underline{\underline{R}}_{RGB \rightarrow prof}$) i el vector de translació ($\vec{T}_{RGB \rightarrow prof}$) que relacionen el sistema de coordenades del núvol de punts amb el sistema de coordenades del sensor RGB.

4.6.2. Transformació de l'escena

Per tal de treballar còmodament amb el núvol de punts, és necessari que el seu sistema de coordenades siga “natural” en relació a l'escena, és a dir, que un dels eixos siga perpendicular al terra. És per això que cal rotar l'escena per a aconseguir situar el terra en el pla $y = 0$. Si es captura el núvol de punts des d'una orientació natural, ja hi està prop, però cal ajustar-ho i, a més, cobrir altres possibles orientacions.

Per a això, prenem la meitat del núvol on es troba el terra d'acord amb la informació d'orientació, i ens aprofitem del fet que el pla més gran que formen els punts del núvol és el del terra per a ajustar-hi un pla model (figura 4.36). La funció `pcfitplane` ajusta plans ideals a conjunts de punts mitjançant el mètode MSAC, una variant de RANSAC, i li especifiquem restriccions com ara l'orientació que esperem que tinga el pla detectat dins d'una tolerància (que en el nostre cas és de 45°).

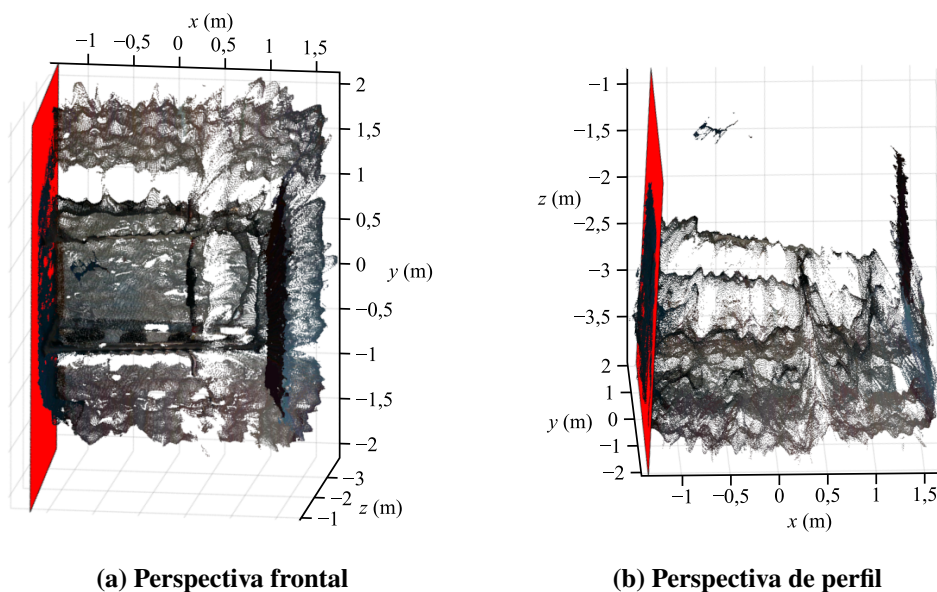


Figura 4.36: Núvol de punts original de l'escena amb un model de pla ajustat al terra

La funció `normalRotation` permet calcular la matriu de rotació que cal aplicar a un vector perquè coincidisca amb un altre, i això ho aprofitem per a esbrinar la rotació que hem d'aplicar al pla (i, per tant, al núvol) perquè siga paral·lel al pla XZ del sistema de coordenades. Després, calculem també la translació que cal aplicar per a fer que el terra se situe específicament en $y = 0$, i l'apliquem també tant al pla ideal com al núvol sencer, de manera que el núvol ja té la posició i l'orientació que desitjàvem, tal com mostra la figura 4.37 (equivalent a la figura 4.23 i repetida per comoditat).

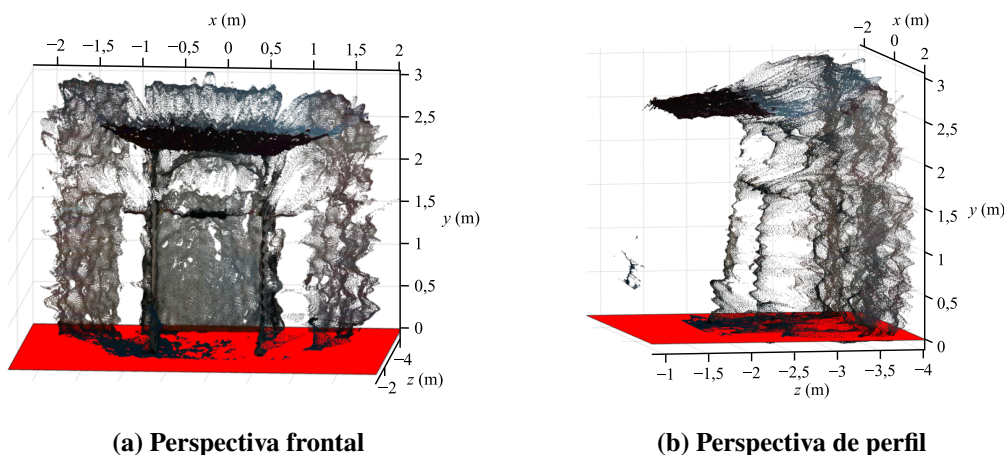


Figura 4.37: Núvol de punts de l'escena després de fer coincidir el terra amb $y = 0$

4.6.3. Aïllament de la màquina de Smith

Per tal de poder reconèixer les diferents parts de la màquina de Smith, és necessari manipular el núvol de punts de l'escena i anar descartant i eliminant punts fins que quede únicament la màquina. Es tracta d'una de les parts més crítiques de l'algoritme. Per a dur-la a terme, ens servim de diverses funcions de MATLAB, però fonamentalment de les d'ajust de models geomètrics i les de segmentació.

La funció `pcsegdist` permet segmentar un núvol de punts per distància, és a dir, classificar-ne els punts en grups (anomenats segments) segons la proximitat entre ells. Funciona per connectivitat, de tal manera que si un punt A es troba a una distància menor que un cert llindar d'un altre punt B, el punt A forma part del mateix segment a què pertany el punt B. El llindar de proximitat s'especifica en cridar la funció. Com que tots els elements de l'escena estan connectats pel terra, abans de segmentar-la descartem els 15 cm inferiors del núvol de punts i eliminem també el sostre i la paret del fons mitjançant la tècnica d'ajustar un pla amb una orientació aproximada a eixes parts del núvol de punts. Per a assegurar que els segments creats són raonables, a més del llindar de proximitat s'ha especificat la quantitat mínima de punts de què ha de constar un segment per a tindre's en consideració. Açò evita que apareguen molts segments de pocs punts.

Així, tenim el contingut de l'escena dividit per segments, amb un per a cada objecte (figura 4.38, repetició per comoditat de la 4.21). Idealment, en l'escena tan sols apareixeria la màquina de Smith, però per si de cas hi ha més objectes, s'ha implementat un filtratge per mida. Així, els que tenen una mida molt diferent a la típica d'una màquina de Smith són descartats. Després d'això és quasi impossible que quede més d'un objecte (figura 4.39), però per si es donara el cas, s'ha implementat com a criteri arbitrari que el més proper a la càmera es considere que és la màquina.

4.6.4. Detecció i modelatge de les parts de la màquina

Una vegada ja hem obtingut un núvol de punts que conté únicament la màquina de Smith, cal reconèixer les seues parts (pilars i barra transversal) i modelar-les geomètricament per tal de poder realitzar posteriorment les operacions necessàries de projecció, homografia, etc. Per tal de facilitar les tasques, treballem amb una estructura en la part central de la qual només està present la barra

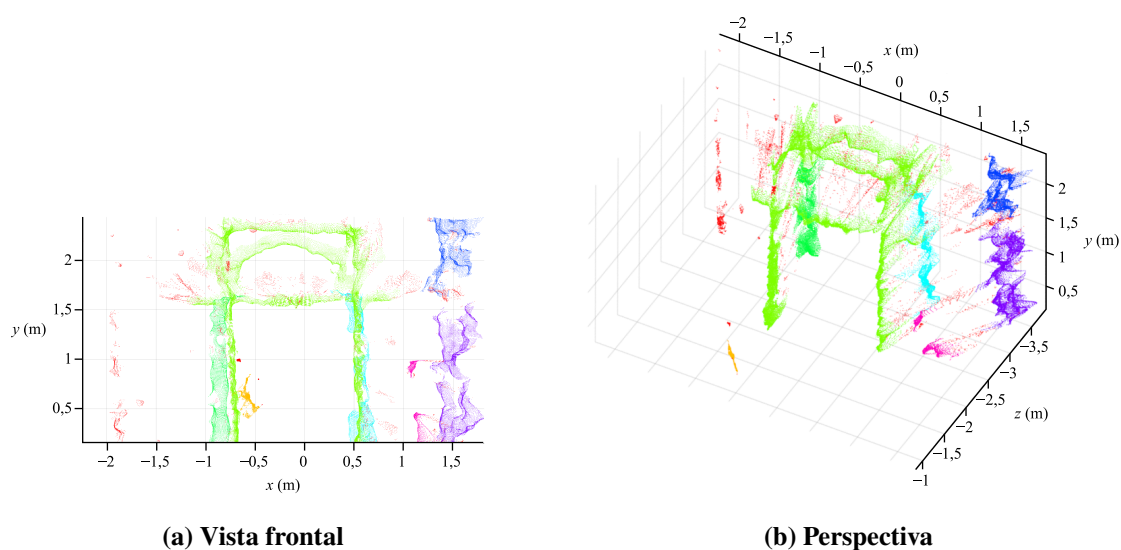


Figura 4.38: La màquina de Smith i el seu entorn en segmentar-los per connectivitat

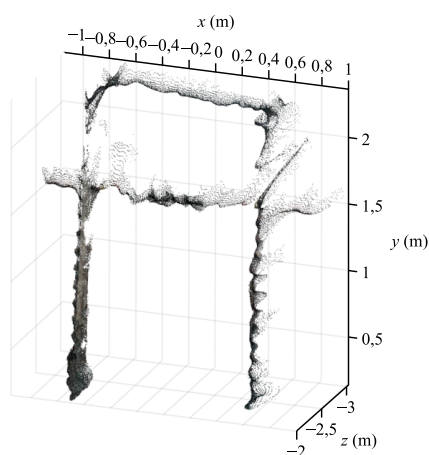


Figura 4.39: Màquina de Smith aïllada

transversal i en els laterals de la qual només trobem els pilars, és a dir, una estructura en forma de H. Per a aconseguir-la, eliminem la part superior, que sol contindre un pilar transversal, i tota la part posterior de la màquina, que no és rellevant en el nostre algoritme.

A continuació, prenem com a mostra de la barra transversal la part central de la màquina en les direccions x i z . En la direcció y prenem tota l'extensió de la màquina perquè desconecim a quina altura està la barra. Com que hem realitzat les operacions descrites en el paràgraf anterior, el més normal és que quede un sol grup de punts, però per si de cas, fem una segmentació i un filtratge per mida tal com hem fet abans per a trobar la màquina en l'escena. Esta mostra de la barra transversal ens permet conèixer-ne informació com ara l'altura a la qual se situa o el gruix.

En concret, estes dos dades permeten prendre una franja transversal (aplicant certes toleràncies) de la màquina per a poder ajustar-hi una recta en l'espai que modele la barra. L'ajust es fa determinant el centroid de la franja mitjançant una mitjana aritmètica de tots els seus punts i la

direcció dominant mitjançant una descomposició en valors singulars. Per tal de trobar els extrems de la barra, es va recorrent la recta que hem determinat abans per passos d'1 cm i comprovant a cada pas si hi ha punts del núvol a menys de 10 cm de distància. Quan no se'n troben, es considera que s'ha traspassat en eixos 10 cm l'extrem de la barra. Així, amb el centre de la barra, la direcció, els extrems i el gruix, el model de la barra queda completament determinat.

En les franges de la màquina en les quals no hem trobat la barra només ha d'haver-hi els dos pilars. De la mateixa manera que en els casos anteriors, efectuem una segmentació i un filtratge per mida per a assegurar-nos que es descarten altres possibles objectes o grups de punts diferents dels pilars i que n'hi ha exactament dos. A més, reconeixem quin és l'esquerre i quin és el dret a partir de la posició de cadascun. Coneixent la posició de cada pilar, es prenen sengles franges verticals de la màquina i s'hi ajusten sengles rectes mitjançant el mateix mètode que s'ha utilitzat per a la barra transversal. Així, ja tenim la barra transversal i els dos pilars modelats per trams de rectes (figura 4.40, repetició per comoditat de la 4.24).

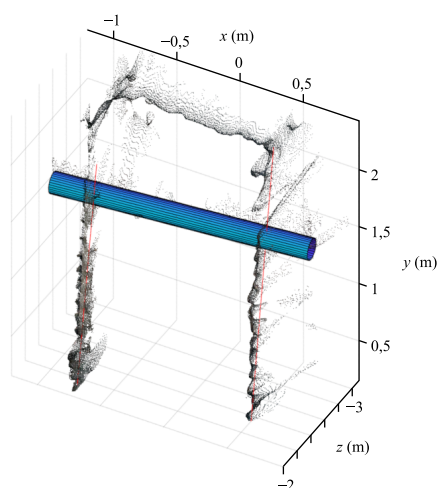


Figura 4.40: Modelatge de la barra i dels pilars per elements geomètrics

4.6.5. Determinació del pla de moviment i del seu sistema de coordenades

El pla en què té lloc el moviment de la barra és una peça clau per a poder dur a terme més endavant la part relacionada amb la projecció dels punts detectats. El primer pas és detectar el pla que formen els dos pilars de la màquina. Com que pot ocórrer que les rectes que modelen els pilars no siguin paral·leles o inclús no pertanguen a un mateix pla, no és viable utilitzar-les per a trobar el pla de moviment. En canvi, s'ha comprovat tant amb la maqueta de simulació com amb la màquina de Smith real que la funció d'ajust d'un pla a un núvol de punts és bona a l'hora de trobar el pla que formen els pilars en la seua cara davantera (figura 4.41, repetició per comoditat de la 4.22), ja que la majoria de punts que apareixen en els núvols de punts dels pilars hi pertanyen. Així, la funció `pcfitplane` és la que s'utilitza amb este objectiu. Com sempre, s'especifiquen paràmetres d'orientació esperada, tolerància i distància al model per a considerar un punt com a pertanyent al pla. Per a este últim paràmetre s'especifica un valor xicotet per a intentar que el pla quede tan cap avant com siga possible.

Sabem que el pla de moviment o pla de la barra és paral·lel al pla dels pilars, ja que en les

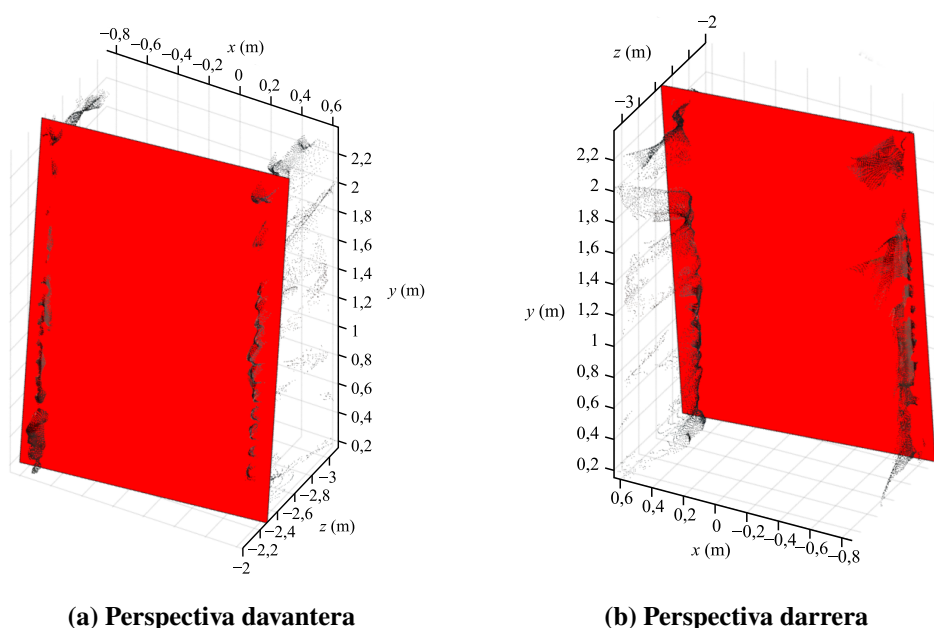


Figura 4.41: Ajust d'un pla als pilars de la màquina de Smith

màquines de Smith les guies de la barra transversal són paral·leles als pilars i la distància entre la guia i la part davantera dels pilars és igual en els dos pilars. Així, determinem el pla de moviment de la barra com a aquell que és paral·lel al dels pilars i conté el centroide de la barra.

Un dels sistemes de coordenades que intervenen en l'homografia que es calcula més avant és el del pla de la barra, de manera que aprofitem que acabem de determinar eixe pla per a especificar el sistema de coordenades associat a ell. Els eixos d'este sistema es configuren a partir de les interseccions del terra, el pla de moviment i un pla auxiliar perpendicular als altres dos i que passa pel centroide del pilar esquerre (figura 4.42, repetició per comoditat de la 4.32), tal com s'explica en la pàgina 48. En concret, els plans i els eixos es determinen a través dels productes vectorials adjacents i de la resolució del sistema d'equacions format per les equacions dels tres plans. A més, s'assegura que el sentit dels eixos siga el correcte.

4.6.6. Detecció de la marca de la barra en les imatges 2D

L'altre ingredient que necessitem per a calcular la posició de la barra en l'espai és la posició de la marca de la barra en la seqüència temporal d'imatges 2D de color. La marca es busca en les imatges per mitjà de tècniques de processament com ara operacions morfològiques, i estes operacions suposen una càrrega computacional notable, de manera que s'estableix una zona de cerca més reduïda que elimina el cost de buscar la marca en llocs de la imatge on és impossible que aparega.

Esta zona es defineix a partir de l'escena 3D. En la direcció horitzontal, desplaçant-nos cap a l'esquerra al llarg de la direcció de la barra, especifiquem la distància comptada des del centroide de la barra a partir de la qual comença la zona de cerca, i especifiquem una segona distància en la qual acaba eixa zona. Estes distàncies han de ser tals que abasten aproximadament des del pilar esquerre fins a una quarta part de la barra transversal en totes les màquines de Smith. En la direcció vertical,

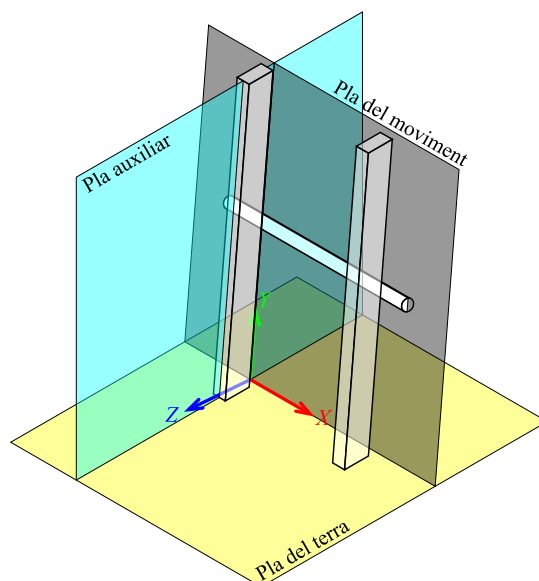


Figura 4.42: Plans de l'escena que configuren el sistema del pla de la barra

busquem la barra des d'una altura de 60 cm fins a una altura de 2 m o l'alçada de la màquina menys 20 cm.

A partir d'estes fites, determinem els quatre vèrtexs d'un quadrilàter situat en l'espai i paral·lel al pilar esquerre. A continuació, hem de dur estos quatre punts a la imatge 2D de color. Per a això, desfem sobre els punts la translació i la rotació que hem aplicat sobre el núvol complet al principi de l'algoritme. A continuació, cal passar d'expressar-los en el sistema de coordenades 3D del núvol (centrat en el sensor esquerre de profunditat) a expressar-los en el sistema de coordenades 3D centrat en el sensor RGB, per a la qual cosa operem de la mateixa manera que en l'equació 4.8. Finalment, per tal d'obtenir les coordenades dels punts en la imatge 2D de color, simplement fem una projecció de la mateixa manera que en les equacions 3.3 i 3.4. Cal tindre en compte, això sí, que Z és negativa i a Y cal canviar-li el signe perquè els nostres sistemes de referència tenen direccions contràries per a la coordenada vertical.

Quan ja coneixem la posició dels quatre cantons del quadrilàter de cerca en la imatge 2D de color (que en general formen un trapezoide), trobem el rectangle alineat amb els eixos que circumscriu a eixe quadrilàter (figura 4.43), i eixa serà la regió on es busca la marca groga de la barra.

A continuació, es prenen tots els fotogrames que s'han extret del *bag* de la gravació al principi de l'algoritme i es passen, un per un, per la rutina de detecció de la marca (figura 4.44, repetició per comoditat de la 4.26). El funcionament d'esta rutina és el que hem explicat en l'apartat 4.4.2. A continuació, sobre el vector que hem obtingut de coordenades de la marca al llarg del temps apliquem un filtre gaussià de suavitzat per tal de reduir el soroll i possibles irregularitats. Finalment, en cas que l'usuari l'haja demanat a través del paràmetre corresponent, es crea un vídeo amb tots els fotogrames de la gravació i un senyal en la posició on s'ha detectat la marca. Per a fer açò es reordenen les dades crues de la gravació en el format d'imatge amb què treballa MATLAB i s'inserixen en el fitxer tenint en compte l'orientació de la càmera en la gravació original.

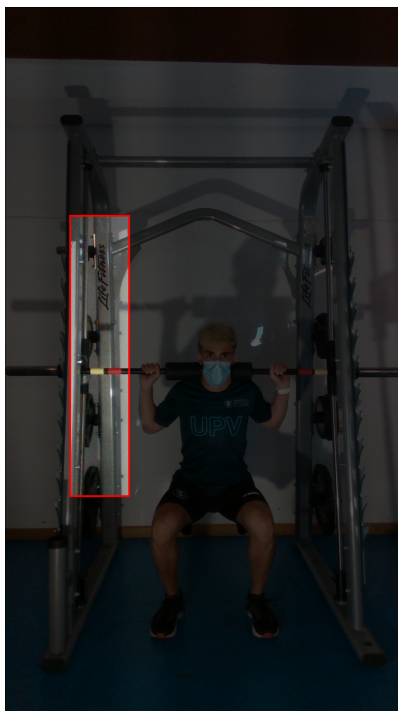


Figura 4.43: Exemple de zona on es cerca la marca de la barra dins d'un fotograma

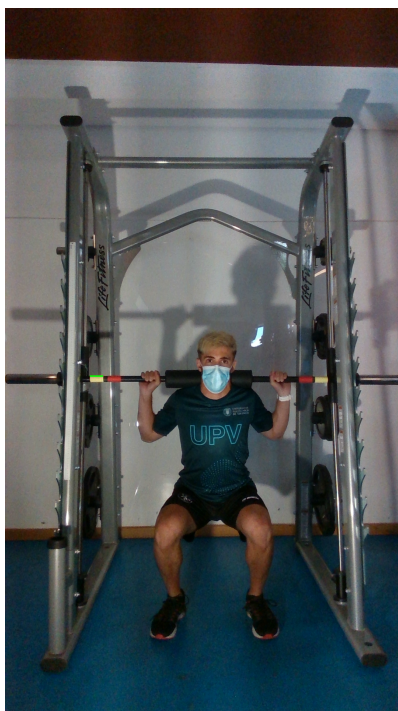


Figura 4.44: Detecció de la marca en un fotograma de la gravació

4.6.7. Determinació de la posició de la marca en l'espai i en el pla de moviment

Ja hem trobat la marca de la barra en els fotogrames de la gravació, i el que ens interessa ara és saber quina és la seua posició en l'espai i també en el sistema de coordenades del pla de moviment. El procediment per a aconseguir això és fer una projecció inversa i establir una homografia amb eixe pla.

Per simplicitat, totes eixes operacions es duen a terme en el sistema de coordenades de la càmera RGB. La detecció del pla de moviment, però, s'ha realitzat sobre el núvol transformat en què el terra coincideix amb el pla $y = 0$. Per tant, prenem el pla de moviment i li apliquem les transformacions (translació i rotació) contràries a aquelles que es duen a terme sobre el núvol complet al principi de l'algoritme (de pas, es fa el mateix amb els tres vectors que representen els eixos del sistema de coordenades del pla de la barra) i, tot seguit, li apliquem la rotació i la translació necessàries per a expressar-lo en el sistema de referència 3D de la càmera RGB. Els punts detectats els manipulem i hi afegim les coordenades que toquen perquè també estiguen expressats en eixe mateix sistema.

A continuació, es realitza la projecció inversa traçant els rajos que passen per l'origen de coordenades (que és el centre òptic de la càmera) i pel punt detectat en cada fotograma. Els rajos intersecten el pla de moviment en uns altres punts que representen la posició espacial de la marca en cada fotograma, expressada en el sistema de referència de la càmera RGB. Per a expressar-los en el sistema de coordenades de l'espai (el centrat en el sensor esquerre de profunditat), els apliquem la rotació i la translació contràries a les que hem aplicat abans. Ja tenim, doncs, els punts de localització de la marca de la barra en cada fotograma expressats en les coordenades del núvol de punts (figura 4.45).

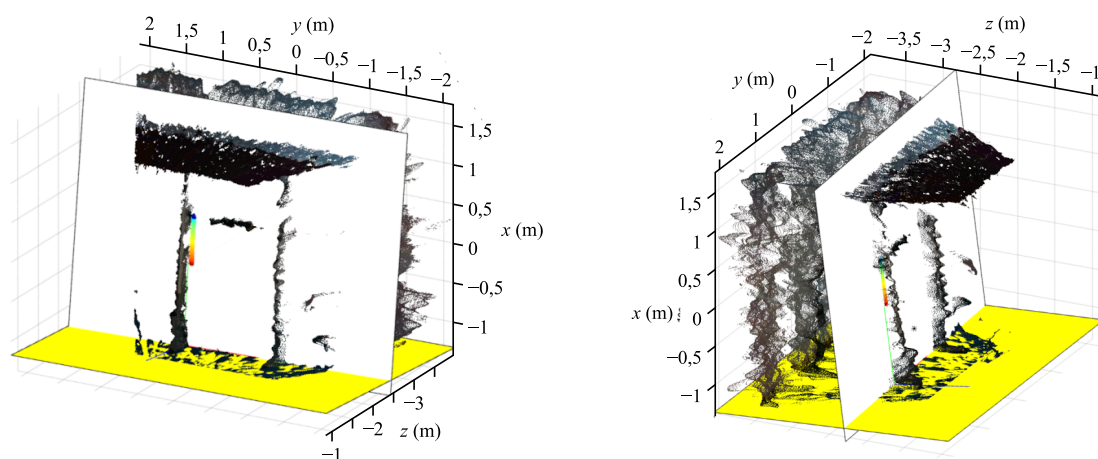


Figura 4.45: Cent primeres deteccions de la marca de la barra (punts de colors) sobre el núvol de l'escena

Per a expressar eixos punts en les coordenades del pla de la barra, cal trobar la relació entre estos dos sistemes. Com que coneixem l'expressió, en el sistema del núvol, dels tres vectors unitaris corresponents als tres eixos del sistema del pla de la barra, trobar la rotació i la translació implica resoldre un sistema d'equacions i fer una resta. Una volta conegudes, s'apliquen als punts i s'obtenen finalment els punts de localització de la marca de la barra en cada fotograma expressats en el sistema de referència del pla de la barra.

Tot este processament geomètric, del qual s'ha oferit ací un resum, s'ha explicat més detalladament i mostrant les operacions matemàtiques pertinents entre les pàgines 48 i 53.

4.6.8. Càlcul de les magnituds d'interés

De la localització de la marca en el pla prenem la coordenada y , ja que el moviment té lloc exclusivament en esta direcció, i canviem la posició de referència (amb valor 0 m) a un punt detectat abans de començar el moviment. A més, com que els instants de temps que proporciona el *bag* que hem llegit al principi del fitxer no estan equiespaciats, interpolem els valors de la posició en una base de temps amb un mostratge uniforme.

A continuació, calculem les diferents magnituds tal com hem explicat abans en la secció 4.5: la velocitat, com a quocient de diferències finites de la posició front al temps; l'acceleració, com a quocient de diferències finites de la velocitat front al temps; la força exercida per l'esportista, mitjançant la 2a llei de Newton i realitzant un càlcul vectorial de forces on incorporem les dades de massa de les peses i l'esportista configurades per l'usuari; i la potència, com a producte escalar dels vectors força i velocitat. La velocitat i l'acceleració les passem per filtres Savitzky-Golay per tal de reduir el soroll. També determinem, segons la posició i la velocitat, les diferents repeticions amb una fase excèntrica i una fase concèntrica cadascuna i el període de descans entre repeticions. Calculem també la velocitat mitjana en la fase concèntrica de cada repetició.

Capítol 5

Validació i resultats

Mentre s'ha desenvolupat el sistema, s'han anat provant les seues funcions sobre un model més o menys ideal, com és la maqueta feta de llistons de fusta, per tal de poder comprovar el funcionament del codi de l'algoritme d'una manera senzilla i accessible. Tanmateix, sempre cal provar que l'algoritme funciona en una situació d'ús tan semblant a la real com siga possible i, a més, validar-ne els resultats front a una medició completament independent a la qual atorguem més confiança. Com s'han dut a terme aquestes dos tasques és el que expliquem en este capítol.

5.1. Preparació de les gravacions

En les gravacions pròpiament dites en un escenari real i amb un esportista calia que el funcionament del sistema estiguera perfectament provat i que s'hagueren tingut en compte abans les necessitats específiques d'una situació real que no apareixen en provar amb una maqueta en condicions favorables. Per tant, i per tal d'optimitzar els resultats de les gravacions en escenari real, s'han fet abans algunes proves per a adquirir familiaritat amb el comportament del sistema en condicions reals i poder ajustar-hi els paràmetres oportuns.

En primer lloc, s'han provat en un entorn controlat els paràmetres tècnics de la càmera dels quals depén que la imatge de color es veja nítida i tinga la lluminositat adequada, és a dir, temps d'exposició, guany, etc. En definitiva, és en esta fase quan s'han estudiat els factors que mencionàvem en l'apartat 4.2.4, amb la conclusió que primer cal ajustar el temps d'exposició als valors 39 o 78 per tal que la barra aparega nítida, i després seleccionar el guany adequat (i, si cal, afegir la il·luminació externa necessària) perquè la lluminositat de la imatge siga la correcta i es puguen identificar bé els diferents colors.

La segona part de les preparacions ha consistit en una visita a la localització de les gravacions (el gimnàs de l'UPV) prèvia a la seua realització. En esta visita s'ha dut a terme una espècie de "simulacre" de les gravacions amb els següents objectius:

- Determinar els valors dels paràmetres tècnics estudiats anteriorment que proporcionen la millor imatge concretament en l'escenari de les gravacions
- Determinar quines gravacions cal dur a terme

- Detectar algun factor o situació no previstos durant el desenvolupament de l'algoritme i que poden causar-ne un mal funcionament

Per a realitzar este simulacre hem dispostat de la càmera estèreo, un trípede, un ordinador portàtil per a recollir les gravacions i un xicotet focus addicional. Amb el temps d'exposició ajustat al valor 78, el guany al màxim i el focus encés s'han pres estes captures estàtiques i gravacions de moviment:

- Captura estàtica d'imatge de color (figura 5.1a)
- Captura estàtica amb projecció infraroja d'imatges monocromes, imatge de profunditat i núvol de punts sense textura (figures 5.1b i 5.1c)
- Captura estàtica sense projecció infraroja d'imatges monocromes, imatge de profunditat i núvol de punts sense textura
- Captura estàtica amb projecció infraroja de núvol de punts amb textura de color i amb il·luminació addicional
- Captura estàtica amb projecció infraroja de núvol de punts amb textura de color i sense il·luminació addicional (figura 5.2)
- Gravació d'imatge de color
- Gravació d'imatge de color i imatges infraroges
- Gravació d'imatge de color i imatge de profunditat
- Gravació d'imatge de color i imatge de profunditat, moviment lent
- Gravació d'imatge de color i imatge de profunditat, 3 repeticions
- Gravació d'imatge de color i imatge de profunditat, només la part esquerra de la màquina i amb il·luminació addicional

S'han examinat les imatges i els núvols adquirits i s'han passat per l'algoritme. Després de l'anàlisi de les dades i de xicotets ajustos en l'algoritme, hem fet estes consideracions per a tindre-les en compte en la gravació real:

- A prop de la màquina de Smith hi ha una porta que du a l'exterior de l'edifici. Tal com esperàvem, amb la porta oberta, la llum solar crea alguns artefactes en el núvol de punts, ja que interfereix en la detecció del patró infraroig; en canvi, el núvol de punts millora amb la porta tancada.
- A conseqüència d'això, la situació de la màquina de Smith en un lloc poc il·luminat del gimnàs requereix l'ús d'un focus més potent per a optimitzar la detecció de la barra, però només en gravar el moviment, que és on s'ha de detectar la marca groga. La captura del núvol de punts és millor sense il·luminació addicional.

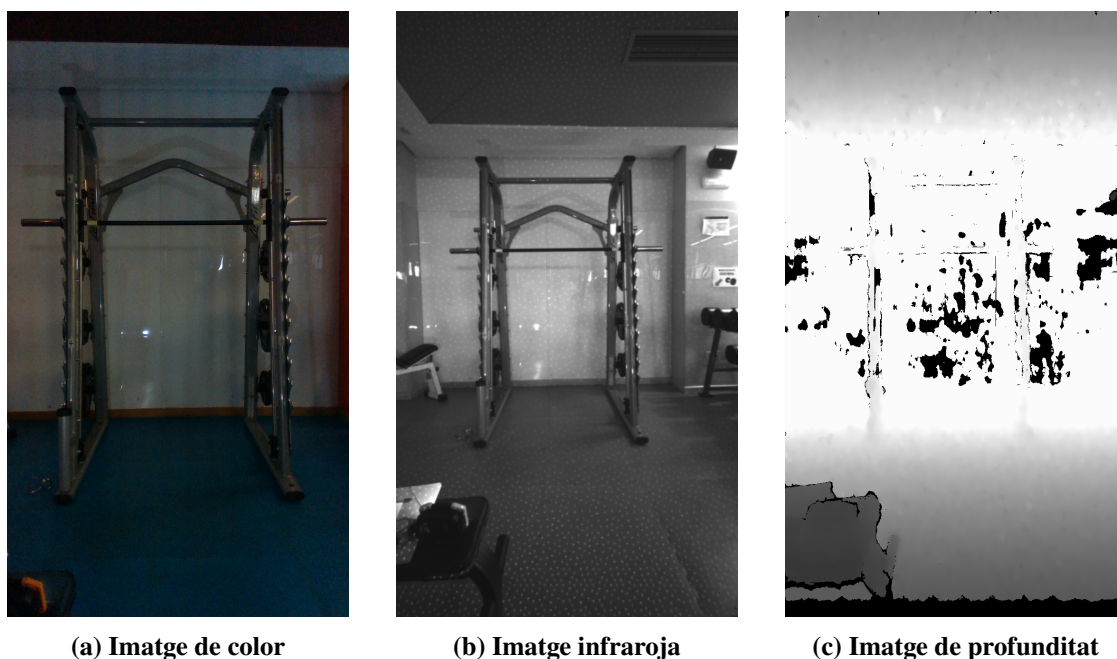


Figura 5.1: Imatges bidimensionals de la màquina de Smith del gimnàs de l'UPV

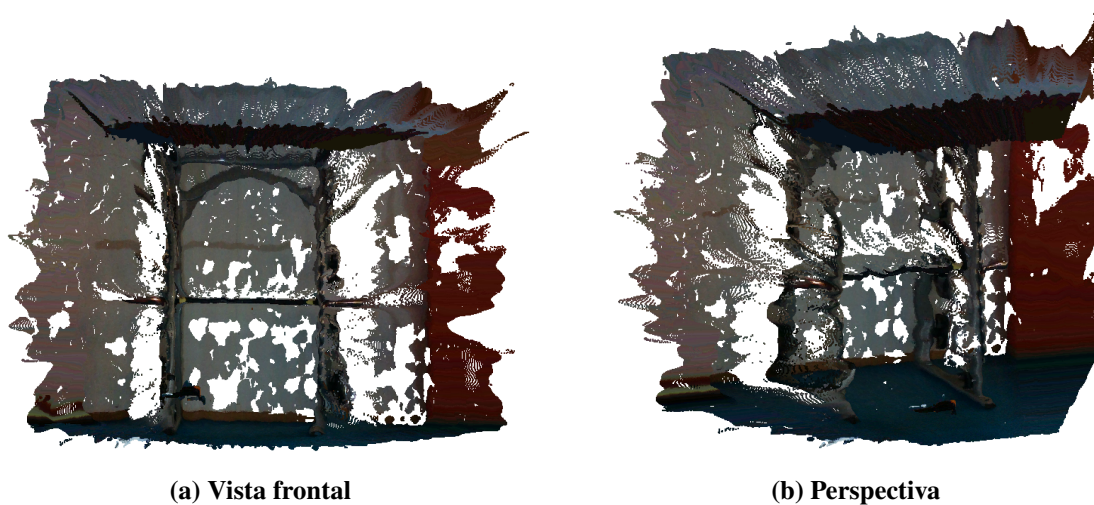


Figura 5.2: Núvol de punts de la màquina de Smith del gimnàs de l'UPV amb textura de color

- A diferència de la maqueta, les màquines reals tenen quatre pilars (els dos de davant i els dos de darrere). Els pilars de darrere, que només tenen una finalitat estructural i no juguen cap paper en el nostre algoritme, no interferixen en absolut en les operacions que realitzem en els de davant.
- Esta màquina en concret disposa d'elements per a recolzar la barra que estan situats per davant dels pilars. Això redueix un poc la qualitat del núvol de punts, però l'algoritme funciona correctament.

- El fet que el fons de la màquina se situe a prop d'una paret reflectant i la llunyania de la càmera a la màquina (per a poder captar-la sencera) també degraden el núvol sense que arribe a afectar a l'algoritme més enllà de xicotets canvis en alguns paràmetres.
- La barra es veu bastant nítida, cosa que obri la porta a treballar amb un temps d'exposició més gran per tal de compensar un poc la falta de llum.

5.2. Gravacions en condicions reals

Havent après algunes lliçons del simulacre, s'han realitzat algunes gravacions i captures estàtiques amb la col·laboració de David Linares, estudiant de l'ETSET, que ha pres el paper d'esportista i ha fet algunes repeticions de *back squat* amb el pes de la barra (figura 5.3). Per comoditat s'han pres per separat les captures del núvol de punts i les gravacions del moviment de l'esportista, però en una aplicació final seria fàcil realitzar les dos accions amb una única intervenció de l'usuari. L'ús d'este focus més potent com a font de llum tan directa ha causat l'aparició de lleus patrons d'interferència relacionats amb la diferència entre la freqüència de quadre i la freqüència de la xarxa elèctrica. Estos patrons no han suposat finalment cap obstacle per a la detecció correcta de la marca de la barra, però de totes maneres s'han captat les imatges amb i sense la il·luminació del focus per a poder triar les millors.



Figura 5.3: Gravació en condicions reals amb un esportista

Finalment, s'ha triat com a òptima la gravació sense il·luminació, ja que ha resultat ser l'òptima pel que fa a l'enregistrament del mètode de validació que s'explica en la secció següent. El criteri que s'ha seguit per a prendre esta decisió ha sigut donar-li una importància major a la bona visibilitat del mètode de validació, ja que la detecció de la marca groga de la barra està dissenyada per a funcionar de manera acceptable sobre imatges fosques amplificant-ne artificialment la brillantor.

En qualsevol cas, la totalitat de l'algoritme s'ha pogut aplicar sobre la imatge i el núvol triats de manera satisfactòria.

5.3. Validació

Per “validació” entenem la comparació dels resultats del nostre algoritme (tant la magnitud que es detecta, o siga, la posició, com les magnituds que en deriven) amb resultats obtinguts mitjançant o a partir d'un mètode de referència.

Pel que fa a eixe mètode de validació, ha consistit en una cinta mètrica fixada a la part davantera d'un dels pilars i assenyalada per un element metàl·lic filamentós fixat fermament a la barra i que es desplaça amb ella (figura 5.4). Així, el valor assenyalat en la cinta mètrica en cada moment és una bona referència del moviment de la barra, entre altres motius perquè minimitza la paral·laxi. La precisió mil·limètrica és més que suficient per a esta aplicació.



Figura 5.4: Mètode de validació format per una cinta mètrica i un fil d'aram

Com que la càmera se situa lluny de la màquina, la resolució de la imatge no és suficient com per a llegir els valors de la cinta mètrica, de manera que s'ha col·locat un telèfon mòbil en un tríode en una posició que permet gravar el moviment del fil d'aram sobre la cinta mètrica sense interferir la captura de l'escena i, a més, permet distingir bé les xifres i les marques de mil·límetres. La lectura s'ha realitzat fotograma per fotograma i, a continuació, s'ha aplicat un filtre de suavitzat a les dades per a mitigar possibles errors de quantització.

La qüestió que sorgix immediatament, que és la de la sincronització del vídeo de gravació i

el de validació, s'ha solucionat amb una claqueta visible des de les dos. Així, les dades de les dos fonts s'han "alineat" per tal de dotar-los de la mateixa base temporal i poder comparar-les adequadament. Esta alineació s'ha dissenyat de manera que es du a terme en dos passos: en el primer, a nivell de fotograma, i en el segon, de manera més fina en fraccions de fotograma, ja que el pas entre fotogrames és suficientment gran com perquè una bona alineació requereisca d'un ajust fi. L'alineació és important perquè l'error no es mostre major del que realment és.

Cal apuntar una qüestió: les gravacions realitzades amb la càmera estèreo i guardades com a *bag* de ROS inclouen missatges amb les marques de temps en què s'ha capturat cada fotograma. En examinar-les s'ha descobert que la configuració de freqüència de quadre de la gravació resulta ser un valor mitjà, o en altres paraules, que els diferents fotogrames de la gravació no estan equiespaciats en el temps, segons les marques (probablement pels requisits computacionals). En la figura 5.5 es pot observar que l'instant de captura de cada fotograma no es desvia pràcticament mai més d'un 10 % de l'interval teòric entre fotogrames i, sobretot, que la mitjana és nul·la, de manera que el retard d'alguns fotogrames es compensa amb l'avançament d'altres.

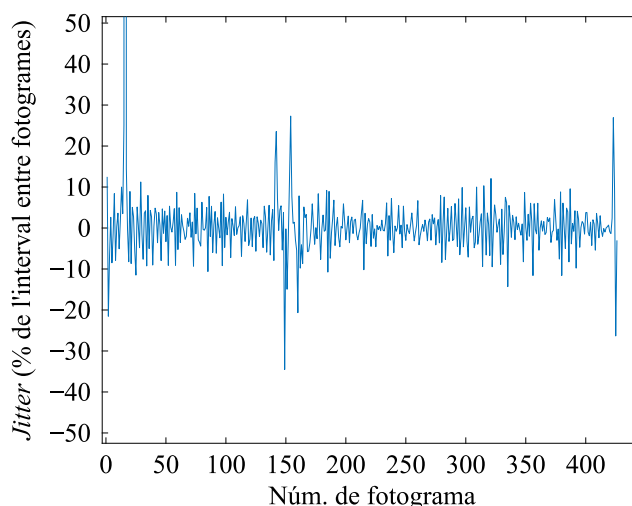


Figura 5.5: *Jitter* relatiu dels fotogrames de la gravació de la càmera estèreo

Encara que s'ha valorat la possibilitat de que realment eixes irregularitats en el temps només estigueren referides al processament i en realitat les captures dels fotogrames sí que hagueren sigut regulars, finalment no s'ha apreciat cap motiu per a desconfiar de les marques de temps que proporciona la càmera, de manera que s'ha considerat que corresponen a l'instant de captura de les imatges.

Això, juntament amb l'ajust fi que comentàvem abans, ha implicat que els instants per als quals hi ha dades de gravació no són en general els mateixos per als quals hi ha dades de validació. Com que la validació sí que té una base de temps regular, s'ha pres com a referència i s'ha obtingut per interpolació el valor de la gravació en els instants per als quals hi ha validació, cosa que ha facilitat el treball posterior amb les dades i especialment el càlcul de l'error.

5.4. Resultats

Les magnituds obtingudes a partir de les dades de validació ens permeten avaluar l'algoritme per comparació amb les magnituds que s'obtenen a partir d'ell. Així, tot seguit es comparen les de la gravació i les de la validació referents a la posició, la velocitat i l'acceleració de la barra transversal, així com la força i la potència exercides per l'esportista. També s'han determinat les diferents fases de l'exercici i s'ha calculat l'error de les magnituds d'interès per als esportistes (velocitat, força i potència) i de la posició de la barra, que és a partir de la qual es calculen totes les altres.

Pel que fa a la posició, en la figura 5.6 s'aprecia com els valors obtinguts a través de l'algoritme que s'ha dissenyat s'ajusten bastant bé a la realitat (representada per la corba de validació). La diferència és de 2,18 cm en el pitjor dels casos (figura 5.7a) i l'error se situa entre l'1 % i el 5 % (figura 5.7b), que es considera acceptable per a la nostra aplicació. La detecció de les diferents fases de l'exercici és correcta.

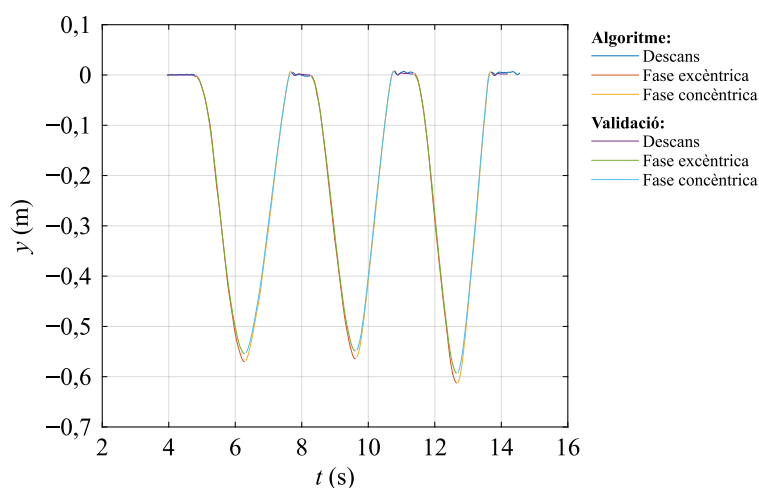


Figura 5.6: Posició de la barra transversal al llarg del temps

La velocitat (figura 5.8) i l'acceleració (figura 5.10) són magnituds resultat de derivar respecte al temps i, com hem explicat, han passat per un filtre per a reduir el soroll que presenten en cru. És per això que trobem diferències observables visualment entre les dades derivades de l'algoritme i les derivades de la validació. Tanmateix, entren dins del que és habitual i acceptable en el camp de les ciències de l'esport. Concretament, en el cas de la velocitat, que és una de les magnituds d'interès per als esportistes, l'estimació de la velocitat mitjana té un error inferior al 3 % en el pitjor dels casos, tal com mostren la figura 5.9b. A més, en [42] es menciona que esta classe de sistemes han de poder detectar com a mínims canvis d'entre $0,05 \frac{m}{s}$ i $0,1 \frac{m}{s}$ en exercicis semblants al d'este treball. En la figura 5.9b comprovem com l'error de la velocitat instantània en termes absoluts queda en quasi la totalitat del temps per baix de $0,05 \frac{m}{s}$, i el de la velocitat mitjana en fase concèntrica no arriba a $0,015 \frac{m}{s}$, és a dir, que són resultats bons.

La força exercida per l'esportista (figura 5.11) es calcula a partir de l'acceleració. Sempre presenta valors positius perquè l'esportista sempre fa força cap a dalt, tot i que en l'oscil·lació que fa, en alguns moments venç a la gravetat (i la barra puja) i en altres moments la gravetat és major (i

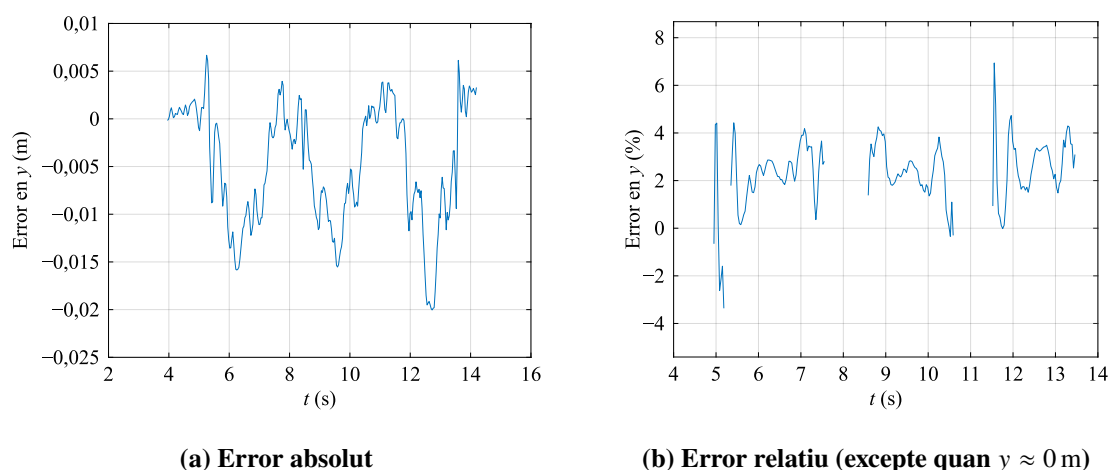


Figura 5.7: Error en la detecció de la barra

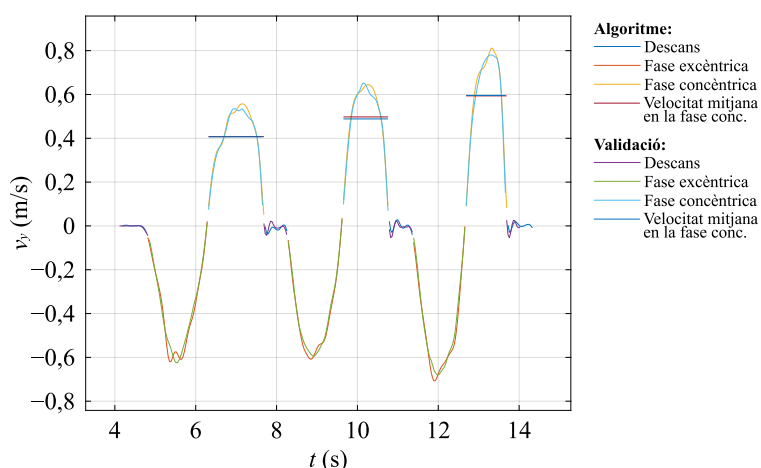


Figura 5.8: Velocitat de la barra transversal al llarg del temps

la barra baixa). L'error es manté en nivells bastant bons (figura 5.12), resultat del fet que els valors tinguin eixe *offset*.

Finalment, la potència desenvolupada per l'esportista (figura 5.13) és positiva quan la barra puja i negativa quan la barra baixa. Tal com s'aprecia en la figura 5.14, l'error es manté la major part del temps sobre el 10%. Este valor és major que en altres magnituds a causa de l'acumulació de l'error en operacions successives.

L'error que presenten la velocitat, la força i la potència es pot retrotraure al que ja existix en l'estimació de la posició de la barra. Els dos factors principals que hi influïxen són la detecció de la marca groga de la barra en la imatge bidimensional de color i la projecció d'este punt sobre el pla dels pilars. Els possibles errors en la detecció de la marca ja es mitiguen en l'algorisme per mitjà d'un filtre de suavitzat i, a més, que l'error augmente quan la barra baixa apunta que l'explicació dels errors la podem trobar més bé en la projecció dels punts.

En desenvolupar la part de l'algorisme dedicada a la projecció dels punts on es troba la marca

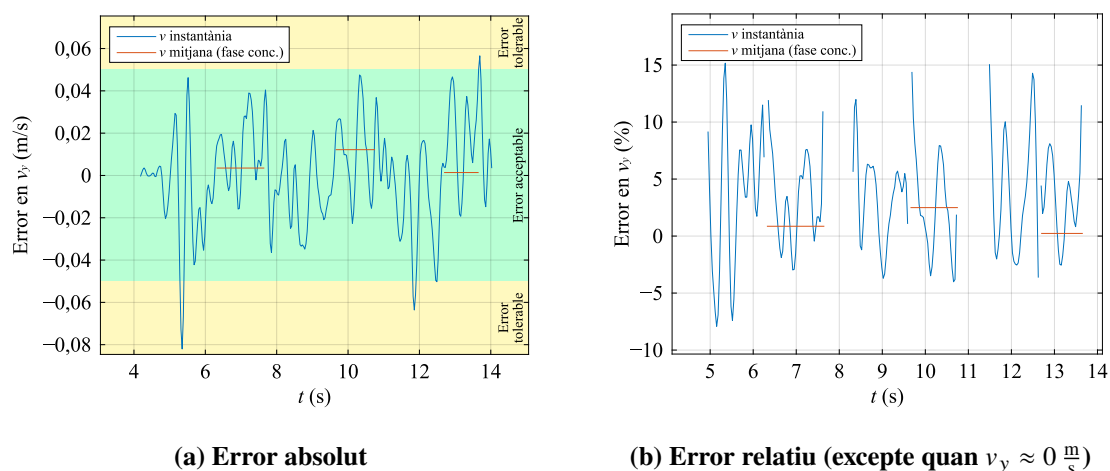


Figura 5.9: Error en l'estimació de la velocitat de la barra

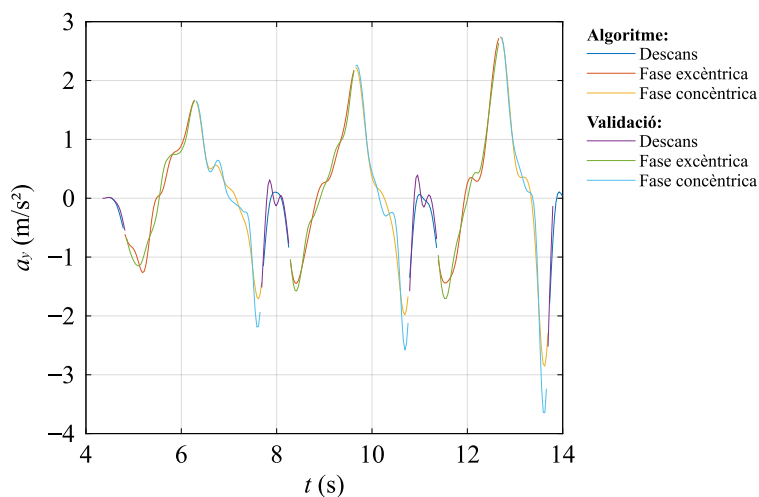


Figura 5.10: Acceleració de la barra transversal al llarg del temps

(per a localitzar eixos punts en l'espai), es van realitzar unes mesures que servixen per a avaluar exclusivament la precisió d'este mètode de projecció. Sobre els pilars de la maqueta utilitzada per a desenvolupar l'algoritme es van col·locar tres marques A, B i C formant un triangle. Es van seleccionar manualment en una imatge bidimensional de color els píxels on es trobaven les marques, i es van projectar estos punts sobre el pla format per la cara davantera dels pilars de la maqueta, que és on es van fixar les marques, de manera analògica a com s'ha dut a terme en les gravacions reals amb el pla de moviment de la barra. Es van mesurar les distàncies entre estos punts projectats i es van comparar amb les reals. Els resultats obtinguts van ser els de la taula 5.1.

En comparar les dades de la taula 5.1 amb les de la taula 4.5, trobem percentatges d'error similars (entre un 2% i un 4%) per a la distància de 2,5 m, que és més o menys a la qual s'han realitzat les gravacions. Això vol dir que les imprecisions del mètode de projecció de punts es deuen a les imprecisions de la càmera en capturar el núvol (concretament, un lleuger factor d'escala).

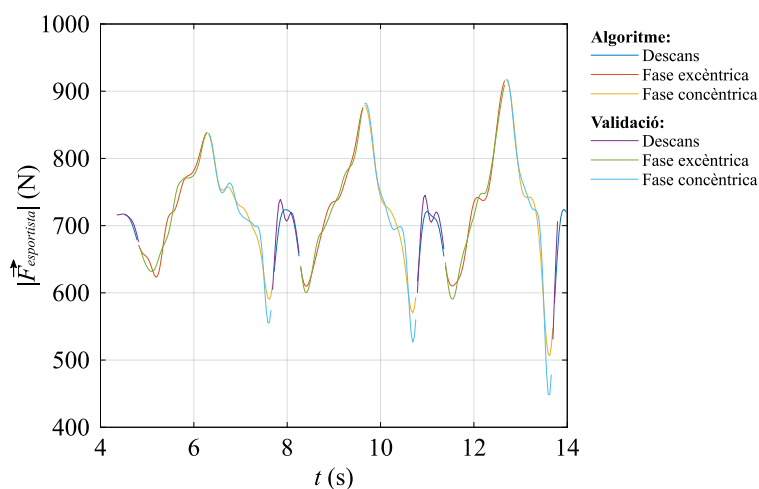


Figura 5.11: Força exercida per l'esportista al llarg del temps

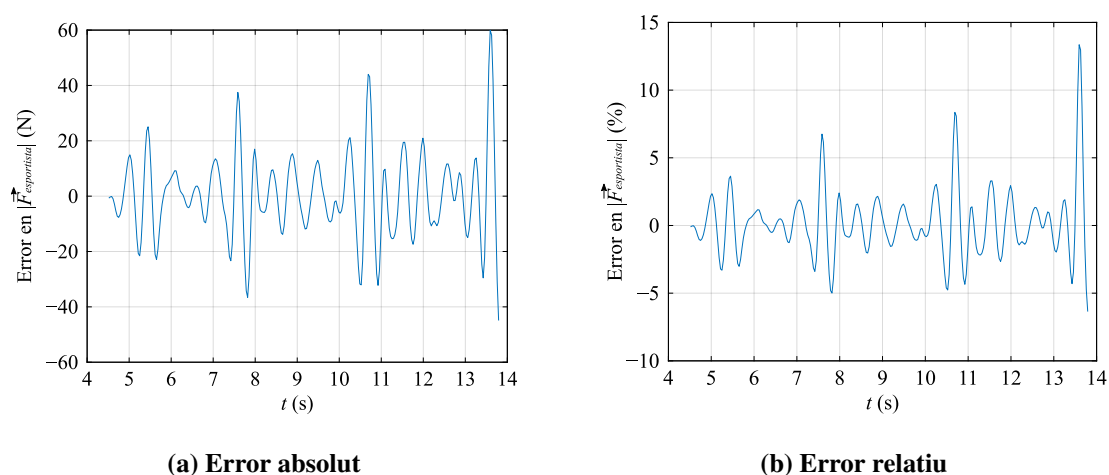


Figura 5.12: Error en l'estimació de la força exercida

Situacions i distàncies	Distància real	Maqueta frontal		Maqueta en perspectiva	
		Dist. mesurada	Error	Dist. mesurada	Error
Segment AB	55,4 cm	57,57 cm	+3,91 %	57,39 cm	+3,60 %
Segment BC	105,3 cm	108,30 cm	+2,85 %	108,66 cm	+3,19 %
Segment AC	119,2 cm	122,45 cm	+2,73 %	123,05 cm	+3,23 %
Mitjana			+3,16 %		+3,34 %

Taula 5.1: Error en la projecció de punts a un pla de l'escena

Si ens fixem en l'estimació de la posició de la barra d'este algoritme, apreciem una pujada lleugera de l'error fins a entre un 2 % i un 5 %. Així, malgrat que la major part de l'error es deu a la càmera en si, també intervenen altres factors, com ara que les condicions i l'entorn en què es troba la màquina de Smith han resultat en un núvol de punts en què la determinació del pla que formen els pilars davanters de la màquina és més delicada i pot ser menys exacta.

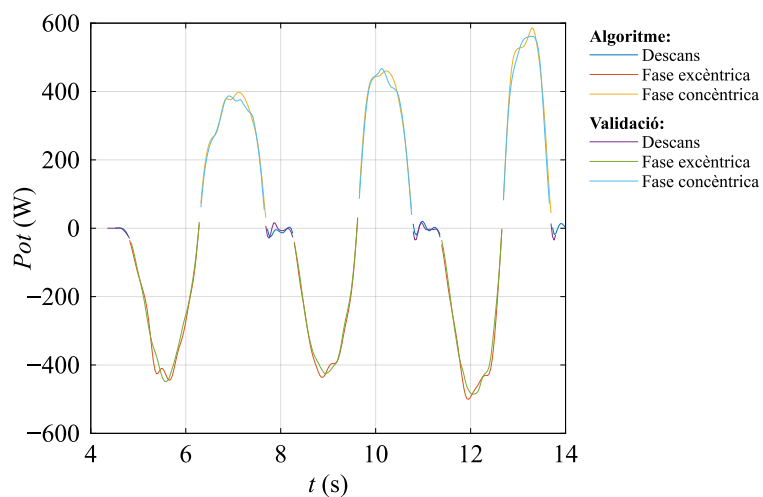


Figura 5.13: Potència desenvolupada per l'esportista al llarg del temps

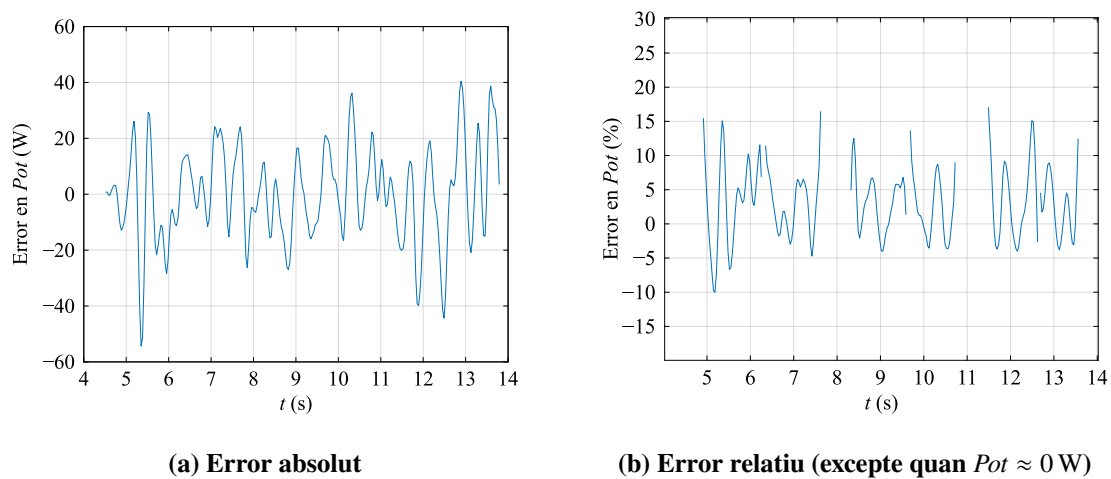


Figura 5.14: Error en l'estimació de la potència

Capítol 6

Conclusions i proposta de treball futur

Després de mostrar com s'ha plantejat i com s'ha desenvolupat el sistema, podem afirmar que els resultats obtinguts són satisfactoris. Tots els objectius plantejats al principi del projecte s'han complert i tant l'algoritme en el seu conjunt com les diferents parts de què està compost servixen correctament els seus fins.

A més, el procés de desenvolupament de l'algoritme i de redacció d'esta memòria ha constituït una experiència valuosa pel que fa a la manera de gestionar projectes de complexitat mitjana i que s'estenen en el temps. De segur que este aprenentatge serà útil en el futur.

Quant a l'algoritme, és adequat per a formar la base o la inspiració per a aplicacions de telèfon mòbil viables que milloraran les rutines d'entrenament de molts esportistes. Per a arribar a eixe punt, això sí, cal treballar en alguns aspectes que malauradament han quedat fora de l'abast d'este treball:

- Optimitzar la cerca de la marca de la barra, de tal manera que en compte de definir una gran zona fixa on es busca la marqueta, es definisca una zona més xicoteta en una posició o una altra de la imatge depenent d'on s'haja trobat la marca en imatges anteriors. Això reduiria la càrrega computacional i faria l'algoritme més ràpid, ja que és actualment una de les seues parts més lentes.
- Millorar la robustesa de l'algoritme, i en particular de la part de segmentació, que de vegades patix un poc en funció de les condicions de l'entorn en què s'haja capturat el núvol de punts
- Implementar l'algoritme amb la biblioteca PCL (Point Cloud Library) sobre C i fer que el processament s'execute en temps real a mesura que es van gravant les imatges
- Implementar l'algoritme de tal manera que el processament el duga a terme la GPU en comptes de la CPU, cosa que gràcies al maquinari especialitzat millorarà molt la velocitat d'execució
- Inserir l'algoritme en una aplicació per a telèfons mòbils que actue com a interfície fàcil d'usuaris i, a més, puga proporcionar funcionalitats addicionals de gestió d'entrenaments

Bibliografia

- [1] *Titanium Strength Smith Machine 680C Linear Bearing*. Mundo Fitness. URL: <https://www.mundofitness.es/smith-machine-titanium-strength-680c-linear-bearing.html> (cons. 25-02-2022).
- [2] *Computer vision*. Wikipedia. URL: https://en.wikipedia.org/wiki/Computer_vision (cons. 17-10-2021).
- [3] *Computer vision*. DeepAI. URL: <https://deepai.org/machine-learning-glossary-and-terms/computer-vision> (cons. 18-10-2021).
- [4] *What is computer vision?* IBM. URL: <https://www.ibm.com/topics/computer-vision> (cons. 18-10-2021).
- [5] *What is stereo vision*. IGI Global. URL: <https://www.igi-global.com/dictionary/stereo-vision/28281> (cons. 20-10-2021).
- [6] Gary Bradsky i Adrian Kaehler. *Learning OpenCV*. Ed. de Mike Loukides. 1a ed. Sebastopol, Califòrnia, EUA: O'Reilly, set. de 2008, pàg. 371 - 373.
- [7] Gary Bradsky i Adrian Kaehler. *Learning OpenCV*. Ed. de Mike Loukides. 1a ed. Sebastopol, Califòrnia, EUA: O'Reilly, set. de 2008, pàg. 415 - 421.
- [8] José Manuel Mossi García. *Procesamiento de señal en comunicaciones y audiovisuales. Cámara dual*. Apunts de classe distribuïts electrònicament a l'alumnat. Universitat Politècnica de València, des. de 2019.
- [9] *Basic 3D machine vision techniques and principles*. Zivid. URL: <https://www.zivid.com/3d-vision-technology-principles> (cons. 24-10-2021).
- [10] *What is a stereo vision camera?* e-con Systems. 10 d'oct. de 2018. URL: <https://www.e-consystems.com/blog/camera/what-is-a-stereo-vision-camera/> (cons. 24-10-2021).
- [11] David Cardinal. *Multi-camera smartphones: benefits and challenges*. DXOMARK. 21 de febr. de 2019. URL: <https://www.dxomark.com/multi-camera-smartphones-benefits-and-challenges/> (cons. 25-10-2021).
- [12] Idris Abdul Ghaffar i Mohd Norzali Haji Mohd. "Selection of Android smartphones with built-in dual lens camera for stereo vision Android app development". A: *2018 7th International Conference on Computer and Communication Engineering (ICCCCE)*. 2018, pàg. 74 - 78. DOI: 10.1109/ICCCCE.2018.8539271.
- [13] *Point Cloud Library*. Read the Docs. Cap. "The PCD (Point Cloud Data) file format". URL: https://pcl.readthedocs.io/en/latest/pcd_file_format.html (cons. 30-10-2021).

- [14] *PLY (file format)*. Wikipedia. URL: [https://en.wikipedia.org/wiki/PLY_\(file_format\)](https://en.wikipedia.org/wiki/PLY_(file_format)) (cons. 29-10-2021).
- [15] *MathWorks Help Center – The PLY format*. MathWorks. URL: <https://es.mathworks.com/help/vision/ug/the-ply-format.html> (cons. 29-10-2021).
- [16] *Research areas*. Technical University of Munich – Computer Vision Group. URL: <https://vision.in.tum.de/research> (cons. 30-10-2021).
- [17] Chul-hong Kim, Tae-jae Lee i Dong-il “Dan” Cho. “An application of stereo camera with two different FoVs for SLAM and obstacle detection”. A: *IFAC-PapersOnLine* 51.22 (2018). 12th IFAC Symposium on Robot Control SYROCO 2018, pàg. 148 - 153. ISSN: 2405-8963. DOI: <https://doi.org/10.1016/j.ifacol.2018.11.533>.
- [18] Sevil A. Ahmed et al. “Model-free detection and following of moving objects by an omnidirectional mobile robot using 2D range data”. A: *IFAC-PapersOnLine* 51.22 (2018). 12th IFAC Symposium on Robot Control SYROCO 2018, pàg. 226 - 231. ISSN: 2405-8963. DOI: <https://doi.org/10.1016/j.ifacol.2018.11.546>.
- [19] Nan Ning i Tele Tan. “A framework for tracking moving target in a heterogeneous camera suite”. A: *2006 9th International Conference on Control, Automation, Robotics and Vision*. 2006, pàg. 1 - 5. DOI: 10.1109/ICARCV.2006.345213.
- [20] J.M. Coughlan i A.L. Yuille. “Manhattan World: compass direction from a single image by Bayesian inference”. A: *Proceedings of the Seventh IEEE International Conference on Computer Vision*. Vol. 2. 1999, pàg. 941 - 947. DOI: 10.1109/ICCV.1999.790349.
- [21] Xiaozhi Chen et al. “3D Object Proposals for Accurate Object Class Detection”. A: *Advances in neural information processing systems*. Ed. de C. Cortes et al. Vol. 28. Curran Associates, Inc., 2015.
- [22] Andrea Tagliasacchi, Hao Zhang i Daniel Cohen-Or. “Curve skeleton extraction from incomplete point cloud”. A: *ACM Transactions on Graphics* 28.3 (ag. de 2009). ISSN: 0730-0301. DOI: 10.1145/1531326.1531377.
- [23] Fadilla ‘Atyka Nor Rashid, Nor Surayahani Suriani i Ain Nazari. “Kinect-based physiotherapy and assessment: a comprehensive review”. A: *Indonesian Journal of Electrical Engineering and Computer Science* 11 (2018), pàg. 1176 - 1187.
- [24] Fabrice Caillette i Toby Howard. “Real-time markerless human body tracking with multi-view 3-D voxel reconstruction”. A: (gen. de 2004). DOI: 10.5244/C.18.62.
- [25] Pathum Rathnayaka, Seung-Hae Baek i Soon-Yong Park. “An efficient calibration method for a stereo camera system with heterogeneous lenses using an embedded checkerboard pattern”. A: *Journal of Sensors* 2017 (set. de 2017). ISSN: 1687-725X. DOI: 10.1155/2017/6742615.
- [26] Katja Wolff et al. “Point cloud noise and outlier removal for image-based 3D reconstruction”. A: *2016 Fourth International Conference on 3D Vision (3DV)*. 2016, pàg. 118 - 127. DOI: 10.1109/3DV.2016.20.
- [27] Hao Song i Hsi-Yung Feng. “A progressive point cloud simplification algorithm with preserved sharp edge data”. A: *The International Journal of Advanced Manufacturing Technology* 45.5 (set. de 2009), pàg. 583 - 592. ISSN: 1433-3015. DOI: 10.1007/s00170-009-1980-4.

- [28] Jian Zhao. “Camera planning and fusion in a heterogeneous camera network”. Tesi doctoral. University of Kentucky, 2011. URL: https://uknowledge.uky.edu/ece_etds/2/ (cons. 31-10-2021).
- [29] Chung-Hao Chen et al. “Heterogeneous fusion of omnidirectional and PTZ cameras for multiple object tracking”. A: *IEEE Transactions on Circuits and Systems for Video Technology* 18.8 (2008), pàg. 1052 - 1063. DOI: 10.1109/TCSVT.2008.928223.
- [30] Alexander Berestov. “Detection and removal of image occlusion errors”. US6865289B1 (USA). Canon Inc. 8 de març de 2005. URL: <https://patentimages.storage.googleapis.com/f3/80/36/91710ded555c25/US6865289.pdf> (cons. 31-10-2021).
- [31] Dahua Lin i Xiaoou Tang. “Quality-driven face occlusion detection and recovery”. A: *2007 IEEE Conference on Computer Vision and Pattern Recognition*. 2007, pàg. 1 - 7. DOI: 10.1109/CVPR.2007.383052.
- [32] Charles Lawrence Zitnick i Takeo Kanade. “A cooperative algorithm for stereo matching and occlusion detection”. A: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22.7 (2000), pàg. 675 - 684. DOI: 10.1109/34.865184.
- [33] Charles R. Qi et al. “PointNet: deep learning on point sets for 3D classification and segmentation”. A: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Jul. de 2017.
- [34] Li Yi et al. “GSPN: Generative Shape Proposal Network for 3D instance segmentation in point cloud”. A: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Juny de 2019.
- [35] Shaoshuai Shi, Xiaogang Wang i Hongsheng Li. “PointRCNN: 3D object proposal generation and detection from point cloud”. A: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Juny de 2019.
- [36] Jianbing Huang i Chia-Hsiang Menq. “Automatic data segmentation for geometric feature extraction from unorganized 3-D coordinate points”. A: *IEEE Transactions on Robotics and Automation* 17.3 (2001), pàg. 268 - 279. DOI: 10.1109/70.938384.
- [37] Chao Wang, Yong K. Cho i Changwan Kim. “Automatic BIM component extraction from point clouds of existing buildings for sustainability applications”. A: *Automation in Construction* 56 (2015), pàg. 1 - 13. ISSN: 0926-5805. DOI: <https://doi.org/10.1016/j.autcon.2015.04.001>.
- [38] Yongtao Yu et al. “3-D feature matching for point cloud object extraction”. A: *IEEE Geoscience and Remote Sensing Letters* 17.2 (2020), pàg. 322 - 326. DOI: 10.1109/LGRS.2019.2918073.
- [39] Antonio Albiol et al. “Robust real-time traffic monitoring using corners motion statistics”. A: *7th International Conference on Imaging for Crime Detection and Prevention (ICDP 2016)*. 2016, pàg. 1 - 6. DOI: 10.1049/ic.2016.0094.
- [40] Nitin L. Gavankar i Sanjay Kumar Ghosh. “Automatic building footprint extraction from high-resolution satellite image using mathematical morphology”. A: *European Journal of Remote Sensing* 51.1 (2018), pàg. 182 - 193. DOI: 10.1080/22797254.2017.1416676.
- [41] Junsheng Zhang et al. “Defect detection of aluminum alloy wheels in radiography images using adaptive threshold and morphological reconstruction”. A: *Applied Sciences* 8.12 (2018). ISSN: 2076-3417. DOI: 10.3390/app8122365.

- [42] Alejandro Martínez Cava et al. “Reliability of technologies to measure the barbell velocity: implications for monitoring resistance training”. A: *PLOS ONE* 15.6 (juny de 2020), pàg. 1 - 17. DOI: 10.1371/journal.pone.0232465.
- [43] Alejandro Pérez Castilla et al. “Precision of 7 commercially available devices for predicting bench-press 1-repetition maximum from the individual load–velocity relationship”. A: *International Journal of Sports Physiology and Performance* 14.10 (2019), pàg. 1442 - 1446. DOI: 10.1123/ijsp.2018-0801.
- [44] Jesús G. Pallarés et al. “Imposing a pause between the eccentric and concentric phases increases the reliability of isoinertial strength assessments”. A: *Journal of Sports Sciences* 32.12 (2014), pàg. 1165 - 1175. DOI: 10.1080/02640414.2014.889844.
- [45] Steve W. Thompson et al. “The reliability and validity of current technologies for measuring barbell velocity in the free-weight back squat and power clean”. A: *Sports* 8.7 (2020). ISSN: 2075-4663. DOI: 10.3390/sports8070094.
- [46] *Manual de usuario – Velowin*. Vers. 1.3. Deportec. URL: https://www.deportec.es/wp-content/uploads/2017/03/MANUAL_DE_USUARIO_v_1_3.pdf (cons. 31-10-2021).
- [47] Basilio Pueo et al. “Video-based system for automatic measurement of barbell velocity in back squat”. A: *Sensors* 21.3 (2021). ISSN: 1424-8220. DOI: 10.3390/s21030925.
- [48] *Intel® RealSense™ product family D400 series. Datasheet*. 337029-010. Vers. 010. Intel. Febr. de 2021. URL: <https://www.intelrealsense.com/download/15409> (cons. 06-05-2021).
- [49] Anders Grunnet-Jepsen et al. “Intel® RealSense™ self-calibration for D400 series depth cameras”. Vers. 2.1. A: *Intel RealSense documentation* (2021). URL: <https://dev.intelrealsense.com/docs/self-calibration-for-depth-cameras> (cons. 08-11-2021).
- [50] *Intel® RealSense™ dynamic calibration print target with fixed width (10 mm) bars*. Vers. 2.1. Intel. Juny de 2017. URL: <https://www.intel.com/content/dam/support/us/en/documents/emerging-technologies/intel-realsense-technology/print-target-fixed-width.pdf> (cons. 08-11-2021).
- [51] Zhengyou Zhang. “A flexible new technique for camera calibration”. A: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22.11 (2000), pàg. 1330 - 1334. DOI: 10.1109/34.888718.
- [52] *MathWorks Help Center – pcsegdist: Segment point cloud into clusters based on Euclidean distance*. MathWorks. URL: <https://es.mathworks.com/help/vision/ref/pcsegdist.html> (cons. 14-11-2021).
- [53] *MathWorks Help Center – Segmentation: Segment point cloud data using deep learning and geometric algorithms*. MathWorks. URL: <https://es.mathworks.com/help/lidar/deeplearning.html> (cons. 14-11-2021).
- [54] *MathWorks Help Center – pcfitylinder: Fit cylinder to 3-D point cloud*. MathWorks. URL: <https://es.mathworks.com/help/vision/ref/pcfitylinder.html> (cons. 14-11-2021).
- [55] *MathWorks Help Center – pcfityplane: Fit plane to 3-D point cloud*. MathWorks. URL: <https://es.mathworks.com/help/vision/ref/pcfityplane.html> (cons. 14-11-2021).

- [56] *MathWorks Help Center – pcfitsphere: Fit sphere to 3-D point cloud.* MathWorks. URL: <https://es.mathworks.com/help/vision/ref/pcfitsphere.html> (cons. 14-11-2021).
- [57] *MathWorks Help Center – ransac: Fit model to noisy data.* MathWorks. URL: <https://es.mathworks.com/help/vision/ref/ransac.html> (cons. 14-11-2021).
- [58] *MathWorks Help Center – segmentGroundSMRF: Segment ground from lidar data using a Simple morphological filter (SMRF) algorithm.* MathWorks. URL: <https://es.mathworks.com/help/lidar/ref/segmentgroundsmrf.html> (cons. 14-11-2021).
- [59] *MATLAB Answers – 3D line of best fit.* Mathworks. 7 de juny de 2013. URL: <https://es.mathworks.com/matlabcentral/answers/78363-3d-line-of-best-fit> (cons. 20-11-2021).
- [60] *MathWorks Help Center – findNearestNeighbors: Find nearest neighbors of a point in point cloud.* MathWorks. URL: <https://es.mathworks.com/help/vision/ref/pointcloud.findnearestneighbors.html> (cons. 20-11-2021).
- [61] Antonio Albiol Colomer. *Tratamiento de imágenes. Transformaciones geométricas.* Apunts de classe distribuïts electrònicament a l'alumnat. Universitat Politècnica de València, març de 2018.
- [62] Gary Bradsky i Adrian Kaehler. *Learning OpenCV.* Ed. de Mike Loukides. 1a ed. Sebastopol, Califòrnia, EUA: O'Reilly, set. de 2008, pàg. 163 - 174.
- [63] Gary Bradsky i Adrian Kaehler. *Learning OpenCV.* Ed. de Mike Loukides. 1a ed. Sebastopol, Califòrnia, EUA: O'Reilly, set. de 2008, pàg. 384 - 387.