Universitat de Vic
Escola Politècnica Superior

Treball Final de Màster

Anàlisi d'imatges estereoscòpiques utilitzant mètodes de baixa càrrega computacional

Enric López i Rocafiguera

Màster en Tecnologies Aplicades i de la Informació

Directors: Pere Martí i Puig, Ramon Reig i Bolaño

Vic, setembre de 2012

Universitat de Vic
Escola Politècnica Superior

Resum del Treball Final de Màster
Màster en Tecnologies Aplicades de la Informació

Títol: Anàlisi d'imatges estereoscòpiques utilitzant mètodes de baixa càrrega computacional
Paraules clau: imatges, disparitat binocular, llindars, correspondència, estereoscòpia, temps real, profunditat perceptual.
Autor: Enric López Rocafiguera
Direcció: Pere Martí Puig, Ramon Reig Bolaño
Data: setembre de 2012

Resum

Mitjançant imatges estereoscòpiques es poden detectar la posició respecte de la càmera dels objectes que apareixen en una escena. A partir de les diferències entre les imatges captades pels dos objectius es pot determinar la profunditat dels objectes. Existeixen diversitat de tècniques de visió artificial que permeten calcular la localització dels objectes, habitualment amb l'objectiu de reconstruir l'escena en 3D. Aquestes tècniques necessiten una gran càrrega computacional, ja que utilitzen mètodes de comparació bidimensionals, i per tant, no es poden utilitzar per aplicacions en temps real.

En aquest treball proposem un nou mètode d'anàlisi de les imatges estereoscòpiques que ens permeti obtenir la profunditat dels objectes d'una escena amb uns resultats acceptables. Aquest nou mètode es basa en transformar la informació bidimensional de la imatge en una informació unidimensional per tal de poder fer la comparació de les imatges amb un baix cost computacional, i dels resultats de la comparació extreure'n la profunditat dels objectes dins l'escena. Això ha de permetre, per exemple, que aquest mètode es pugui implementar en un dispositiu autònom i li permeti realitzar operacions de guiatge a través d'espais interiors i exteriors.

Universitat de Vic
Escola Politècnica Superior

Abstract of final work of Master
Master of Applied Information Technology

Title: Analysis of stereoscopic images using low computational methods
Keywords: images, binocular disparity, thresholds, correspondence, stereoscopy, real-time, depth perception.
Author: Enric Lopez Rocafiguera
Director: Pere Martí Puig, Ramon Reig Bolaño
Date: September 2012

Abstract

Using stereoscopic images, the position with regard to the camera of the objects appear in a scene, can be detected. From differences between the captured images by the two objectives we can determine the depth of objects. Diversity of computer vision techniques exists that allow you to calculate the location of the objects, usually with the aim to reconstruct the scene in 3D. These techniques require large computational burden, since they use two dimensional comparison methods, and therefore can not be used for real-time applications.

In this project we propose a new method of analysis of stereoscopic images that allows us to obtain the depth of objects of a scene with acceptable results. This new method is based on transforming the two-dimensional image information into one-dimensional information in order to make a comparison of the images with a low computational cost, and use the comparison results to extract the depth of objects in the scene. This should allow, for example, this method can be implemented in a standalone device and allows doing guidance operations through indoor and outdoor spaces.

Agraïments

Primerament, vull donar les gràcies a en Pere Martí i en Ramon Reig per haver-me dirigit aquest treball final de Màster. Han estat molts mesos aconsellant-me i apretant-me mig en broma mig seriosament fins que l'he finalitzat. També he de dir que la resta de membres del departament hi ha posat cullerada.

També vull dedicar un espai per agrair a l'Alba pel cop de ma que em va donar tot i la ressaca que portava aquell dissabte.

Moltes gràcies pel suport que m'han donat la Merci, la M. Àngels i la Meritxell amb els seus missatges animant-me a través de totes les xarxes socials possibles.

Una menció molt especial es mereix la Beti, ja la coneixeu, hem estat companys de Màster, després de feina i ara puc dir que tinc una amiga. Gràcies per haver estat sempre aquí per oferir-me suport.

Per últim vull dedicar aquest treball a l'Elisabet, la Carla i la Marta, les meves dones, que saben perfectament el que m'ha costat compaginar el Màster amb la feina, les reunions i la vida al seu costat, que aquest any s'han quedat sense vacances, m'han hagut de suportar en els moments baixos, i tot i això, em segueixen estimant.

Gràcies a tots!!

Índex

Índex de figures

1. Introducció

El sistema de visió humana pot descriure automàticament una textura en detall, un contorn, un color, una representació bidimensional d'una tridimensional, ja que pot diferenciar imatges de diferents persones, de signatures, colors, pot diagnosticar malalties à partir de radiografies, etc. De totes maneres, encara que algunes d'aquestes tasques es poden dur a terme utilitzant visió artificial, el software o el hardware no arriben a aconseguir els resultats desitjats.

El tema de la visió estereoscòpica, durant molts anys ha rebut molta atenció des del punt de vista psicofísic, i en els darrers anys s'ha produït un considerable augment de l'interès en els temes de processament d'imatge, això ha comportat l'aparició de nous mètodes teòrics i desenvolupaments pràctics per al disseny d'aquest tipus de sistemes.

Les aplicacions del processament d'imatge són molt variades, i inclouen aspectes com el mesurament remot, l'anàlisi d'imatges biomèdiques, la simulació de cirurgia guiada remota, el reconeixement de caràcters, aplicacions de realitat virtual i realitat augmentada en sistemes col·laboratius, entre d'altres. Una anàlisi ràpida de les publicacions recents revela que una gran quantitat de contribucions tracten el problema de l'anàlisi d'imatges i escenes.

L'objectiu general del processament d'imatge és analitzar imatges d'una determinada escena i reconèixer el seu contingut per poder prendre decisions en funció dels resultats. Molts tipus d'escenes són bàsicament bidimensionals, com el reconeixement de caràcters, o el tractament de la majoria d'imatges biomèdiques, però en altres situacions les escenes que es volen descriure o interpretar són tridimensionals.

Les imatges que es reben en cadascun dels ulls en els humans són pràcticament iguals, amb una diferència en la posició relativa dels objectes. Això és degut a la posició dels ulls i la forma de moure'ls. Aquestes diferències relatives a la posició a cada imatge (disparitat), tenen una relació directa amb la distància (profunditat) a la qual es troben els objectes entre si, i de l'observador. El cervell és capaç d'interpretar aquesta diferència i reconstruir l'estructura de l'escena que veu l'observador.

Hi ha tres fases en el procés de recuperació de l'estructura d'una escena segons *Marr i Poggio* [1]:

- Seleccionar un punt característic d'un objecte en una de les imatges.
- Trobar el mateix punt característic en l'altra imatge.
- Mesurar la diferència relativa (disparitat) entre la posició d'aquests dos punts.

La visió estereoscòpica consisteix en aquesta capacitat de recuperar l'estructura tridimensional d'una escena a partir de dues vistes o de dues imatges diferents d'ella. L'estructura que es recupera és la posició dels objectes presents a l'escena, bàsicament recuperant la profunditat (distància a l'observador) dels objectes.

Una altra alternativa és la de localitzar per a cada punt de cadascuna de les imatges el seu corresponent en l'altra imatge, entenent per punts corresponents aquells que són projeccions d'aquest punt de l'espai en cadascuna de les imatges. D'aquesta manera es recupera una imatge plena de profunditats per a cadascun dels punts de l'escena projectats en ambdues imatges.

En els darrers anys s'han realitzat molts treballs [2] sobre el tema de la visió estereoscòpica basats en el càlcul de la disparitat, amb solucions ajustades a diferents requeriments i utilitzant diferents tècniques. Dins de les tècniques més utilitzades es troben les que es basen en Programació Dinàmica [1], que tracta d'obtenir la solució òptima subdividint el problema i trobant les solucions òptimes dels subproblemes, i les més recents les basades en Tall de Grafs[3], consistents en obtenir un graf d'un node per a cada píxel de la imatge representant la superfície de profunditats de l'escena.

Des del punt de vista d'errors generats, les tècniques de Tall de Grafs donen millors resultats. Una de les aplicacions d'una imatge densa de profunditats de l'escena és la descomposició d'una imatge en capes d'igual profunditat per al seu posterior processament i generació de noves vistes (*Image Based Rendering*). S'utilitza en la reconstrucció tridimensional d'un objecte a partir de diverses vistes o una seqüència de vídeo. També en la navegació de robots, creació de realitat virtual, codificació d'imatges estereoscòpiques, seguiment i vigilància (comptatge de persones), etc.

La memòria està organitzada en sis capítols. En aquest primer capítol fem una introducció general i el plantejament dels objectius que intenten resoldre el problema que es planteja. En el capítol 2 realitzem una revisió dels principis bàsics i l'estat de l'art de l'estereoscòpia. En el capítol 3 hem exposat els motius perquè hem decidit plantejar aquest mètode d'anàlisi. En el capítol 4 expliquem els conceptes fonamentals d'estereoscòpia i les principals tècniques de processament d'imatge utilitzades en aquest camp. En el capítol 5 exposem els mètodes utilitzats i els resultats obtinguts. Finalment en el capítol 6 extraurem les conclusions i exposarem les possibles línies de treball.

1.1. Objectius

L'objectiu principal d'aquest treball és l'anàlisi d'imatges estereoscòpiques per tal de poder identificar-ne el contingut i obtenir la profunditat en que es troben els objectes que apareixen a l'escena, amb la particularitat de que el mètode tingui un baix nivell computacional. Això permetria poder implementar aquest mètode en aplicacions en temps real com pot ser la navegació de robots. El desenvolupament el realitzarem mitjançant el programa MATLAB.

Així doncs, primer analitzarem algunes de les tècniques d'anàlisi d'imatges estereoscòpiques i veurem com podem aplicar-les per complementar el mètode proposat. Finalment analitzarem les avantatges i desavantatges del mètode aplicat en el nostre treball.

2. Estereoscòpia

No cal dir que la visió és el més complet i alhora complex dels sentits i el que permet a l'home interactuar amb el medi i desenvolupar-s'hi amb eficiència gràcies a la seva adaptabilitat i robustesa com a sistema de percepció. Entre altres característiques, la informació perceptual permet recuperar color, textura, forma i ubicació dels objectes sense necessitat d'arribar a un contacte físic. Degut a la importància de la informació visual com a font de dades del món real, és interessant plantejar la possibilitat de proveir a un ordinador d'aquest 'sentit', que li permeti d'interactuar eficientment amb el medi ambient dinàmic en què es desenvolupa.

La visió artificial ha tractat de reproduir les funcions del sistema visual mitjançant l'anàlisi i processament d'imatges obtingudes des de càmeres de vídeo. S'han proposat moltes tècniques per intentar arribar a aquest objectiu i, tot i que cadascuna proposa enfocaments diferents totes es poden considerar complementàries [4]. Aquest treball està basat en la tècnica de detecció d'objectes a partir de estereoscòpia, la qual té com a model funcional l'estructura binocular sobre la qual opera el sistema visual humà.

2.1. Visió estereoscòpica

Donada la disposició dels ulls en els humans, les imatges que es reben en cada retina són pràcticament iguals, amb una petita diferència en la posició relativa dels objectes. Aquestes diferències en la posició relativa dels objectes en cada imatge és el que coneixem amb el nom de **disparitat binocular**, i té una relació directa amb la distància (profunditat) a la qual es troben els objectes entre si, i per a l'observador. El cervell és capaç d'interpretar aquesta diferència i reconstruir l'estructura d'aquesta escena.

La visió estereoscòpica la podem entendre com si es tractés de la visió dels dos ulls, esquerre i dret, situats en dues finestres situades una al costat de l'altra. Les vistes de cada ull s'envien per separat al cervell, el qual s'encarrega de combinar-les aparellant les similituds i afegint les diferències, per produir finalment una imatge en estèreo, de manera que percebem la sensació de profunditat, llunyania o proximitat dels objectes que ens envolten (figura 1).

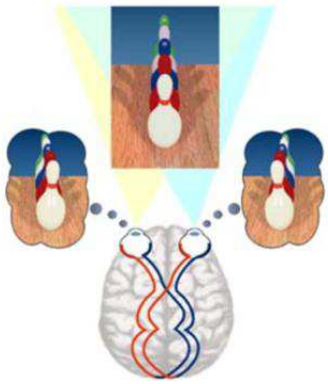


Figura 1. Visió estereoscòpica humana

L'agudesa estereoscòpica és la capacitat de discernir, mitjançant aquest procés, detalls situats en plans diferents i a una distància mínima entre ells. Hi ha una distància límit a partir de la qual no es pot apreciar la separació de plans, i que varia d'unes persones a altres. Així, la distància límit a què es deixa de percebre la sensació estereoscòpica pot variar des d'uns 60 metres fins a centenars de metres. Un factor que intervé directament en aquesta capacitat és la separació interocular o distància interpupil·lar (*DIP*). A major separació entre els ulls, més gran és la distància a la qual apreciem l'efecte de relleu. La distància interocular més habitual és de 65 mm, però pot variar des dels 45 als 75 mm [5].

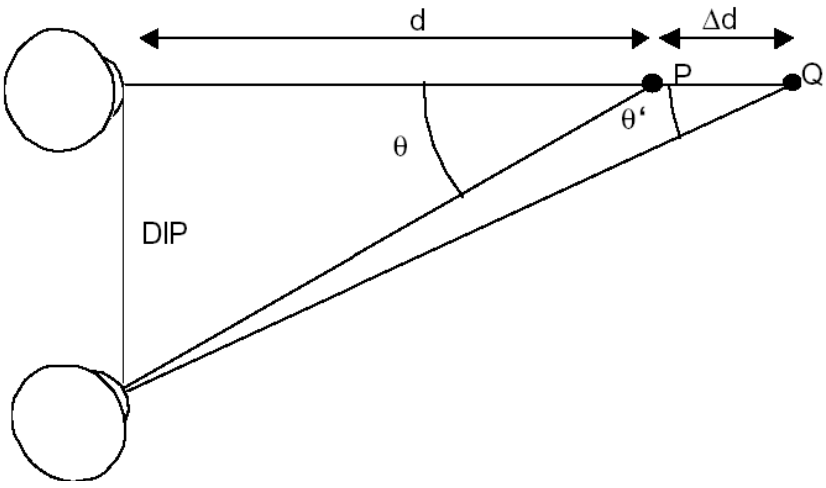


Figura 2. Pla format per l'objecte i els ulls

L'agudesa visual estereoscòpica es defineix com la mínima disparitat binocular necessària per donar sensació de profunditat. Donats dos punts, *P* i *Q*, que formen angles θ i θ', tal i com es mostra a la figura 2, la disparitat binocular serà $\eta = \theta - \theta'$. Si la distància entre els dos objectes és Δd, la disparitat binocular entre aquestes es pot obtenir com:

$$\eta = \frac{\Delta d}{d^2} DIP \qquad (2.1.)$$

Per als ordinadors i altres màquines, aquesta tridimensionalitat és més senzilla de recuperar utilitzant, almenys, dues càmeres orientades cap a la mateixa escena, i que estiguin mirant en perspectives properes però diferents. Analitzant dues imatges, i la geometria del sistema format per les dues càmeres i l'escena, és com es pot recuperar la profunditat de cada un dels objectes visibles. Aquest és l'objectiu de la visió estèreo.

A la figura 3 s'observa l'eix òptic que és la línia imaginària ortogonal al pla imatge i que conté el centre òptic de l'objectiu.

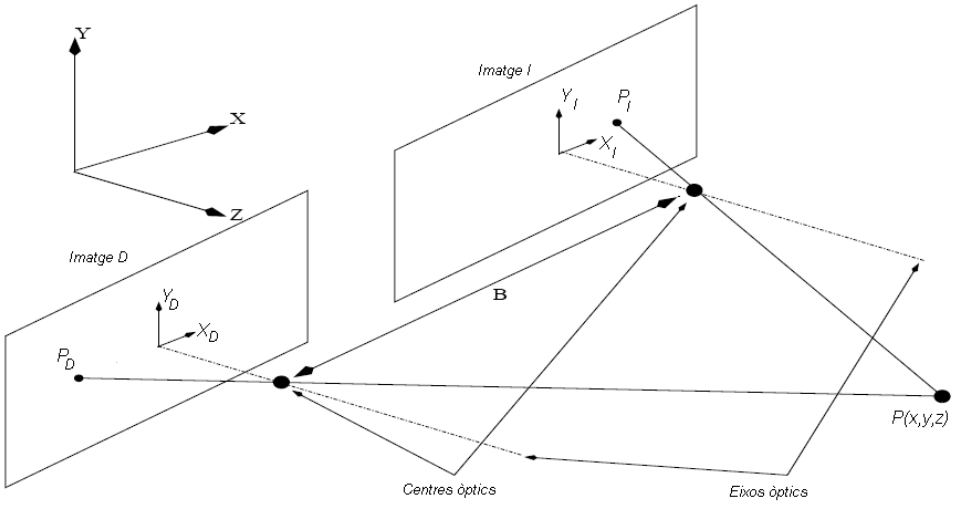


Figura 3. Visió estereoscòpica. Sistemes de referència

En l'anàlisi de l'estereovisió ens trobem amb dos problemes diferents. Partint de dos imatges bidimensionals en l'espai de coordenades (*x,y*), la imatge esquerra (*I*) i la dreta (*D*) de la figura 3:

- El problema de la **correspondència:** tracta de buscar dos punts $p_I(x_I,y_I)$ de la imatge esquerra i $p_D(x_D,y_D)$ de la imatge dreta corresponents a un mateix punt *P* de l'espai tridimensional *(X,Y,Z).*

- El problema de la **reconstrucció:** tracta de trobar les coordenades del punt *P,* un cop trobats aquests dos punts.

L'objectiu més difícil és sens dubte respondre al problema de la correspondència. Com que en general hi ha diverses possibilitats per escollir

l'element corresponent a la imatge *D* d'un element de la imatge *I*, el problema de la correspondència estèreo es diu que és ambigu. Degut a aquesta ambigüitat, cal esbrinar quins elements, quines característiques, quines restriccions i quines consideracions es poden aplicar per reduir-la al màxim.

Primer, cal descriure una sèrie d'elements fonamentals i comuns a totes les geometries emprades en estereoscòpia. Aquests elements són el **pla epipolar**, les **línies epipolars** i els **epipols**. A la figura 4, considerem el pla epipolar aquell que formen els dos centres òptics O_I i O_D dels objectius de les càmeres amb qualsevol punt *P* de l'espai objecte. D'altra banda, el pla epipolar $(O_I\ P\ O_D)$ talla a les dues superfícies imatge *I* i imatge *D* per les rectes ep_I i ep_D, que són les anomenades línies epipolars. Finalment, la projecció del centre òptic de cada càmera sobre l'altra càmera defineix l'anomenat epipol *e*.

Els epipols de cadascuna de les càmeres (e_I i e_D) seran els punts pels que passaran totes les línies epipolars. Utilitzant qualsevol pla epipolar com a base, tots els punts de l'escena que pertanyen a aquest pla tindran la seva imatge en cadascuna de les línies epipolars de les dues imatges. Això implica que qualsevol dels píxels d'una línia epipolar, sigui dreta o esquerra, tindrà la seva corresponent dins de la línia epipolar corresponent en l'altra imatge.

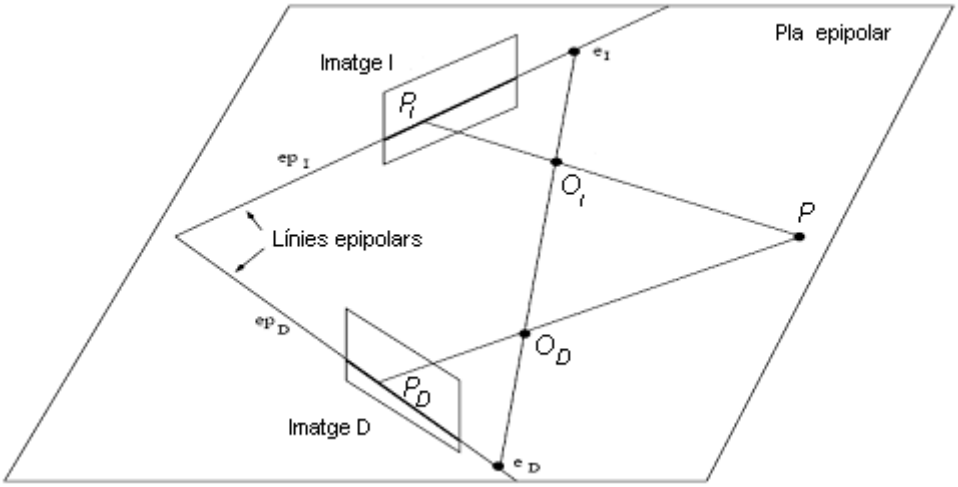


Figura 4. Pla epipolar, línies epipolars i epipols

En general, les línies epipolars són obliqües respecte del pla retinal o superfície fotosensible de la càmera, excepte el cas que considerem que els plans retinals són coincidents entre si, i paral·lels a la línia base (O_I,O_D) que uneix els dos centres òptics de les càmeres. Observarem que els epipols d'ambdues càmeres se situaran en l'infinit, i per tant, les línies epipolars seran totes paral·leles entre si, i paral·leles a la línia base. Aquesta configuració especial s'anomena configuració de **càmeres paral·leles**. Amb aquesta configuració, i un posicionament adequat dels plans retinals es pot aconseguir

que les línies epipolars coincideixin amb les files de les imatges digitals obtingudes.

Les dues geometries bàsiques en visió estèreo, són per doncs, la geometria de càmeres paral·leles i la geometria de càmeres convergents.

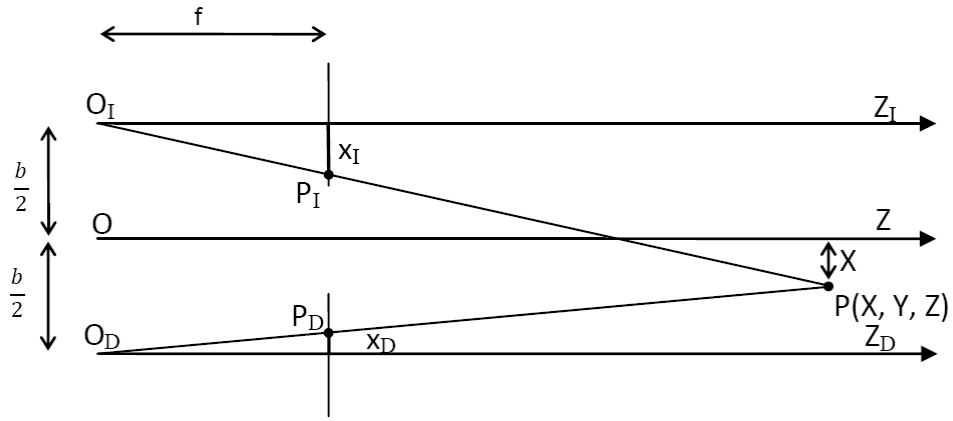


Figura 5. Configuració de càmeres paral·leles

La configuració de càmeres paral·leles és la més senzilla de tractar quant a geometria, i és la que s'analitza en aquest treball. Per realitzar aquesta anàlisi es considera que el sistema de referència de l'escena té el seu origen en el centre òptic de la càmera esquerra, el seu eix *X* coincideix amb la línia base, i el seu eix *Z* coincideix amb l'eix òptic de la càmera. A la figura 5 es representa aquesta configuració considerant exclusivament un pla epipolar.

L'objectiu és trobar la posició del punt *P(x,y,z)* partint de les coordenades de les projeccions d'aquest punt sobre els plans d'imatge (x_I,y_I) i (x_D,y_D). Per això es necessita la distància focal *f* de les càmeres i la distància entre els seus dos centres òptics o línia base *B*. La reconstrucció tridimensional es pot resoldre per geometria euclidiana de forma senzilla i amb uns resultats prou aproximats.

Per la imatge *I*, per semblança de triangles:

$$\frac{x_I}{f} = \frac{x}{z} \quad \rightarrow \quad x = \frac{x_I}{f} \cdot z \qquad (2.2.)$$

I de la mateixa manera per a la direcció *Y*:

$$\frac{y_I}{f} = \frac{y}{z} \quad \rightarrow \quad y = \frac{y_I}{f} \cdot z \qquad (2.3.)$$

Per la imatge *D* es té:

$$x = \frac{x_D}{f} \cdot z - B \quad , \quad y = \frac{y_D}{f} \cdot z \qquad (2.4.)$$

Desenvolupant aquestes equacions podem arribar a les expressions següents:

$$x = \frac{x_I B}{d} \quad , \quad y = \frac{y_I B}{d} \quad , \quad z = \frac{f B}{d} \qquad (2.5.)$$

On el valor *d* és la disparitat, que fa referència a la diferència entre les coordenades x_I i x_D respecte del centre de les seves imatges. Per a l'obtenció d'aquestes coordenades x_I i x_D s'ha pres com a origen de coordenades de cada imatge *(X,Y)* el punt de tall de l'eix òptic i el pla retinal.

$$d = x_D - x_I \qquad (2.6.)$$

Al conjunt de totes les disparitats entre dues imatges d'un parell estèreo s'anomena **mapa de disparitat**. Clarament, les disparitats només es poden calcular a partir d'aquelles característiques que són visibles en les dues imatges. Les característiques que només es veuen en una imatge i no en l'altra es denominen **oclusions**. Amb aquestes expressions podem dir que, un cop conegudes la distància focal de les càmeres, la línia base i la disparitat entre els píxels corresponents, és senzill calcular les coordenades *(x,y,z)* del punt *P* de l'espai per a la configuració de càmeres paral·leles.

La configuració de càmeres paral·leles s'utilitza sovint per la seva simplicitat, però essent físicament possible, a la pràctica resulta difícil alinear dos sistemes òptics de forma tan precisa i prou estable. En el cas general en què els eixos òptics de les càmeres convergeixen cap a un punt finit de l'espai objecte, el procediment a seguir seria el següent: en primer lloc calcular la posició dels epipols e_I i e_D de cadascuna de les càmeres. Amb això, qualsevol punt de la imatge esquerra m_I estarà contingut en una línia epipolar ep_I que l'uneix amb l'epipol esquerre. Aquesta línia epipolar tallarà el pla imatge de la càmera dreta en un punt, que juntament amb l'epipol dret e_d formarà la línia epipolar dreta ep_D, sobre la qual caldrà buscar la seva correspondència. A més, la relació entre cada sistema de referència *(X,Y)* de cadascuna de les imatges amb el marc de referència *(X,Y,Z)* no és tan trivial com en el cas anterior. Cal transformar les components dels vectors *(x,y)* en coordenades *(x,y,z)* de l'escena.

Tenint en compte aquesta complexitat, el que se sol fer és aplicar un procés de rectificat per convertir una geometria general de càmeres convergents en una geometria més simple de càmeres paral·leles.

Si partim del gràfic de la figura 3, el problema de la reconstrucció es pot resoldre mitjançant la intersecció de les rectes $<O_I,P_I>$ i $<O_D,P_D>$. El resultat dependrà de amb quina precisió es coneixen les posicions de O_I i O_D i els plans dret i esquerre en el sistema de coordenades. Això ens porta al problema

del calibratge, ja que, si les posicions P_I i P_D no es coneixen amb precisió, les rectes $<O_I,P_I>$ i $<O_D,P_D>$ podrien no tallar-se.

Mitjançant el calibratge d'un sistema d'estereovisió s'estimen els paràmetres intrínsecs (distància focal, centre òptic i distorsions de les lents) i extrínsecs (posicions relatives i orientacions) de les càmeres que el componen. Hi ha dos mètodes usats habitualment per al calibratge: l'autocalibratge i el calibratge fotogramètric. A l'autocalibratge es prenen diverses imatges d'una mateixa escena i mitjançant la correspondència entre punts de diferents imatges es poden trobar els millors paràmetres del model que puguin atorgar aquesta correspondència. La reconstrucció de l'escena tridimensional realitzada amb el model trobat no cal, ja que aquesta està afectada per un factor d'escala. Amb aquest mètode no es pot saber quina és la mida real dels objectes captats per les càmeres. Això és així ja que un objecte petit a prop del centre òptic podrà tenir la mateixa imatge que el mateix objecte més gran allunyat del centre òptic. Si el que es busca és una reconstrucció precisa, com és el cas en moltes de les aplicacions de la robòtica, és recomanable utilitzar el calibratge fotogramètric. Aquest calibratge utilitza un objecte tridimensional de referència la geometria del qual és coneguda a la perfecció

2.2. Antecedents i estat de l'art

Euclides i Leonardo da Vinci ja van observar i estudiar el fenomen de la visió binocular. També l'astrònom Kepler va deixar uns estudis que comentaven els seus principis [6].

Posteriorment, al 1838, el físic Sir Charles Wheatstone va construir el primer aparell que permetia percebre la tridimensionalitat partint de dues imatges (visor estereoscòpic). Aquest fet curiosament succeí abans del descobriment de la fotografia.

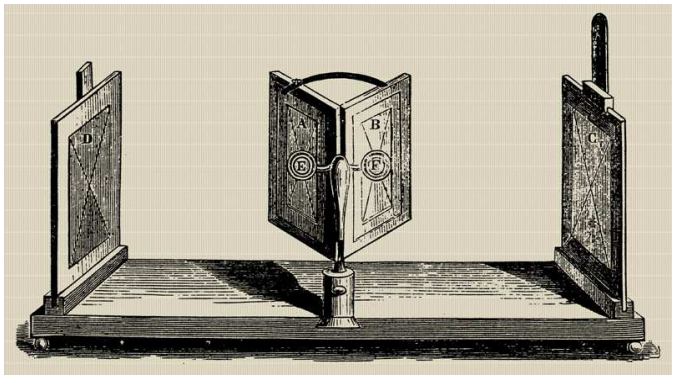


Figura 6. Estereoscopi amb miralls de Wheatstone (Font Wikipèdia)

A l'any 1849, Sir David Brewster va dissenyar i construir la primera càmera estereoscòpica. La càmera disposava de un visor que permetia veure les imatges que passaven per les lents. Alguns anys mes tard, el 1862, Oliver Wendell Holmes construí el que seria l'estereoscopi de mà més popular del s. XIX.

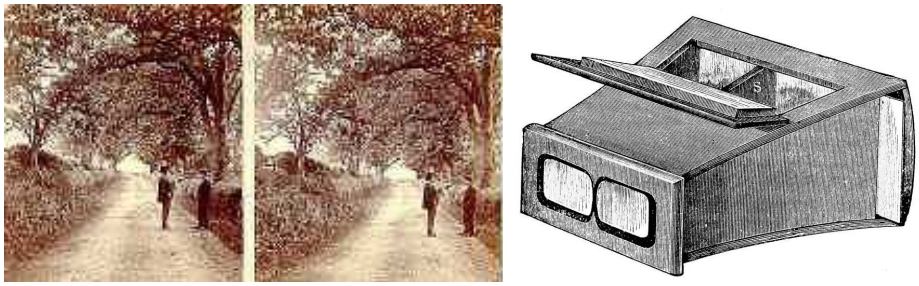


Figura 7. Imatge estereoscòpica. Càmera de Brewster (Font Wikipèdia)

Als anys 30 varen ressorgir la estereofotografia amb l'aparició de les càmeres 3D, amb pel·lícules de 35mm. com Realist o la ViewMaster.

A mitjans de segle XX hi van haver diferents intents d'impulsar les pel·lícules 3D sense gaire èxit, ja que les tècniques utilitzades provocaven problemes de visió. No va ser fins als 80s que amb la creació de pel·lícules de gran format i amb alta resolució, com IMAX 3D, que no s'ha popularitzat.

El principal problema és trobar la forma perquè cada ull vegi la imatge que li pertoca. Per això hi ha diversos sistemes, alguns dels quals descriurem a continuació:

2.2.1. Visió lliure

Hi ha dos mètodes [7]:

- Paral·lela:

  Els ulls observen cada un la seva imatge corresponent, mantenint els seus eixos òptics paral·lels, és a dir, com si miréssim a l'infinit. Només es pot utilitzar aquest mètode amb imatges amb menys de 65 mil·límetres de separació. És el mètode usat per veure les imatges dels llibres amb estereogrames de punts aleatoris.

- Creuada

  Les imatges s'observen creuant els eixos òptics dels ulls. Amb l'ull dret mirar la imatge esquerra i al revés. Aquest mètode es pot utilitzar amb imatges de dimensions grans, tot i que la imatge virtual apareix més petita.

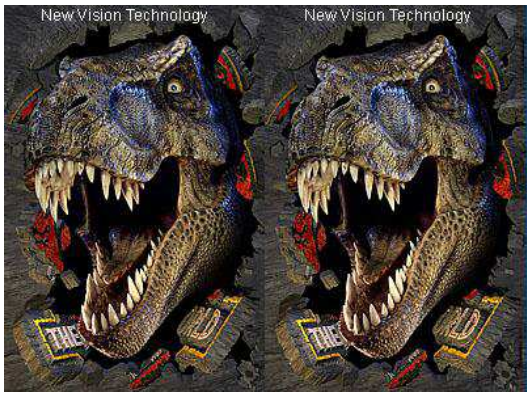


Figura 8. Visió lliure creuada

2.2.2. Anàglif

Les imatges d'anàglif es componen de dues capes de color, superposades però mogudes lleugerament una respecte a l'altra per a produir l'efecte de profunditat. Usualment, l'objecte principal està en el centre, mentre que el al voltant i el fons estan moguts lateralment en direccions oposades. La imatge conté dues imatges filtrades per color, una per a cada ull. La imatge presentada per exemple en vermell no és vista per l'ull que té un filtre del mateix color, però sí que veu l'altra imatge en blau o verd. Aquest sistema, pel seu baix cost, s'utilitza sobretot en publicacions, així com també en monitors d'ordinador i en el cinema. Presenta el problema de l'alteració dels colors, pèrdua de lluminositat i cansament visual després d'un ús prolongat.

Figura 9. Imatge d'anàglif (Font Wikipèdia)

2.2.3. Head Mounted Display (HMD)

Un HMD es un casc estereoscòpic que porta dues petites pantalles LCD amb dues lents, una per a cada ull. Aquest sistema genera les imatges en el propi dispositiu. Pot utilitzar-se per mostrar pel·lícules, imatges o videojocs, però el principal us fins ara ha estat la realitat virtual degut al seu cost prohibitiu.

Figura 10. Un HMD amb les dues pantalles davant de cada ull (Font Wikipèdia)

2.2.4. Sistemes de polarització

Utilitzen un tipus d'ulleres 3D polaritzades amb la mateixa alineació que els filtres polaritzats de les lents dels projectors. En la majoria de projectors i ulleres, les orientacions de polarització són en forma de "V" (45º/-45º). S'utilitza llum polaritzada per separar les imatges esquerra i dreta. El sistema de polarització no altera els colors, encara que hi ha una certa pèrdua de lluminositat. S'usa tant en projecció de cinema 3D com en monitors d'ordinador mitjançant pantalles de polarització alternativa. Avui dia és el sistema més econòmic per a una qualitat d'imatge acceptable. La pantalla ha de ser platejada (alumini), ja que preserva la polarització de la llum projectada. No obstant això no totes les pantalles platejades serveixen.

2.2.5. Imatge alternativa

Amb aquest sistema es presenten en seqüència i alternativament les imatges esquerra i dreta, sincronitzadament amb unes ulleres dotades amb obturadors de cristall líquid (anomenades LCS, *Liquid Crystal Shutter glasses* o LCD, *Liquid Crystal Display glasses*), de manera que cada ull veu només la seva corresponent imatge. A una freqüència elevada, el parpelleig és imperceptible. S'utilitza en monitors d'ordinador, TV i cinemes 3D d'última generació.

Figura 11. Ulleres LCS d'última generació (Font Wikipèdia)

2.2.6. Autoesteroscòpia

És el mètode per reproduir imatges tridimensionals que puguin ser visualitzades sense que l'usuari hagi d'utilitzar cap dispositiu especial (com ulleres o cascos especials) ni necessiti condicions especials de llum. Gràcies a aquest mètode, l'observador pot apreciar profunditat encara que la imatge està produïda per un dispositiu pla. Actualment s'estan desenvolupant monitors que no utilitzant variants del sistema lenticular, es a dir, microlents cilíndriques col·locades sobre la pantalla del monitor, que generen una certa desviació a partir de dos o més imatges.

2.3. Càmera estereoscòpica

En aquest apartat es descriu la càmera estereoscòpica amb la qual s'han obtingut les imatges i les seves característiques. Tot i que en el treball partim de la imatge estèreo i els resultats no estan directament vinculats a la càmera utilitzada, si que indirectament ens afecta i és interessant de conèixer.

Figura 12. Imatge de la càmera estereoscòpica Bumblebee2

La càmera emprada és la Bumblebee2, model BB2-COL-ICX424 (640x480 Color 3.8mm) del fabricant canadenc Point Grey [8]. Les principals característiques d'aquesta càmera són les que indiquem a continuació:

- Dues CCDs Sony ICX204 d'1/3 ", *Color progressive scan*.
- 640x480 píxels quadrats fins a 48 fotogrames per segon.
- Convertidor analògic / digital de 12 bits.
- Línia base (distància entre càmeres) 120mm.
- Distància focal de l'òptica: 3,8 mm amb 70º de camp de visió horitzontal.
- Dimensions: 157mm x 36mm x 47.4mm.
- Pes: 342 grams.
- Interface: 6-pin IEEE-1394 (FireWire) per al control de la càmera i transmissió de les dades de vídeo.
- Alimentació subministrada a través del port IEEE-1394.
- Consum < 3W.

Aquesta càmera, que observem a la figura 12, és ideal per a aplicacions com ara el seguiment de persones, reconeixement de gestos i postures, robòtica mòbil i altres aplicacions de visió per computador. Està precalibrada per corregir les distorsions de les lents i els desajustos temporals i espacials, de manera que no cal fer-ho manualment. La informació de calibratge està precarregada a la càmera, permetent que el programari recuperi la correcció de la imatge. Això admet l'intercanvi entre diferents càmeres, o recuperar la informació correcta quan hi ha múltiples càmeres al mateix bus de connexió FireWire, i totes les càmeres es sincronitzen automàticament, el que és molt important per a la reconstrucció 3D a partir de múltiples punts de vista.

En la configuració de la càmera hi ha dos paràmetres que afecten de manera important a l'enregistrament de les imatges. Un és el balanç de blancs que és una funció de la càmera per compensar els colors de llum emesa per diferents fonts d'il·luminació. La raó per ajustar el balanç de blancs és eliminar els colors no reals de l'escena, de manera que els objectes que són blancs es vegin blancs a la imatge. Es realitza per via electrònica, sobre la base d'un objecte blanc de l'escena gravada.

Els ulls humans són molt bons en el que és blanc sota diferents fonts de llum. No obstant això, les càmeres solen tenir grans dificultats amb l'autobalanç de blancs. Un balanç de blancs incorrecte pot crear imatges tintades de blau, taronja, o fins i tot verd, que no són realistes i perjudicials per al seu tractament i obtenció de característiques, tal i com observem a la figura 13.

Figura 13. Efecte del balanç de blancs

L'altre és la velocitat d'obturació (*shutter*) que mesura el temps que es manté obert l'obturador de la càmera quan es pren una imatge. Com més lenta sigui més gran és el temps d'exposició i més llum arriba al sensor de la càmera. Generalment, una velocitat d'obturació alta congela l'acció de la seqüència gravada mentre que velocitats baixes la desenfoca.

Figura 14. Efecte de la velocitat d'obturació

En aplicacions de visió artificial com més nítida sigui la imatge més informació se'n pot extreure, pel que s'utilitzarà la major velocitat d'obturació possible.

Per al programari de control i reconeixement s'emprarà un sistema dividit en tres parts relacionades entre elles. Primerament, Digiclops proporciona el control de la càmera i la transmissió via firewire a l'ordinador receptor de les imatges. També permet configurar la càmera ajustant el balanç de blancs, la velocitat d'obturació, exposició, guany, etc. El Triclops SDK és una biblioteca de funcions C++ que permet interactuar amb la informació 3D de les imatges. Proporciona una ràpida i precisa generació del mapa de profunditat de l'escena. Això es pot aconseguir aplicant múltiples algoritmes especificant totes les característiques del processat estèreo. Aquests dos elements estan inclosos amb la càmera. La tercera part consisteix en l'obtenció per part de l'usuari dels resultats i la seva presentació en el format desitjat de forma optimitzada i en temps real.

3. Motivació del treball: Obtenció d'un mètode de baix cost computacional

Fent un repàs a les tècniques que hi ha actualment d'anàlisi d'imatges estereoscòpiques i que comentem en el punt 4.2.5., observem que es basen en l'anàlisi bidimensional de les imatges mitjançant l'aplicació de correlacions bidimensionals en els anomenats mètodes basats en àrea, o bé, en extreure característiques de les imatges i buscar similituds en els mètodes basats en característiques.

Per a obtenir la disparitat entre els píxels de les dues imatges i així poder calcular la profunditat dels objectes, les tècniques basades en àrea han de comparar, bàsicament la intensitat, de cadascun dels píxels d'una imatge amb l'altra per trobar el seu píxel corresponent. Aquest càlcul s'ha de fer per tots els píxels de la imatge, per tant, comporta haver de realitzar una infinitat de productes. Per evitar realitzar fer tants càlculs, el que es fa és dividir la imatge en finestres per fer la comparació amb un nombre de píxels més reduït i evitar fer alguns càlculs innecessaris. La clau està en trobar la finestra adequada perquè es pugui establir la correspondència.

Si suposem que les intensitats dels píxels en el punt *(x,y)* de les dues imatges són *I(x,y)* (esquerra) i *D(x,y)* (dreta) i considerem finestres d'amplada *2N+1* píxels, una forma de calcular la correlació creuada [9]:

$$C(d_x, d_y) = \sum_{j=-N}^{N} \sum_{i=-N}^{N} I(x+i, y+j) D(x+i-d_x, y+j-d_y) \qquad (3.1.)$$

Pel que fa als mètodes basats en característiques cal una correcta descripció de les característiques o descriptors que ens permetin trobar les correspondències en les diferents propietats dels objectes. Cal un procés de cerca de cada característica d'una imatge en l'altra per obtenir la corresponent i així calcular la disparitat. Aquest és un procés una mica més ràpid perquè no cal fer la comparació de tots els píxels però tot i així és lent perquè cal realitzar la comparació de les característiques obtingudes prèviament.

La particularitat del mètode que volem implementar és el seu baix nivell computacional. La idea consisteix en convertir una informació bidimensional en una informació unidimensional de forma senzilla i així evitar haver de fer la quantitat de càlculs que aquests mètodes impliquen.

Cal tenir en compte que inicialment és un mètode nou, sense cap referència, però que ens sembla que pot tenir uns resultats interessants. La idea és la següent, enlloc d'utilitzar la intensitat de cadascun dels píxels individualment per a obtenir la correspondència, utilitzar la intensitat d'un conjunt de píxels. Per això, dividirem la imatge en blocs rectangulars petits. Per cada bloc farem la suma dels nivells de gris dels píxels de cada columna fins a obtenir una funció i el mateix farem per cada fila. Obtindrem els resultats per la imatge sintètica de la figura 15.

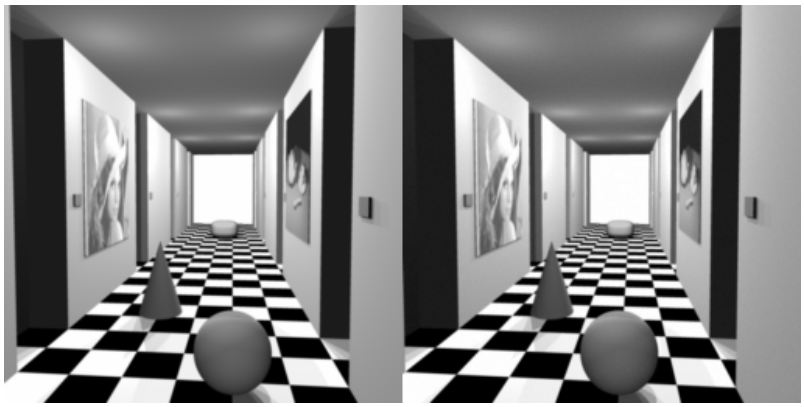


Figura 15. Imatges sintètiques (corridor)

Si dividim cadascuna de les imatges en blocs rectangulars petits i en calculem la suma dels nivells de gris dels píxels per cada columna, obtindrem pel conjunt de blocs d'una filera, les funcions suma mostrades a la figura 16. Podem veure que aquestes funcions ens mostren clarament on hi ha transicions de nivells de grisos, i per tant, on hi ha els límits dels objectes de l'escena. Aquests resultats han estat el motiu que es considerés d'aprofitar aquesta idea per a calcular la disparitat entre els objectes de les dues imatges.

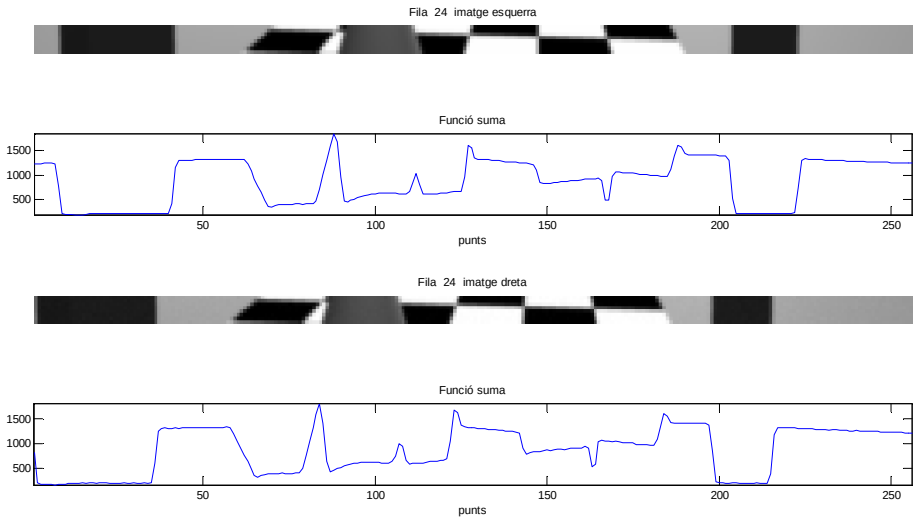


Figura 16. Fila de les imatges esquerra i dreta amb les respectives funcions suma

D'entrada cal dir que aquest mètode servirà per detectar les disparitats a partir de les vores dels objectes, i és probable que si els objectes són grans quedi un buit en el seu interior. Podem intuir també problemes si hi ha objectes que apareixen en el bloc d'una imatge i no apareixen en l'altra.

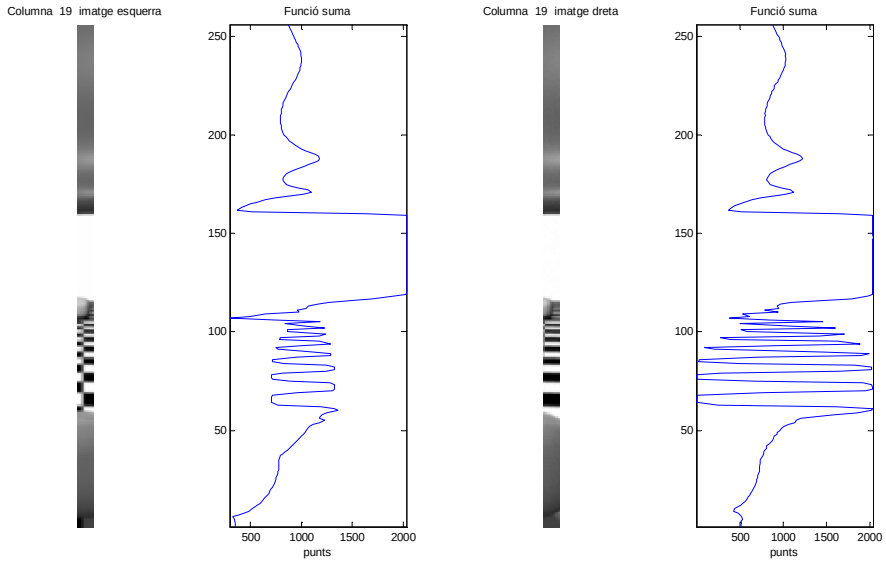


Figura 17. Columna de les imatges esquerra i dreta amb les respectives funcions suma

Aquest mètode doncs ens permetrà calcular la disparitat entre les dues imatges comparant les funcions suma, per tant amb una correlació unidimensional.

Si suposem que les funcions suma d'un bloc de les dues imatges són $s_I(x)$ (esquerra) i $s_D(x)$ (dreta) i considerem finestres d'amplada $2N+1$ píxels, la forma de calcular la correlació creuada serà:

$$C(d) = \sum_{i=-N}^{N} s_I(x+i)s_D(x+i-d) \qquad (3.2.)$$

Aquest càlcul es farà per la fila i per la columna, però el nombre de productes que en resultarà serà molt inferior al dels mètodes actuals.

4. Anàlisi d'imatges

La idea general és que després de processar la imatge, puguem definir una escena com a una representació icònica del món visible, perquè a partir de les dades d'aquesta representació es puguin realitzar les accions pertinents. Aquest model només necessitarà representar els detalls necessaris per a la tasca que vulguem realitzar.

Una vegada ja tenim les imatges a través del sistema de càmeres necessitem adequar les dades al format desitjat per poder realitzar després tot l'anàlisi. Primer cal un filtratge on es duran a terme accions més directes sobre les imatges obtingudes. Aquestes accions poden ser, entre altres, la reducció de soroll, la millora del contrast, el realçament de contorns o la correcció de distorsions. Algunes d'aquestes accions es podran dur a terme a nivell de hardware, és a dir, per mitjà dels recursos proporcionats pel sistema de càmeres.

El processament posterior es pot considerar la part principal del sistema, ja que serà on s'apliquin totes les tècniques d'anàlisi estereoscòpic. Serà aquí on s'implementaran els algoritmes per a l'anàlisi de disparitat i correspondència dels parells estereoscòpics i on s'obtindran les mesures de distància a la càmera, per tant, serà on s'aplicarà la part central d'aquest treball.

En el nostre cas, tenim un entorn interior on el que ens interessa identificar són els objectes de l'escena més propers a la càmera per tal de poder fer, per exemple, el guiat automàtic d'un robot de forma que es produeixin el mínim de col·lisions possibles amb els objectes del seu entorn, com són les portes, mobles, parets, terres, persones, etc.

El resultat de processament, tot i que no és l'objectiu d'aquest treball, ens podrà ajudar a reconstruir l'escena en una fase posterior. Aquesta reconstrucció consisteix en, a partir del model icònic obtingut després de les fases anteriors, representar els objectes de l'entorn.

El primer problema que ens trobem, abans d'arribar a analitzar imatges estereoscòpiques, és el de la pròpia captació de les escenes. El nombre de càmeres que s'utilitzaran, la disposició de les càmeres, i la seva relació entre si, i amb el sistema de coordenades de l'espai tridimensional i de les imatges, possibles defectes de les càmeres, etc. El normal és utilitzar un sistema de captació format únicament per dues càmeres, com és el nostre cas, a

semblança del sistema visual humà. El problema consistirà aleshores en la recuperació de l'estructura 3D a partir de la informació binocular.

La geometria que utilitzem és la d'eixos paral·lels, encara que permet un menor solapament de les imatges projectades que la d'eixos convergents, però, la geometria de projecció és més senzilla, i sovint multitud de casos reals es poden tractar d'aquesta manera.

En el nostre cas, tant la correcció de distorsions, com el calibratge i l'alineació es duen a terme directament des de les funcions proporcionades per la pròpia càmera. La figura 18 mostra el flux de treball per defecte que proporciona el programari distribuït per Point Grey [10].

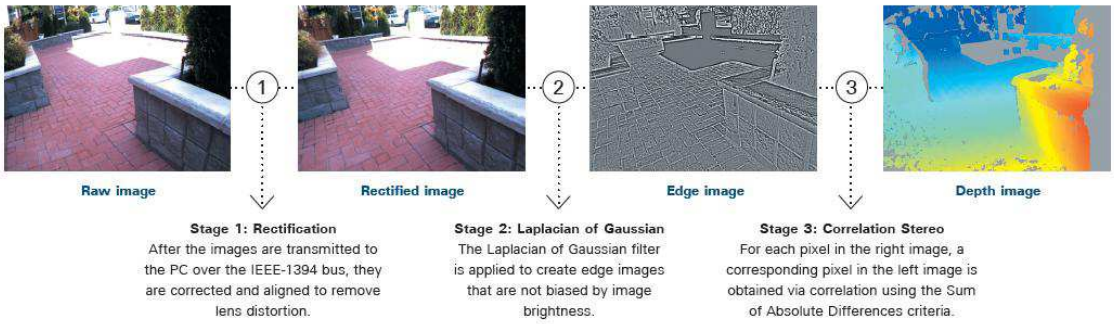


Figura 18. Flux de treball del sistema de captació utilitzat

Com veiem, el primer pas és la rectificació de les imatges transmeses a l'ordinador a través de l'IEEE-1394. Un cop captada la imatge el següent pas és la seva representació en el format adequat per al seu processament. Les eines per a l'adquisició d'imatges transformen la imatge visual d'un objecte físic i les seves característiques intrínseques en un conjunt de dades digitals per processar, així a partir d'aquest moment, una imatge *f(x,y)* estarà donada per les seves coordenades espacials i serà representada matemàticament en una matriu on els índexs de les files i columnes indiquen un punt específic de la imatge.

4.1. Filtratge

Tal i com hem comentat, a vegades pot passar que les imatges que es disposi no siguin adequades per al seu tractament, per diversos motius, o bé simplement sigui convenient realitzar algun tipus de tractament sobre elles per facilitar el treball de les etapes posteriors.

El filtrat és un procés indispensable en tota activitat que involucri el maneig d'imatges. Aquest és el primer procés que s'aplica a les imatges, i fins i tot moltes vegades es realitza en el mateix dispositiu que realitza la captura, per exemple en pràcticament totes les càmeres fotogràfiques digitals s'aplica el filtre de Bayer.

Els filtres permeten destacar característiques particulars de les imatges i descartar les parts no desitjades. El principal objectiu dels filtres consisteix en processar una imatge de manera que resulti més adequada que l'original per a una aplicació específica. De manera general, filtrar una imatge consisteix a aplicar una transformació de manera que s'accentuïn o disminueixin certs aspectes:

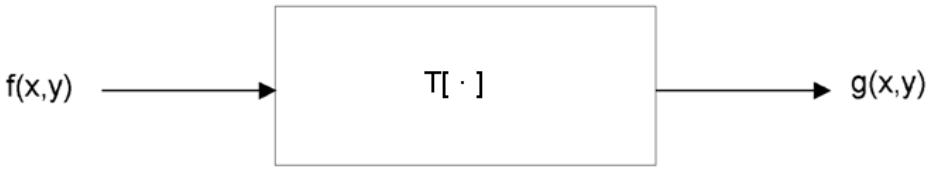


Figura 19. Transformació deguda a un filtre

$$g(x,y) = T[f(x,y)] \qquad (4.1.)$$

Segons el domini on actuen podem dividir-los en filtres de domini espacial (convolució) i filtres en el domini de la freqüència (multiplicació i transformades de Fourier). El filtres en el domini espacial treballen directament sobre els píxels de la imatge i utilitzen generalment matrius anomenades màscares que operen sobre un veïnatge de píxels centrat en el píxel d'interès. Es realitza una convolució (escombrat) de la màscara per la imatge. Aplicant un filtre en una imatge en resulta una de nova obtinguda com el sumatori del producte de la màscara pel veïnatge del píxel. En aquest cas la transformació tindrà la forma:

$$g(x,y) = \sum_i \sum_j f(i,j) w(i,j) \qquad (4.2.)$$

Observem que per a calcular la nova imatge el nombre d'operacions que caldrà fer serà molt gran.

Amb aquestes tècniques bàsiques podem:

- Realçar o emfatitzar determinades característiques. Per exemple, millora del contrast o pseudoacoloriment.
- Filtrar per a eliminar el soroll degut a la funció de dispersió.
- Realçar contorns.
- Realitzar operacions geomètriques d'ampliació, reducció o correcció de distorsions.

4.2. Processament

El processament d'imatges és un processament d'informació en el qual l'entrada és una imatge, com pot ser una fotografia o les seqüències d'un vídeo, però la sortida no és necessàriament una imatge, pot ser per exemple una sèrie de característiques de la imatge. La majoria de les tècniques de processament d'imatges consisteixen en tractar la imatge com un senyal bidimensional i aplicar tècniques estàndard de processament de senyals. Aquestes tècniques requereixen que les característiques de la imatge estiguin ben definides, els contorns ben delimitats i el color i la brillantor siguin uniformes. El tipus de mesures que es vulguin realitzar per a cada característica específica és un factor important per poder determinar els passos apropiats per al seu processament. Els procediments aplicats per al processament d'imatges, per tant, estan orientats a les aplicacions. El que pot ser adequat per a una aplicació pot no ser-ho per a una altra.

4.2.1. Extracció de característiques

Abans de processar la imatge, és necessari dur a terme un preprocessament que tindrà com a objectiu identificar les característiques representatives de cada imatge. Aquestes característiques han de ser elegides acuradament perquè amb elles es duran a terme les anàlisis posteriors (anàlisi de correspondència, disparitat, obtenció de profunditat).

L'extracció de característiques és un procés que, com qualsevol procés d'informació, retorna una sèrie de valors a partir d'una entrada. En aquest cas en particular l'entrada són imatges i la sortida són valors que indiquen la posició de la característica buscada dins de la imatge inicial.

Qualsevol mètode d'extracció de característiques necessita d'algun filtre previ que ressalti les característiques candidates, un exemple molt comú és aplicar filtres de detecció de contorns.

Els contorns són una de les característiques més primàries de qualsevol imatge. Estudis sobre el processament visual humà indiquen que per a la percepció de les escenes, el còrtex visual utilitza informació del color dels objectes, de les textures, de les ombres, però sobretot dels contorns dels objectes. Sembla que hi ha neurones especialitzades a detectar les discontinuïtats de la intensitat lumínica (contorns). Segons la teoria dominant, el còrtex visual treballa en una jerarquia de característiques visuals, on agrupa els elements primaris dels contorns en objectes geomètrics més complexos

fins a la interpretació de les figures. Aquest subsistema biològic és capaç de completar contorns parcialment ocults o explícitament eliminades.

4.2.2. Segmentació

Bàsicament, en qualsevol imatge hi podem trobar un o diversos objectes en un entorn determinat. L'objectiu de la segmentació consisteix en separar aquests objectes del medi en què es troben i distingir-los entre si. En el cas de les imatges amb les que treballarem no és gens simple degut a que presenta una il·luminació que dificulta molt la seva extracció.

La segmentació es basa en els següents principis [11]:

- Similitud: tots els píxels d'un element tenen valors semblants respecte a alguna propietat determinada (nivell de gris, color, textura,..)
- Discontinuïtat: els objectes destaquen de l'entorn i tenen per tant uns contorns definits.
- Connectivitat: els píxels que pertanyen al mateix objecte o regió han de ser contigus, és a dir, tendeixen a agrupar-se constituint regions contínues.

En el nostre cas, els elements que volem separar són els diferents objectes d'una habitació amb una il·luminació deficient, amb reflexos a les parets i finestres. Després del procés de segmentació tindrem una sèrie de regions que indicaran el lloc de la imatge on es troben els diferents objectes cosa que ens permetrà obtenir la seva proximitat al punt d'observació.

Hi ha dos formes clàssiques de segmentar:

i. Segmentació basada en els contorns dels objectes

Una tècnica utilitzada és la detecció de contorns per poder separar lels objectes del fons. De totes maneres aquesta etapa no permet segmentar els objectes de la imatge per la qual cosa li caldrà una etapa de processament posterior per acabar d'elaborar les fronteres dels objectes. En el punt 3.2.3. ampliarem l'explicació d'aquesta tècnica que hem utilitzat en el nostre treball.

La segmentació basada en llindars s'utilitza quan s'observa una clara diferència entre els objectes i el fons de l'escena. La similitud és entre els píxels que formen els objectes i el fons. Es sol utilitzar l'histograma de la imatge on es representa la freqüència relativa d'aparició de cadascun dels nivells d'intensitat en la imatge. Es tracta de posar un llindar que divideixin les diferents línies de l'histograma pel punt on hi

hagi el mínim. Obtindrem una imatge binaritzada tal i com es mostra a la figura 20.

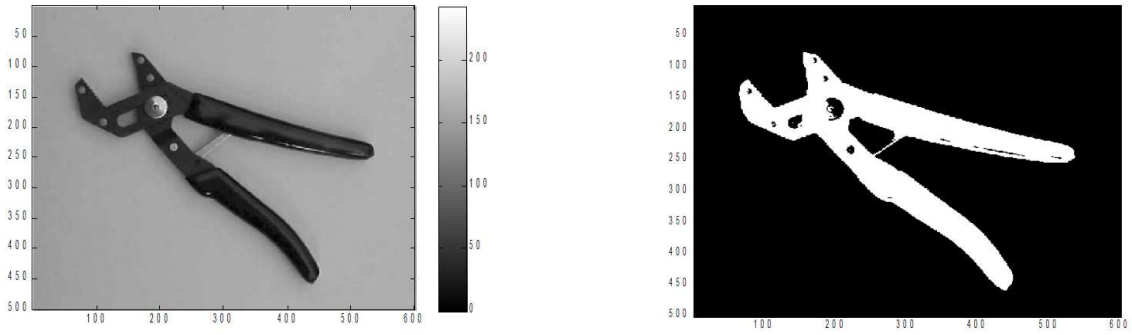


Figura 20. Exemple de segmentació per llindar

ii. Segmentació basada en regions homogènies

Intenta dividir la imatge en particions amb certes característiques comuns segons algun criteri com és la proximitat i similitud de píxels que formen la regió. La imatge es considera que està formada per *n* regions disjuntes, cadascuna d'elles agrupa un conjunt de píxels amb alguna característica en comú. Es sol utilitzar la idea del creixement de regions a partir d'un píxel seguint algun criteri de similitud (figura 21). El problema que té és com escollir la llavor? I quin criteri de similitud seguim?

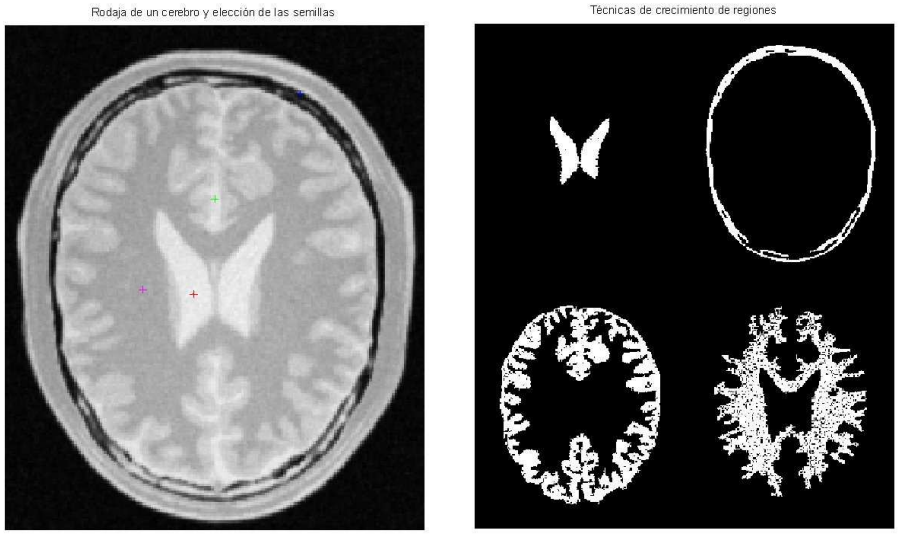


Figura 21. Imatge amb les llavors escollides i resultat de la tècnica de creixement

4.2.3. Detecció de contorns

Un contorn (o vora) és la frontera entre dues regions on els tons de gris difereixen significativament o tenen propietats diferents, com passa en el cas de textures. Si volem detectar els contorns hem de posar èmfasi en els canvis bruscos dels nivells de gris de píxels veïns i suprimir aquelles àrees amb valors de gris constants

Una vora local (aresta local), és un píxel que el nivell de gris difereix significativament del nivell de gris d'alguns píxels del seu entorn. És a dir, hi ha diferència de contrast local. Això és degut bàsicament a que:

a) El píxel forma part de la vora entre dues regions diferents de la imatge.

b) El píxel forma part d'un arc molt fi sobre un fons de diferent nivell de gris.

Per detectar els contorns començarem comentant la detecció de les vores locals. Les vores locals es detecten mesurant la taxa de canvi dels tons de gris del seu entorn. Per això, l'operador gradient (com a operador derivada de primer ordre) o l'operador Laplacià (com a operador derivada de segon ordre) ens permetran fer aquestes mesures [12] (figura 22).

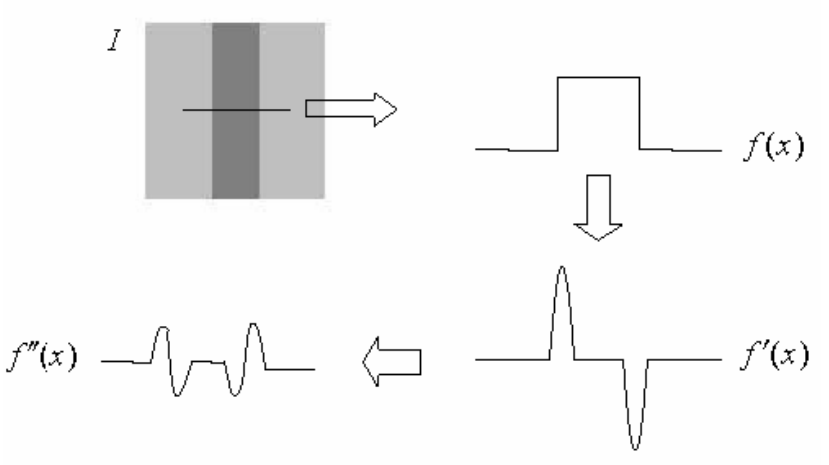


Figura 22. Transició brusca i les seves derivades

Es poden detectar les transicions brusques a partir dels extrems de la primera derivada, o bé, dels passos per zero de la segona derivada. Ho veurem més bé en el gràfic de la figura 23.

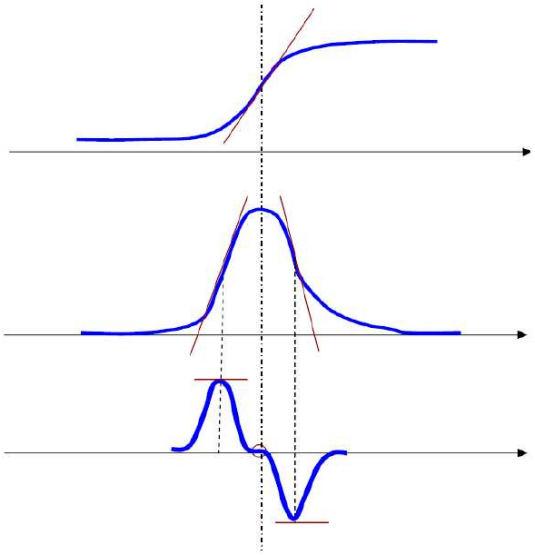


Figura 23. Transició brusca i les seves derivades

Els mètodes basats en el gradient són més adequats quan la transició en els nivells de gris de els píxels de l'entorn és brusca, semblant a una funció esglaó. No obstant això, aquests operadors són pocs sensibles als canvis graduals en els nivells de gris, que corresponen a funcions tipus rampa. En aquests casos és més adequat aplicar operadors basats en les derivades de segon ordre, com, per exemple, el operador Laplacià, definit a partir de la laplaciana de una funció.

Alguns dels algoritmes de detecció de vores més habituals:

- Mètodes basats en gradient:
  - Operador de Roberts
  - Operador de Sobel
  - Operador de Prewitt
  - Operador Frey-Chen
- Operadors basats en creuaments per zero
  - Operador de Marr-Hildreth
  - Operador de Canny

En el nostre cas, i degut a les necessitats de detectar determinats objectes i no detectar ombres degudes a la mala il·luminació, utilitzarem un detector basat en el gradient, en aquest cas hem utilitzat l'operador de Prewitt, tot i que els demés donaven resultats molt similars.

4.2.4. Operador de Prewitt

L'operador de Prewitt, és un operador de gradient, En els píxels dels contorns s'observa un canvi d'intensitat ràpid en alguna direcció, donada pel vector gradient. La magnitud del gradient ens indica quan de marcada és la vora. Si calculem el gradient en regions uniformes obtindrem un valor 0, per tant, que no hi ha vores.

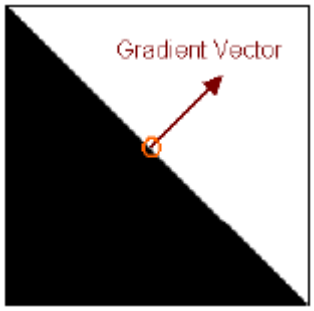


Figura 24. Gradient d'un píxel [13]

El gradient té dos components: el gradient horitzontal i el gradient vertical. Aquests es poden definir a partir de la diferencia entre els píxels veïns en la direcció que marca el gradient. El gradient d'una imatge A serà doncs,

$$\nabla A = [G_V \; G_H] \qquad (4.3.)$$

on

$$G_V = \frac{\Delta A}{\Delta c}, \quad G_H = \frac{\Delta A}{\Delta f} \qquad (4.4.)$$

Són els components horitzontal i vertical.

Matemàticament es poden redefinir els gradients mitjançant filtres que utilitzin un conjunt de píxels per a obtenir els valors corresponents, aquests filtres varien una mica segons el mètode (figura 25) i, tal i com hem comentat anteriorment, donen resultats similars entre ells.

En el nostre cas, tal i com hem comentat, hem implementat l'operador de Prewitt per tal de millorar la detecció dels contorns dels objectes i així millorar el càlcul de la posició de cadascun d'ells.

Roberts	$M_V=\begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix}$	$M_H=\begin{bmatrix}-1 & 0\\ 0 & 1\end{bmatrix}$
Prewitt	$M_V=1/3\cdot\begin{bmatrix}-1 & -1 & -1\\ 0 & 0 & 0\\ 1 & 1 & 1\end{bmatrix}$	$M_H=1/3\cdot\begin{bmatrix}-1 & 0 & 1\\ -1 & 0 & 1\\ -1 & 0 & 1\end{bmatrix}$
Sobel	$M_V=1/4\cdot\begin{bmatrix}-1 & -2 & -1\\ 0 & 0 & 0\\ 1 & 2 & 1\end{bmatrix}$	$M_H=1/4\cdot\begin{bmatrix}-1 & 0 & 1\\ -2 & 0 & 2\\ -1 & 0 & 1\end{bmatrix}$
Frei-Chen	$M_V=1/(2+\sqrt{2})\cdot\begin{bmatrix}-1 & -\sqrt{2} & -1\\ 0 & 0 & 0\\ 1 & \sqrt{2} & 1\end{bmatrix}$	$M_H=1/(2+\sqrt{2})\cdot\begin{bmatrix}-1 & 0 & 1\\ -\sqrt{2} & 0 & \sqrt{2}\\ -1 & 0 & 1\end{bmatrix}$

Figura 25. Matrius gradient pels diferents mètodes

Un cop hem realitzat la detecció de contorns, podem emfatitzar-los mitjançant la **morfologia matemàtica** [11]. La morfologia matemàtica és una tècnica de processament no lineal de la imatge que es basa en la geometria dels objectes. Ens permet extreure components útils per a la caracterització de les regions de la imatge i característiques dels objectes.

La **dilatació** és un operador morfològic que ens permet emfatitzar els contorns obtinguts amb el detector. La dilatació consisteix en escombrar una imatge amb un element estructurant (*EE*) que provoca que s'afegeixin tots els punts del fons que toquin la vora d'un objecte, per tant, és extensiva, alhora reomple els espais on no càpiga l'element estructurant, i per tant, dona continuïtat als contorns. L'element estructurant és un patró que s'utilitza com una màscara de convolució per tal d'examinar l'estructura geomètrica de la imatge. una matriu binaria de píxels amb una forma geomètrica determinada (circular, quadrada, rectangular, etc).

4.2.5. Anàlisi de correspondència

En la visió estereoscòpica, el càlcul de la profunditat d'una escena és l'objectiu final per poder fer una representació tridimensional dels objectes que conté. Per això necessitem calcular la distància que hi ha entre els objectes i la càmera. Aquesta distància s'obté a partir de les disparitats produïdes per la visió binocular. El problema més important a resoldre en aquests sistemes és

el de trobar el conjunt de característiques d'una imatge que es corresponen amb els de l'altra, el que es coneix com el problema de la correspondència.

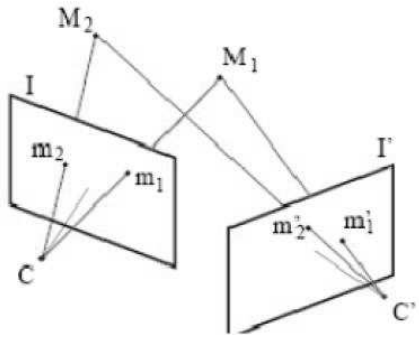


Figura 26. Correspondència de característiques

El problema de la correspondència és, bàsicament, el següent:

Donades dues imatges inicials, trobar les parelles de punts de les dues imatges que es corresponen amb un mateix punt de l'escena o de la imatge en 3D. La posada en correspondència d'imatges és un procés ambigu, ja que, donat un punt en una imatge, hi ha molts punts en l'altra que poden estar en correspondència amb ell. Això pot conduir a diverses interpretacions diferents de la mateixa escena.

El cas més simple seria, tal i com es mostra a la figura 4, que les dues càmeres fossin idèntiques i separades només al llarg de l'eix *X* per una distància de base *B*. És a dir, les imatges fossin coplanars. La informació de profunditat (distància) s'obté a partir del fet que el mateix punt característic en l'escena apareix en una posició lleugerament diferent en els dos plans de les imatges. El desplaçament entre les dues imatges, és el que coneixem com disparitat.

Com ja hem comentat, la relació entre la disparitat d'un punt en el parell d'imatges estèreo és directament proporcional a la profunditat d'aquest punt en l'escena. Si incrementem la distància de la línia de base *B*, ens donarà una estimació més precisa de la profunditat *Z*. No obstant això, com més gran sigui *B*, els angles de vista són més diferents, i es dificulta determinar la correspondència entre les dues imatges.

A més, el problema de la correspondència el dificulten diverses causes:

- Variació de la intensitat: La lluminositat observada d'un punt en l'escena varia en desplaçar la càmera, en funció de l'angle entre la llum incident sobre el punt i l'eix òptic de la càmera.
- Fenomen de l'oclusió: Un punt d'una escena visible en una imatge no és necessàriament visible en l'altra imatge.

- Soroll i errors de mostreig: La digitalització pot comportar imprecisions a causa de la resolució finita de les imatges obtingudes.
- Errors de calibratge: Els errors de calibratge, alineació i rectificació no només dificulten la recerca de punts corresponents, sinó que contribueixen a augmentar els errors en la reconstrucció tridimensional.
- Textura dels objectes: Els objectes observats poden presentar textures que augmenten l'ambigüitat de la recerca de punts corresponents. Una textura de molt alta freqüència (patrons diaris) augmenta el nombre de punts similars, mentre que una textura de molt baixa freqüència (objectes llisos o d'escassa textura) fa la recerca del punt corresponent molt difícil.
- Quan la textura es repeteix, com per exemple en maons d'una paret, poden existir múltiples correspondències. Aquest problema produeix correspondències falses. La majoria dels objectes presents en escenaris naturals té textures. Encara que la textura no necessàriament ajuda a identificar característiques abstractes, com segments, forma la base per als algorismes basats en àrea, ja que presenta un patró estadísticament rellevant.

Les estratègies d'obtenció de correspondència entre dos punts poden classificar-se de diverses formes: Segons les primitives o funcions usades en les tècniques podem classificar-les en:

- Tècniques basades en àrea: Correlació d'àrea.
- Tècniques basades en característiques: Teoria computacional de Marr-Poggio o de Pollard-Mayhew-Frisby, o tècniques basades en contorns.
- Tècniques Jeràrquiques
- Programació Dinàmica: Algorisme de Bircheld i Tomasi.

a) Tècniques basades en àrea

Aquests mètodes consideren les dues imatges captades com a un senyal bidimensional traslladat. Tracten d'obtenir, per a cada punt de la imatge, aquesta translació minimitzant un cert criteri (correlació). Per a cada píxel d'una imatge es calcula la correlació entre la distribució d'intensitats d'una finestra centrada en aquest píxel i una finestra de la mateixa grandària centrada en el píxel a analitzar de l'altra imatge.

Els mètodes basats en correlació i en característiques guarden moltes similituds, doncs el primer que fan tots dos és l'obtenció de punts característics

de la imatge. Obtenen les vores, cantonades o característiques que creuen convenients i tenen com desavantatges problemes com la falta de correspondència entre punts (que un punt visible a la imatge esquerra que està darrere d'un altre objecte a la dreta per exemple) o l'existència de sorolls. La primera diferència que ens trobem entre dos mètodes és que els mètodes de correlació tracten la totalitat dels punts de la imatge mentre que els mètodes basats en característiques només tracten punts característics.

Les tècniques de correlació es basen en minimitzar la diferència en el desplaçament entre trames en un bloc de píxels. En la seva forma bàsica, l'algoritme de correlació de blocs divideix una imatge en una sèrie de regions de la mateixa mida i per a cadascuna de les regions es busca, en la següent trama, la possible correlació en el seu veïnatge, minimitzant un criteri d'error com la diferència en el desplaçament entre trames, o una altra mesura relacionada sobre un conjunt de vectors de moviment, és a dir, el criteri de similitud és una mesura de la correlació entre les finestres en les dues imatges. L'element en correspondència esta donat per la finestra que maximitza el criteri de similitud dins d'una regió de recerca. A la figura 27 es mostra com es defineixen les finestres de correlació i cerca, i les diferents posicions de correlació que es calculen.

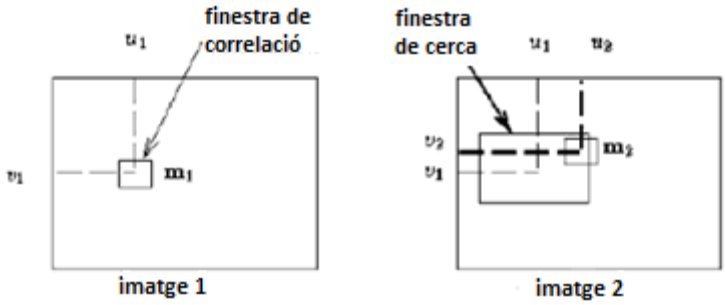


Figura 27. Correlació entre imatges

Una de les tècniques més senzilles és la Suma de Diferències Absolutes *(SDA)* ja que únicament es realitzen operacions amb nombres enters. Donat un píxel de coordenades *(x,y)* en la imatge esquerra, es calcula un índex de correlació *C(x,y,s)* a cada desplaçament *s* de la finestra de correlació en la imatge dreta. La disparitat entre el píxel de la imatge esquerra i del corresponent a la imatge dreta es defineix com el desplaçament *s* que minimitza l'índex de correlació.

El comportament de l'algorisme depèn en gran part de la mida de la finestra de correlació.

Entre els avantatges d'aquestes tècniques estan el obtenir bons resultats sobre imatges amb textura important, permetre crear mapes densos de disparitat i la

seva facilitat de paral·lelitzar processos. Com a inconvenients dir que presenten problemes amb imatges amb elevades discontinuïtats de superfície, que són molt sensibles a variacions fotomètriques degudes a ombres, que requereixen un procés posterior d'eliminació de falses correspondències i que té problemes amb les oclusions. Podem veure els resultats d'aquesta tècnica a la figura 28.

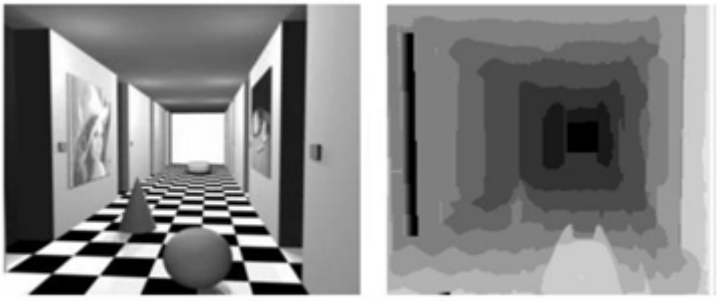


Figura 28. Resultat de l'algorisme SDA aplicat a la imatge corridor [5]

b) Tècniques basades en característiques

Els mètodes basats en característiques obtenen primitives d'alt nivell (contorns, segments, corbes, regions, etc.) que contenen propietats pràcticament invariants a la seva projecció. Les primitives d'alt nivell tenen l'avantatja que contenen millor informació que els nivells d'intensitat dels píxels, permeten utilitzar restriccions geomètriques entre elles i són robustes, tot i que donen una informació molt dispersa.

Entre els mètodes basats en característiques ens trobem la teoria computacional de Marr-Poggio, la teoria de Pollard-Mayhew-Frisby, la tècnica de relaxació de Kim i Aggarwa o les tècniques basades en segments de contorns, per exemple.

Per entendre aquest tipus de tècniques descriurem el model de correspondència basat en característiques de Marr-Poggio [1] basat en una estratègia jeràrquica per solucionar l'establiment de les correspondències. El mètode queda descrit en tres fases:

i) Filtratge de les imatges amb l'operador Laplacià-Gaussià (convolucionar la imatge amb la laplaciana d'una funció gaussiana bidimensional).

ii) Extracció de característiques. En les imatges filtrades es busquen els talls per zero, s'emmagatzema la seva posició, el signe del contrast (si el canvi és de positiu a negatiu o al contrari) i una estimació de l'orientació del contorn en aquest punt.

iii) Correspondència: Per a cada tall per zero a la imatge esquerra es defineix una regió de cerca a la imatge dreta (o al contrari) en línies epipolars corresponents al mateix pla epipolar. S'estableixen com a possibles correspondències aquelles que tenen el mateix signe del contrast i aproximadament la mateixa orientació. Finalment es força la continuïtat de les superfícies.

c) Programació dinàmica

La programació dinàmica es basa en la cerca d'un camí sobre un espai bidimensional, que minimitzi algun tipus de funció de cost. La recerca de correspondències es planteja com un problema d'optimització, descomponent el problema de maneres més senzilles, per això són considerats programació dinàmica. Per a un sistema estereoscòpic d'imatges d'eixos alineats, els punts corresponents s'han de buscar dins de la mateixa línia horitzontal, d'aquesta manera podem definir un espai bidimensional amb els eixos formats per les línies de rastreig de les imatges esquerra i dreta. Els algorismes es basen en la suposició que els contorns conserven l'ordre en un parell d'imatges estereoscòpiques [14].

El major desavantatge d'aquest mètode és que es basa en la recerca entre línies corresponents, i no té en compte les línies adjacents, provocant un efecte ratllat horitzontal. Els intents per utilitzar programació dinàmica amb restriccions en dues dimensions no han donat bon resultats.

Birchfield i Tomasi [15] introdueixen una nova mesura per al càlcul de la semblança entre dos possibles punts corresponents, en lloc d'utilitzar la correlació defineixen el concepte de confiança. Demostren que només modificant aquesta mesura s'obtenen millors resultats, i es pot incloure dins d'altres algoritmes i incrementar el seu rendiment. Presenten doncs un algorisme per detectar les discontinuïtats en la profunditat d'una escena. Propaguen la informació entre línies utilitzant una mesura de fiabilitat del valor de la disparitat en cada punt determinat pels punts veïns. Els punts es classifiquen en tres categories de confiança, i segons aquesta classificació es propaguen els valors de la disparitat. Aquest mètode permet gestionar regions de textura uniforme.

Els avantatges d'utilitzar aquest algorisme és que és bastant ràpid, a causa de la descomposició en subtasques que realitza la programació dinàmica i és un mètode senzill d'implementar. Entre els inconvenients és que no calcula molt bé la disparitat en objectes petits.

4.2.6. Anàlisi de disparitat i obtenció de la distància

Segons *Marr i Poggio* [1], hi ha tres etapes en el procés de recuperació de l'estructura d'una escena. Primer, seleccionar un punt característic d'un objecte en una de les imatges, segon, trobar el mateix punt característic en l'altra imatge complementària, i tercer, mesurar la diferència relativa entre la posició d'aquests dos punts. Això és el que es coneix com el problema de correspondència, i és fonamental per calcular la disparitat. Si trobem un punt en la imatge esquerra, ha d'aparèixer en algun lloc de la imatge dreta, la diferència de la posició d'aquests punts en cadascuna de les imatges és tal i com hem dit, la disparitat.

Una forma d'estimar la profunditat de cada un dels punts en l'escena és mitjançant el càlcul d'aquesta disparitat entre les dues imatges. Assumirem que l'escena és estàtica, és a dir, que els objectes visibles de l'escena no es mouen ni es deformen. Per definir la disparitat assumim una configuració de dues càmeres de característiques similars, com la que es mostra a la figura 29. Aquestes dues càmeres, per tant, formen un parell estèreo amb eixos òptics paral·lels Z_I i Z_D. Suposem que les càmeres tenen la mateixa distància focal, *f*, amb centres O_I i O_D separats una distància o línia base, *B*, de manera que les imatges que es formen, *I* i *D*, estan en plans paral·lels.

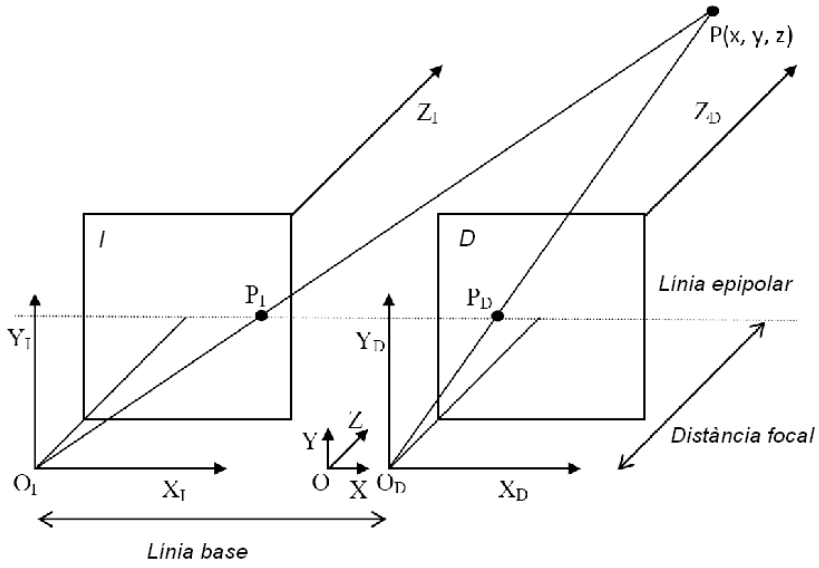


Figura 29. Configuració de càmeres paral·leles. Relació entre els paràmetres per a obtenir la profunditat, Z.

D'aquesta manera la línia base és paral·lela a l'eix *X*. Qualsevol punt en l'espai tridimensional *P*, amb coordenades (x,y,z), es projecta en cadascuna de les

imatges bidimensionals en els punts P_I i P_D, amb coordenades homogènies (x_I,y_I) i (x_D,y_D), respectivament.

El pla (PO_IO_D) creua a les imatges en dues rectes epipolars, e_I i e_D, per tant, un punt, P_I, a la recta e_I de la imatge I_I té el seu corresponent en algun punt de la recta e_D. Això redueix la recerca del corresponent de P_D de tota la imatge I_D a la recta e_D.

A la figura 29 podem veure com es relacionen els paràmetres definits en el parell estèreo que permeten obtenir la relació entre la disparitat d i la profunditat Z del punt P. La disparitat és la diferència en les coordenades horitzontals dels punts P_I i P_D, és a dir,

$$d = x_I - x_D \qquad (4.5.)$$

Per tant, les coordenades de P_I i P_D queden relacionades mitjançant,

$$x_I = x_D - d, \quad y_I = y_D \qquad (4.6.)$$

Per semblança entre els triangles s'arriba a la relació entre d i Z.

$$Z \cdot d = f \cdot B \qquad (4.7.)$$

D'aquesta manera es pot recuperar, excepte d'una constant d'escala, la profunditat de cada píxel en cadascuna de les imatges a partir de la disparitat calculada.

Els algorismes de càlcul de disparitat assumeixen, per simplicitat, que les imatges amb què es treballa estan rectificades, és a dir, que els plans de les imatges en cada càmera són paral·lels entre si, i paral·lels a la direcció en la qual hi ha el desplaçament entre les imatges. D'aquesta forma el corresponent d'un punt de la fila y_I de la imatge esquerra és a la fila $y_D = y_I$ de la imatge dreta.

Per trobar la correspondència de punts s'imposen restriccions basades en propietats físiques i geomètriques raonables dels objectes i superfícies presents en l'escena, i la seva relació. Les restriccions més habituals són:

a) Restricció epipolar: el corresponent d'un punt en una imatge ha d'estar a la recta epipolar del punt en l'altra imatge.

b) Restricció d'ordre: si la projecció de l'objecte Q està a l'esquerra de la projecció de l'objecte P a la imatge esquerra, llavors la projecció de Q ha d'estar a l'esquerra de la projecció de P en la imatge dreta.

c) Restricció d'unicitat: cada punt d'una imatge pot tenir només un corresponent en l'altra imatge.

d) Restricció de semblança: les característiques dels punts en les imatges (intensitat o color, etc.) no ha de canviar molt.

Un altre fenomen a tenir en compte en la visió estèreo i que influeix en el procés de trobar els punts corresponents, són les oclusions, o regions que es veuen en una de les imatges i que no es veuen en l'altra imatge per estar tapades per un objecte. Les oclusions sempre impliquen una discontinuïtat en la profunditat de l'escena i són l'origen d'errors en molts algorismes. Tot i això, aquestes regions ocultes poden ser usades per a la recuperació de l'estructura de l'escena i en donen informació fonamental.

L'objectiu d'un algoritme estèreo de càlcul de disparitat és obtenir la profunditat en tots els píxels de les imatges del parell estèreo. Es mesura la disparitat per a cada un dels punts de l'escena, obtenint una única imatge que s'anomena **mapa de disparitat**. Com que hi ha una relació directa entre la profunditat relativa dels objectes en una imatge i la seva disparitat en un parell estèreo, podem agafar com a valors relatius de la profunditat dels objectes la informació extreta del mapa de disparitat, és a dir, prendrem aquesta imatge com una aproximació vàlida del mapa de profunditat.

El mapa de disparitats es pot entendre com una imatge on les intensitats no representen lluminositat en l'escena, sinó més aviat la disparitat de cada punt en l'escena. Els punts de l'escena més propers a les càmeres tindran disparitats més grans que els punts allunyats. Per això, la majoria dels resultats es presentaran en mapes de disparitats, on els punts propers a les càmeres (punts amb disparitats altes) es representaran amb tonalitats de grisos clars i els llunyans amb intensitats de grisos foscos. El blanc correspon a la disparitat màxima de l'interval de recerca de disparitats i el negre correspon al mínim.

La nostra reconstrucció, ja que partim d'imatges que són 2D, prendrà com a valor aproximat de la profunditat relativa d'un punt respecte de l'observador, el corresponent a aquest punt al mapa de profunditat, i així poder obtenir una tercera coordenada per poder representar en un espai 3D.

Per reconstruir una escena en 3D, cal com a mínim tenir dues imatges 2D per obtenir característiques que no són observables des d'una imatge plana. La principal diferència entre una imatge 2D i una imatge 3D és la sensació de profunditat, és a dir, la distància a l'observador (en el nostre cas a la càmera) de cada un dels elements de l'escena. Cada element té una profunditat determinada, de manera que un dels problemes complexos és donar-li a la imatge plana aquesta distància, és a dir, obtenir el seu mapa de profunditat.

Les característiques i limitacions de cada procés d'obtenció de mapes de profunditat d'una escena porten a la necessitat d'integració, no només de mesures, sinó també de tècniques que facilitin la fusió estereoscòpica. Els estudis sobre la visió humana permeten observar aquest procés d'integració. Per exemple, la fusió binocular humana no es produeix si la disparitat és

superior al límit de Panum, el que suposa que el moviment ocular sigui important en la fusió estereoscòpica.

4.3. Reconstrucció

El procediment de reconstrucció 3D consisteix a calcular la ubicació *(x,y,z)* de cada punt contingut en les dues d'imatges aplicant la tècnica de visió estereoscòpica. Els fonaments es basen en l'operació que realitza el sistema visual humà, en el qual a partir de dos dispositius d'adquisició de dades (globus oculars) adquireixen alhora dues imatges d'una mateixa escena, les quals presenten un desplaçament entre si (disparitat). Aquesta disparitat és calculada pel cervell i utilitzada per a obtenir la profunditat mitjançant triangulació.

El model visual humà posseeix un sistema d'eixos convergents, és a dir, els globus oculars s'orienten en direcció al punt observat, el càlcul de la ubicació d'un punt en aquest tipus de sistema és molt complex, i per tant, computacionalment molt costós, és per això que habitualment s'utilitza un sistema d'eixos de coordenades paral·lels amb dues fonts receptores d'imatges disposades en forma paral·lela i des de les quals s'obtenen dues imatges, l'única diferència és la disparitat existent entre elles.

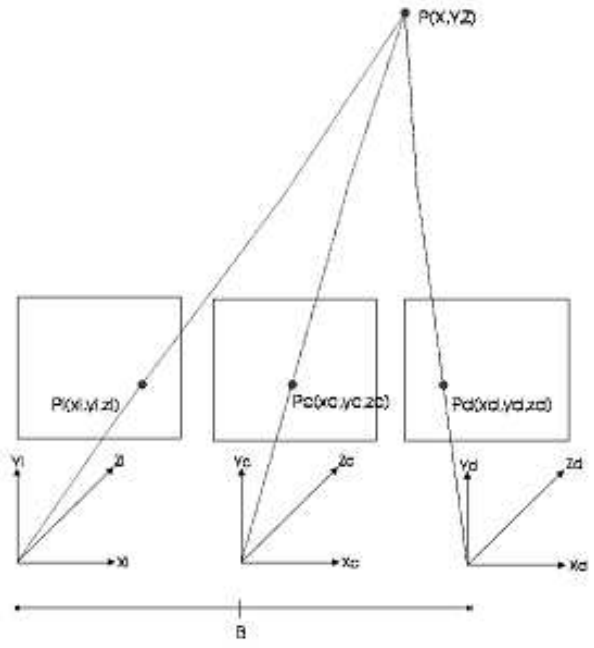


Figura 30. Sistema de coordenades ciclopè

Tal i com s'observa a la figura 30, partint de l'existència de dos sistemes de coordenades paral·lels (un per cada càmera) s'estableix un tercer sistema

situat simètricament entre els dos anteriors anomenat sistema ciclopè (per la seva ubicació similar a la d'un ull de ciclop).

Recolzats en els sistemes de coordenades anteriors i les fórmules de projecció en perspectiva es dedueix:

$$X = B\frac{(x_D + x_I)}{z(x_D - x_I)} \qquad Y = B\frac{y}{x_D - x_I} \qquad Z = B\frac{f}{x_D - x_I} \tag{4.8.}$$

De l'aplicació d'aquestes fórmules i considerant que es coneix la disparitat *(x_d - x_i)* calculada per l'algoritme de correspondència, es pot recuperar la ubicació real *(x,y,z)* de cada punt.

En el procés de reconstrucció 3D, un cop adquirides les imatges, obtingudes les característiques i la cerca de correspondències, és necessari conèixer el modelatge de les càmeres per poder determinar les profunditats dels punts de l'escena. Un model de les càmeres és una representació de les característiques geomètriques i físiques del sistema d'adquisició estèreo. Aquest model té un component relatiu, que relaciona les coordenades de les dues càmeres i que és independent de l'escena, i un component absolut que relaciona el sistema de coordenades d'una càmera amb el sistema de coordenades de l'escena.

El procés d'obtenció d'aquests components és el **calibratge**. L'etapa de calibratge és crucial en el procés de visió estèreo ja que permet simplificar el problema de correspondències i obtenir una representació tridimensional precisa dels resultats.

El model estereoscòpic és el conjunt d'expressions que relacionen punts en l'escena, expressats en coordenades globals, amb els punts projectats expressats en les coordenades locals de cada càmera per a una certa disposició geomètrica dels sensors òptics. En observar una escena des de dos punts de vista diferents, s'obtenen imatges en les quals els objectes es veuen més o menys desplaçats segons la seva profunditat i la seva posició en l'escena. La Figura 30 és un exemple de tres imatges preses amb càmeres paral·leles desplaçades lateralment al llarg de l'eix x. El sistema visual humà té una configuració semblant, però a més té la capacitat d'ajustar l'angle de convergència (angle entre els eixos òptics), que en el cas paral·lel és fix i igual a 0º.

5. Resultats

Tal i com ja hem comentat anteriorment, l'objectiu d'aquest treball és desenvolupar, mitjançant el programa MATLAB, un mètode de càlcul de la profunditat d'una escena de baixa complexitat computacional. Això ho farem a partir del càlcul de la distància entre els diversos objectes i la càmera. Aquesta distància l'obtindrem a partir de les disparitats produïdes per la visió binocular.

Per poder realitzar l'estudi, disposem de diverses imatges d'una escena obtingudes amb la càmera Bumblebee2, de Point Grey. D'altra banda, també disposem d'algunes mesures de l'escena que ens seran útils per a poder calibrar les mesures que anirem fent.

A l'escena de la figura 31 hi podem veure una taula amb els seus complements al costat esquerre i en primer pla, també en un pla molt proper hi ha un moble al costat dret, i en un segon pla, a la part central, hi ha una cadira i un cendrer. Pel que fa a la part superior, a l'esquerra ha dues finestres i al sostre hi ha un focus de llum. Veurem més endavant que aquest focus juntament amb els seus reflexes a les finestres i a la paret de fons ens provocarà alguns problemes de detecció.

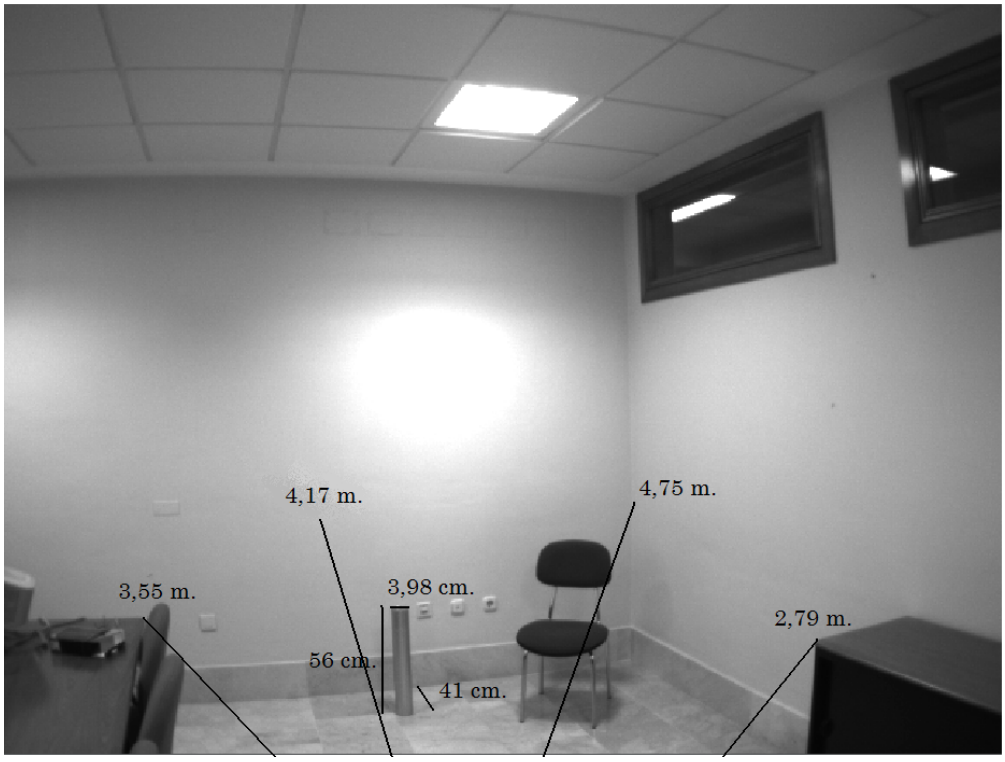


Figura 31. Imatge per al calibratge

5.1. Càrrega de les imatges

Les imatges que tenim són en format pgm. El format pgm és format d'arxiu representat amb escala de grisos. Està dissenyat per tal de poder treballar amb ell de forma molt senzilla. Una imatge pgm es pot considerar una simple matriu d'enters de valors arbitraris que es poden processar amb qualsevol programa. El nom 'PGM' és un acrònim derivat de 'Portable Gray Map'.

En aquest cas, al tractar-se d'una imatge estereoscòpica conté informació de les dues càmeres, per tant, està formada d'una combinació de les imatges de la càmera dreta i de l'esquerra. El primer que hem hagut de fer és separar les dues imatges esquerra (*I*) i dreta(*D*) a partir de la imatge inicial. En el nostre cas, la imatge inicial conté les dues imatges amb les files intercalades, la imatge *D* es troba a les files imparelles i la imatge *I* a les parelles. Això ho realitzem amb la funció *[I,D]=f_separaID('imatge').* A la figura 32 podem observar les petites diferències de desplaçament entre les imatges esquerra i dreta.

Figura 32. Imatges esquerra i dreta

5.2. Obtenció de màxims i mínims de la funció suma.

Un cop disposem de les dues imatges hem d'implementar el mètode de baixa complexitat per tal de calcular les disparitats dels objectes i així poder calcular la seva profunditat.

El processat de les imatges el farem dividint les imatges en parts o blocs més petits amb els quals podrem fer un processat més ràpid. Aquests blocs tindran una forma rectangular i tots ells tindran les mateixes dimensions. Com que les

imatges són de 640x480 píxels inicialment ho dividirem en 16x16 parts, d'aquesta manera cada bloc tindrà les dimensions de 40x30 píxels. Tot i això, aquest serà un paràmetre que podrem canviar fàcilment si volem treballar amb divisions més o menys grans, o bé, volem treballar amb imatges d'altres formats, l'única condició que hem de tenir en compte és que el nombre de divisions que realitzem sigui enter. Per a cadascuna d'aquestes divisions buscarem les possibles disparitats.

Tal i com ja hem comentat, el treball es basa en fer un processament de les imatges unidimensional perquè la quantitat de càlculs es redueixi potencialment.

Les imatges les convertim en escala de grisos de 8 bits, per tant, en una escala que va des del color negre (valor 0) fins al blanc (255). Per cadascun dels blocs obtindrem una funció suma que ens indicarà la quantitat de blanc que hi ha en cada columna i farem el mateix per cadascuna de les files que formen el bloc. Per tant, en el cas de dividir la imatge en 16x16 blocs seria una funció de 40 valors resultat del càlcul horitzontal i una funció de 30 valors del càlcul vertical per a cada bloc. Això ho farem tant per la imatge dreta (*D*) com l'esquerra (*I*). El resultat serà que per al conjunt dels 16 blocs que formen una fila, obtenim una funció com la de la figura 33.

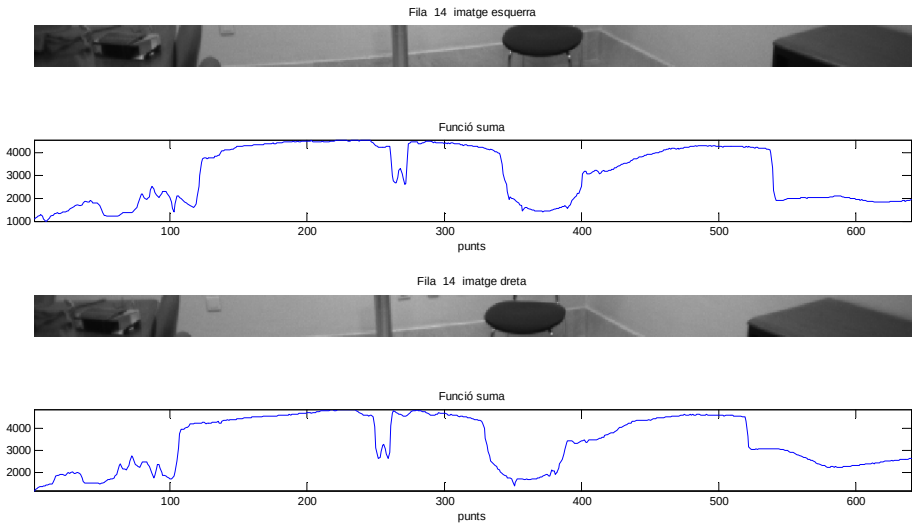


Figura 33. Fila de les imatges esquerra i dreta amb la corresponent funció suma

Tal i com hem comentat en el punt 3, s'observen uns salts importants en les vores dels objectes que hi ha a l'escena que ens han de permetre obtenir el mapa de disparitat de l'escena. Observem que com que la paret és blanca, en aquests punts obtenim un valor elevat a les funcions i on hi ha objectes que són de color fosc la funció disminueix de forma més o menys brusca.

En vertical, a la figura 34, fem el mateix,

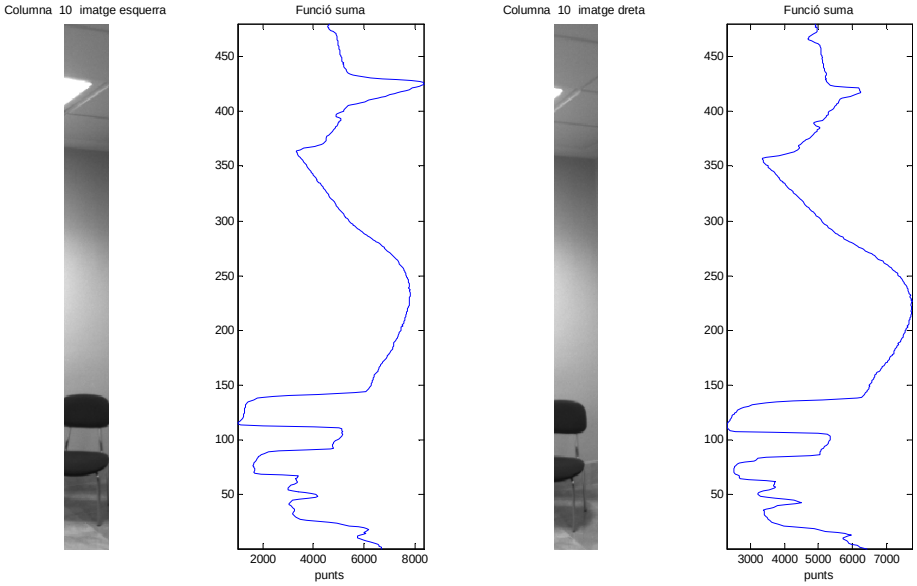


Figura 34. Columna de les imatges esquerra i dreta amb la corresponent funció suma

Un cop tenim les funcions suma horitzontal i vertical de cadascun dels blocs, el següent pas serà obtenir els punts significatius de la funció suma de les dues imatges per poder-los comparar i així obtenir les característiques de la imatge i detectar on apareix cada objecte.

Fixem-nos en la figura 34. En la posició que apareix el focus de llum hi ha un màxim de la funció amb un pendent abrupte, o bé, observem que on hi ha la cadira es produeixen dos mínims importants en les dues funcions. Per tant, podem concloure que és important conèixer els punts on hi ha màxims i mínims de la funció suma.

Per cadascun dels blocs amb els que hem dividit la imatge calculem el valor màxim i el valor mínim de la funció suma, tant per la imatge *D* com per la *I*. Com que els extrems de cada tram poden ser màxims o mínims de la funció en aquell tram degut a que la funció és creixent o decreixent, els descartem juntament amb els seus adjacents. D'aquesta manera obtindrem per a cada bloc un màxim i/o un mínim, o cap dels dos.

El resultat obtingut pel conjunt de blocs que formen una filera és un conjunt de punts on hi ha màxims i mínims parcials com el que tenim a la figura 35. Els màxims locals els hem marcat amb una creu i els mínims locals amb un cercle.

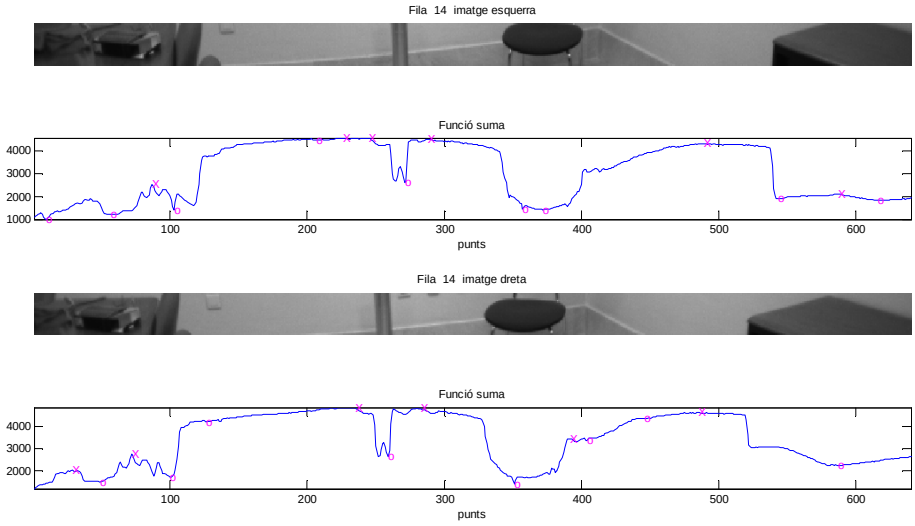


Figura 35. Funcions suma amb els màxims (x) i mínims (o) parcials

Aquest és un punt crític de l'estudi, ja que tot i calcular els màxims i mínims de diferents formes i variant els paràmetres de càlcul, quan els resultats milloren per uns casos, empitjoren en d'altres.

En el nostre cas el problema radica que per detectar un màxim o un mínim en una zona de grans variacions, els blocs haurien de ser petits perquè no ens quedin emmascarats màxims o mínims secundaris, i en canvi, en zones de variació lenta ens caldria utilitzar uns blocs més grans per evitar detectar punts amb una petita variació del pendent.

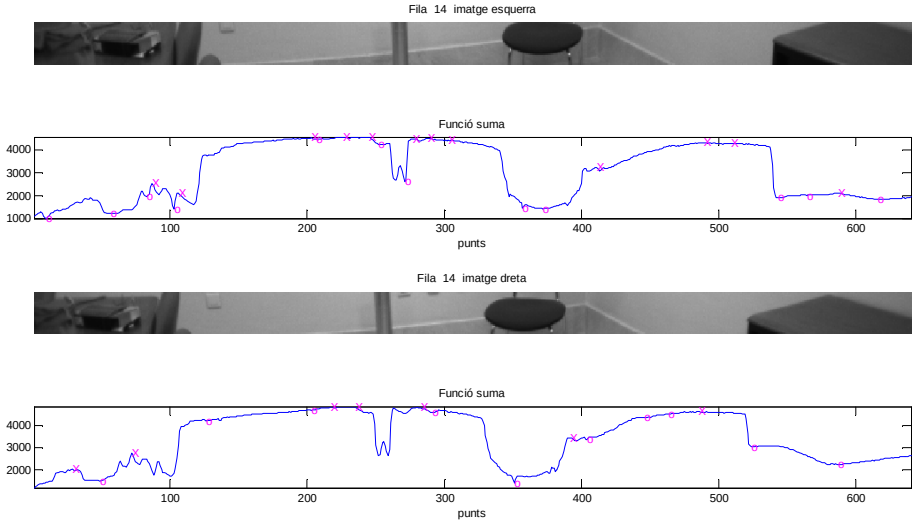


Figura 36. Funcions suma amb els màxims (x) i mínims (o) parcials per intervals petits

Per exemple (figura 36), el nombre de màxims i mínims en el cas de calcular-los per blocs la meitat de grans que en el cas anterior creix però si observem detalladament, tampoc podem assegurar que detecta tots els punts significatius, el que si que es pot observar és que genera molts punts no vàlids.

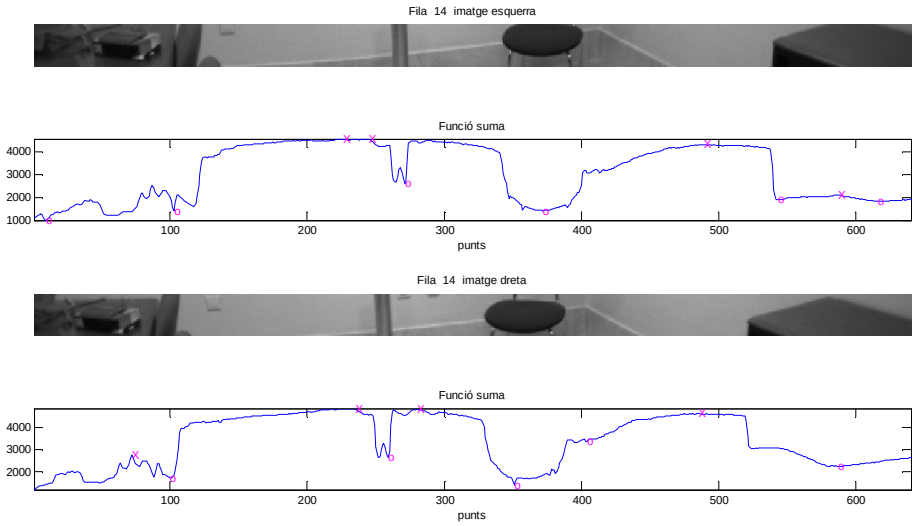


Figura 37. Funcions suma amb els màxims (x) i mínims (o) parcials per intervals grans

Si pel contrari, fem el càlcul per intervals el doble de grans (formats per dos blocs consecutius) el nombre de màxims i de mínims obtinguts es redueix però no podem detectar tots els punts desitjats (figura 37).

La detecció de màxims es faria de la mateixa manera per tots els blocs verticalment. L'exemple per tot un conjunt de blocs que formen una columna és el que tenim a la figura 38.

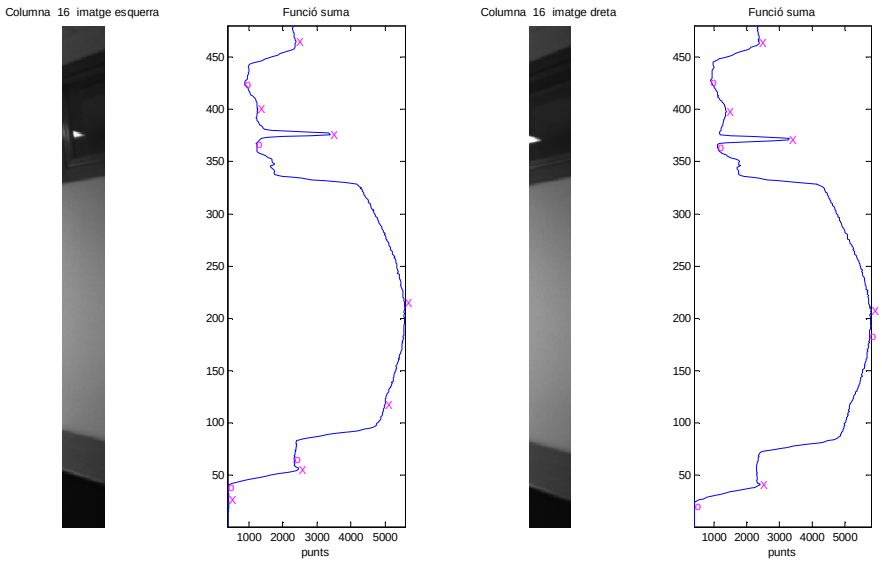


Figura 38. Funcions suma amb els màxims (x) i mínims (o) d'una columna

Tot i que, com es pot observar, els resultats obtinguts són força correctes, ens sorgeix el problema de la correspondència. Quan, per exemple, obtenim un màxim en una de les funcions suma no podem assegurar que trobem el corresponent a l'altra funció, ja que, és possible que l'haguem detectat però, tot i que visualment és molt senzill de certificar a quin màxim correspon, automàticament no sabrà a quin dels màxims detectats es correspon, fins i tot, pot ser que en l'altra imatge o no l'haguem detectat, o bé, que fins i tot no hi sigui.

5.3. Càlcul de la distància.

Un altre mètode d'anàlisi de les imatges pel càlcul de les disparitats és a través del càlcul de la distància entre les funcions obtingudes de les dues imatges. Com que un dels objectius del nostre treball és que l'anàlisi sigui el més simple possible hem calculat la distància restant directament punt a punt la funció suma de les dues imatges *I* i *D*. Aquest és un càlcul molt simple que comporta molt poca càrrega computacional i el realitzarem mitjançant la funció *[tmax]=f_DistaMax(SHI,SHD,i,Wx,Xm).*

$$s(n) = s_I(n) - s_D(n) \qquad (5.1.)$$

Abans de poder fer una comparació dels senyals per calcular la distància per cadascun dels blocs, eliminem el component continu i posteriorment normalitzem els senyals suma. De manera que per un senyal *s(n)* de *N* punts tenim:

$$s_m(n) = s(n) - m_S \qquad (5.2.)$$

On m_s és la mitjana

$$m_S = \frac{\sum_{n=1}^{N} s(n)}{N} \qquad (5.3.)$$

Per normalitzar

$$s_n(n) = \frac{s_m(n)}{P_S} \qquad (5.4.)$$

On P_s és la potència mitjana

$$P_S = \frac{\sum_{n=1}^{N} s_m^2(n)}{N} \qquad (5.5.)$$

Un cop normalitzats els blocs de cada part de la imatge passem a calcular la distància per cadascun d'ells com a diferència entre els senyals enfinestrats al quadrat, per tal d'emfatitzar més les diferències i obtenir valors positius.

$$d(n) = (s_I(n) - s_D(n))^2 \qquad (5.6.)$$

En el cas que en algun dels blocs hi hagi diferències significatives de color (valor de gris) aquesta distància tindrà un pic acusat com en la figura 39.

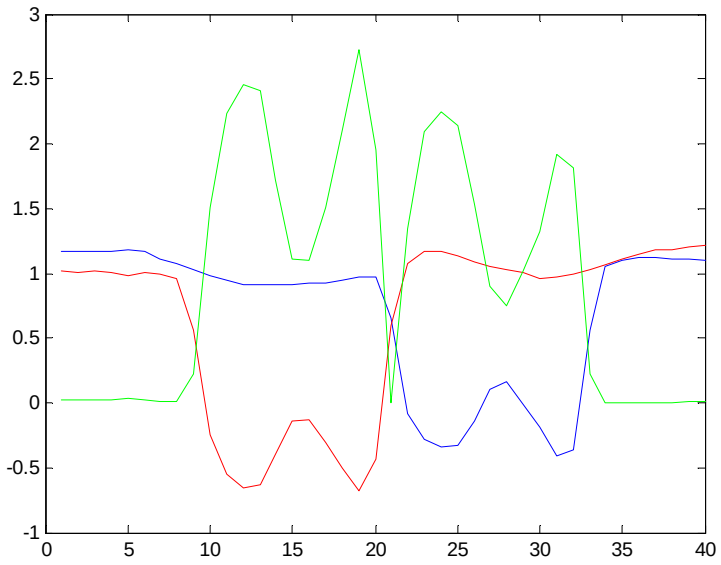


Figura 39. Funcions suma d'un bloc amb un objecte i la distància

Observem en blau la funció corresponent a la imatge esquerra (I) i en vermell la de la imatge dreta (D). En color verd observem la distància amb uns pics que indiquen que en aquest bloc hi ha una notable disparitat. Aquesta imatge correspon a la part del cendrer del centre de l'escena.

En cas de que no hi hagi gaire diferència no apareixerà cap pic com s'observa a la figura 40, que no hi ha massa diferències entre les dues funcions. Aquest gràfic correspon a la part entre el cendrer i la cadira.

En funció de la sensibilitat que ens interessi tenir agafarem només els màxims de valor molt elevat o màxims de valor més reduït. El valor que ens ha donat més bons resultats estaria entorn de 0,7.

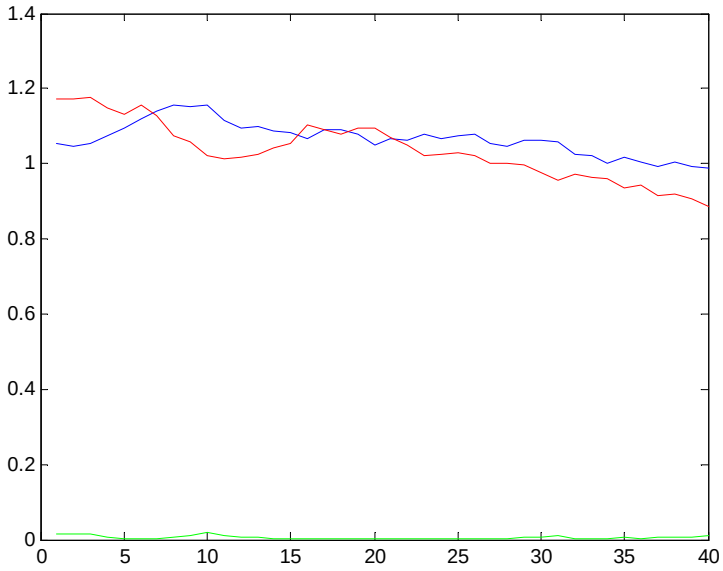


Figura 40. Funcions suma d'un bloc sense cap objecte i la seva correlació

Observem que en el cas d'utilitzar alguns dels resultats obtinguts. Fixant-nos en les figures 41 i 42 observem que si utilitzem blocs més petits els objectes queden més ben definits, observem que el perfil de la persona és força visible, però per contra apareixen zones fosques que ens fusionen alguns dels objectes.

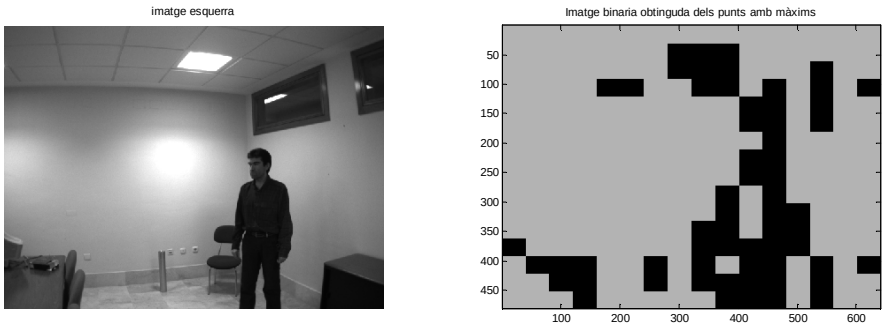


Figura 41. Imatge esquerra i mapa de màxims i mínims per bloc gran amb llindar 0,7

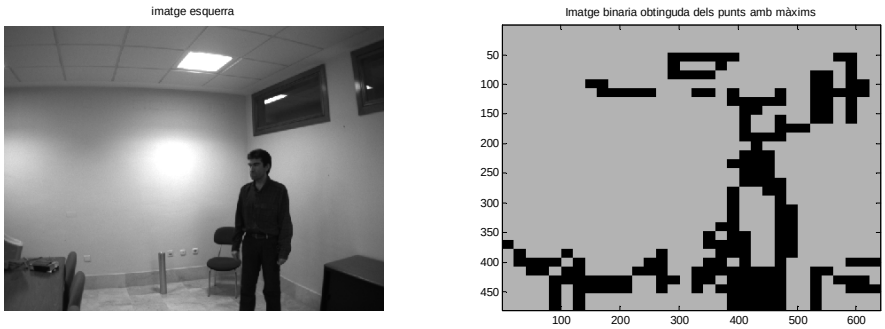


Figura 42. Imatge esquerra i mapa de màxims i mínims per bloc petit amb llindar 0,7

Per les figures 43 i 44 observem un cas similar amb els mateixos resultats.

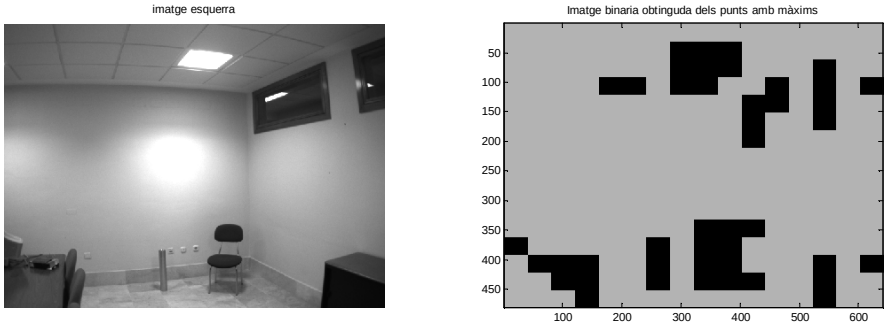


Figura 43. Imatge esquerra i mapa de màxims i mínims per bloc gran amb llindar 0,7

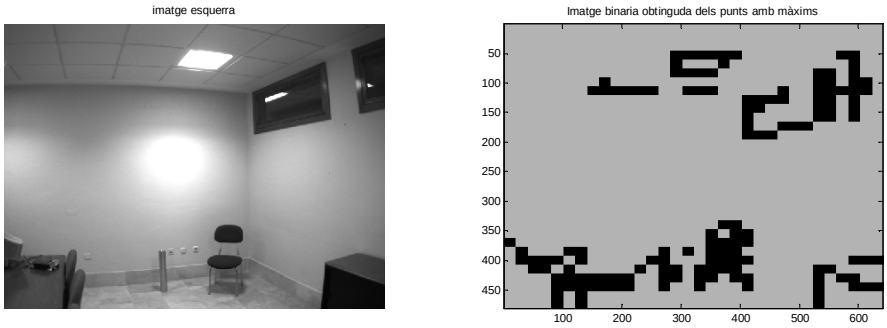


Figura 44. Imatge esquerra i mapa de màxims i mínims per bloc petit amb llindar 0,7

5.4. Càlcul de la correlació.

Mitjançant aquest simple càlcul de la distància de l'apartat anterior detectem els blocs on hi ha disparitat, per tant, els punts on les imatges dreta i esquerra no coincideixen en el to de gris i que corresponen en punts on hi ha objectes. De totes maneres aquest mètode no ens serveix per a detectar la profunditat dels objectes, per això utilitzarem el concepte de correlació.

Un problema que tenim a l'hora de comparar les funcions suma d'un bloc són els seus extrems. Per tal d'evitar l'efecte dels extrems, un cop normalitzats els senyals realitzarem un procés d'enfinestrat i així evitarem les discontinuïtats del inici i del final del bloc analitzat i donarem èmfasi a la part central del senyal que estem comparant. Per aquest efecte hem d'utilitzar una finestra amb un màxim central i amb els extrems nuls. La finestra de Hanning compleix aquest requisit. L'equació que la defineix és:

$$w(n) = \frac{1}{2}\left(1 - \cos\left(2\pi \frac{n}{N}\right)\right), \quad 0 \le n \le N \qquad (5.7.)$$

Pel càlcul de la distància enfinestrarem la funció tenint en compte el bloc i els seus blocs adjacents, per tant, considerarem una longitud equivalent a tres blocs i amb la finestra donarem èmfasi als elements del bloc central. L'amplada del lòbul de la finestra ens marcarà la resolució del sistema. Una finestra estreta ens permetrà diferenciar punts significatius més propers i una finestra ampla ens detectarà punts significatius més llunyans.

Un cop enfinestrats els blocs de cada part de la imatge passem a calcular la distància per cadascun d'ells. Una forma de calcular la distància entre dos funcions és mitjançant l'ús de la funció de correlació. Podem definir la distància com:

$$d(\tau) = E_{SI} + E_{SD} - 2R_{ID}(\tau) \qquad (5.8.)$$

Podem observar que la distància dependrà de la correlació. El resultat més interessant d'aquest càlcul no és el valor del màxim de la funció distància, sinó, la seva posició. Per això calcularem la correlació entre les dues funcions i utilitzarem la posició del seu màxim pel càlcul de la disparitat. La funció de correlació que hem desenvolupat es la *[tau]=f_CorrMax(SHI,SHD,i,Wx,Xm)* amb petites variants per si es tracta de la correlació en vertical o horitzontal.

Observem la figura 45 que correspon a les dues funcions suma obtingudes en el bloc on hi ha el cendrer. A la part inferior observem el resultat per la imatge esquerra i la imatge dreta. En ambdós casos apareixen a la part central dos mínims corresponents a les vores del cendrer. Observem el desplaçament entre una i altra funció. Una vegada obtinguda la correlació (gràfic superior), aquesta tindrà un màxim desplaçat cap a un costat 11 punts, indicant-nos aquesta diferència entre els dos gràfics.

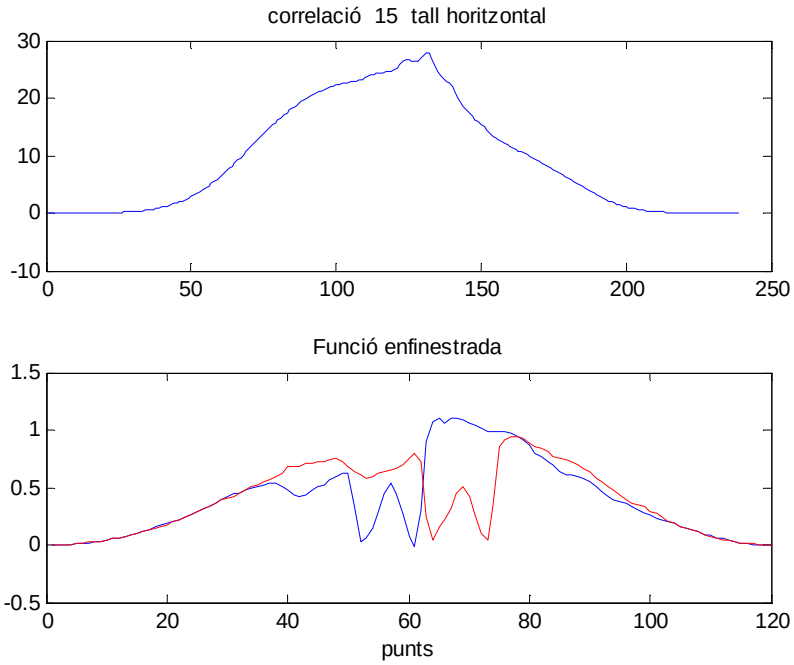


Figura 45. Funció suma d'un bloc enfinestrada i la correlació

Una cosa semblant la podem observar en la part de la imatge on hi ha la cadira en els següents gràfics que corresponen a la part del seient (figura 46) i a la part superior del respatller (figura 47). La correlació també té un desplaçament d'11 punts cap a la dreta.

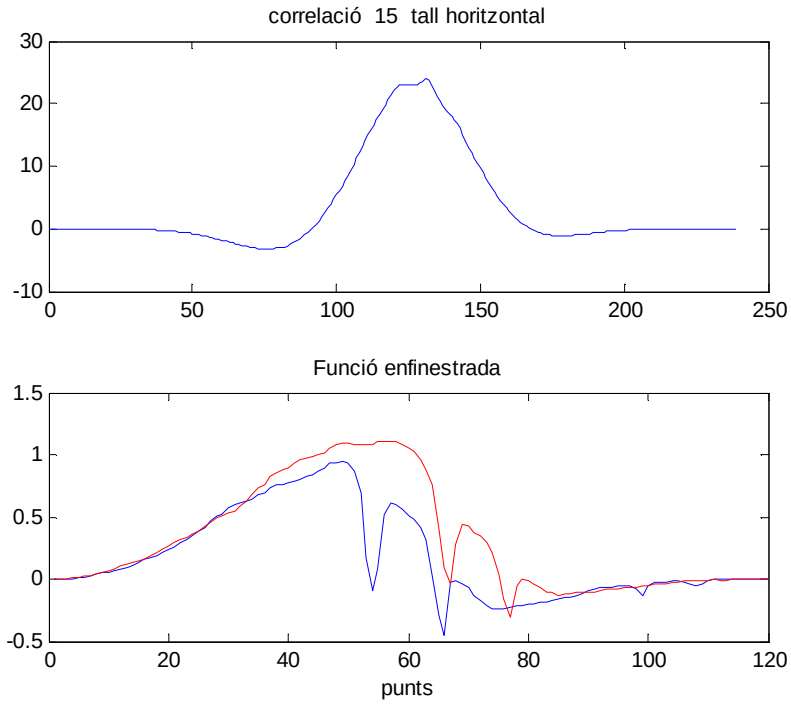


Figura 46. Funció suma d'un bloc enfinestrada i la correlació

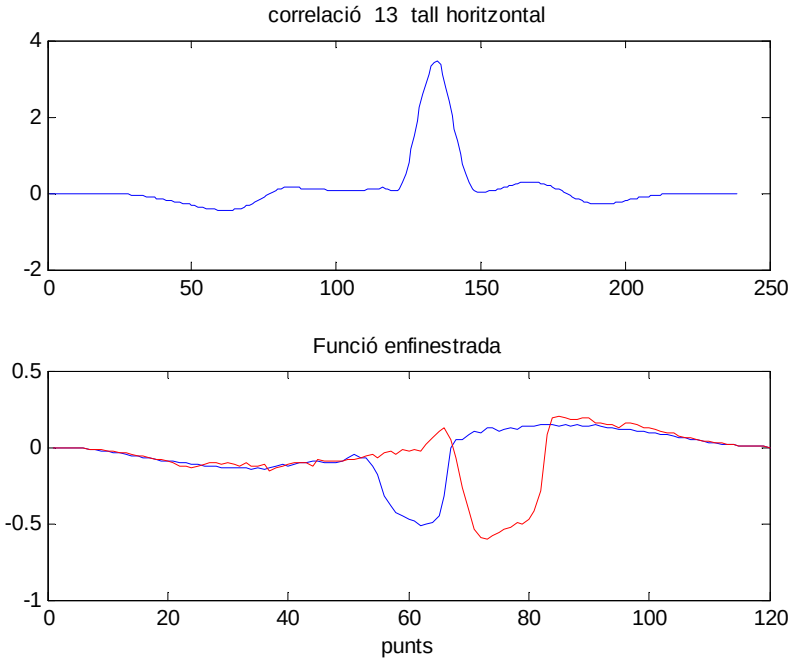


Figura 47. Funció suma d'un bloc enfinestrada i la correlació

Com a exemple, també podem observar què és el que succeeix en una part de la imatge on no apareix cap objecte, com és la paret del fons de l'habitació. Observem a la figura 48 que la correlació entre les dues imatges és simètrica i totalment centrada. Obtenim un valor del desplaçament del màxim de la correlació igual a 0.

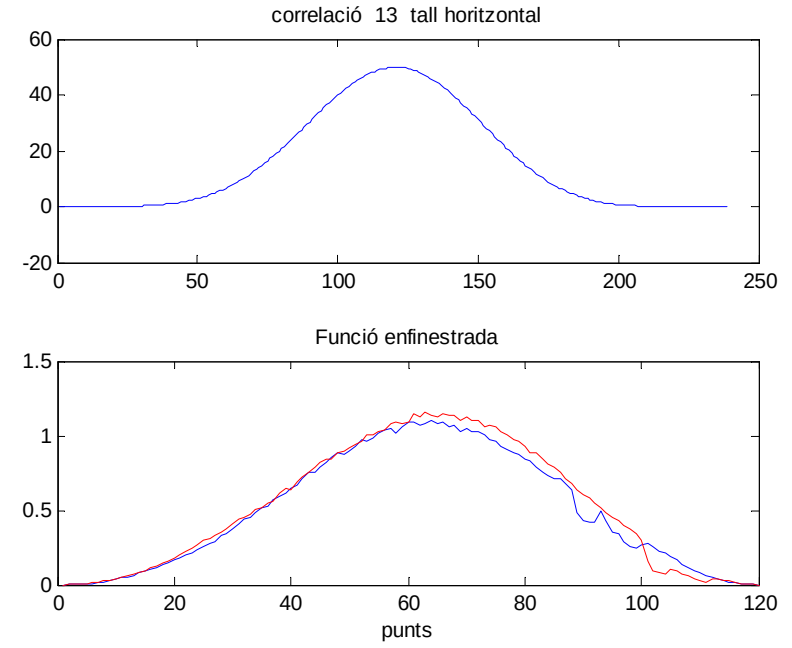


Figura 48. Funció suma d'un bloc enfinestrada i la correlació

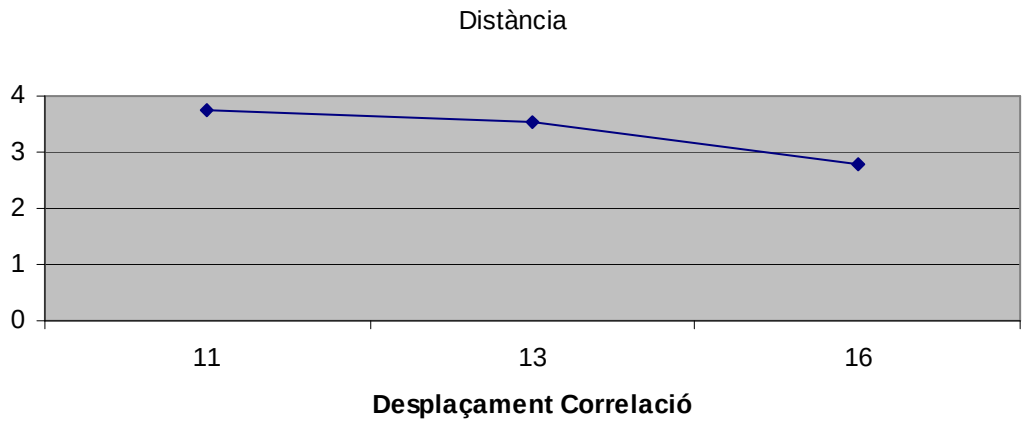
Després de calcular les correlacions entre els senyals suma dels diferents blocs en que hem dividit les imatges, farem el **calibratge** del nostre sistema.

Per tal de poder calibrar el nostre sistema agafarem les distàncies conegudes entre la càmera i els objectes, i les compararem amb els resultats de desplaçament de la correlació en aquests punts.

Observem a la figura 49 que podem trobar una relació de quasi proporcionalitat entre la distància dels objectes a la càmera, i el desplaçament del màxim de la funció de correlació.

Distància	Desplaçament
3,76	11
3,55	13
2,79	16

a)



b)

Figura 49. a) Taula de relació distàncies/Desplaçament, b) Representació gràfica

És important de comentar que les distàncies mesurades de l'escena de les que disposem són escasses i no contemplen variacions en l'alçada dels objectes respecte de la càmera. Tampoc hi ha mesures respecte cadascun dels dos objectius i suposarem que estan obtingudes aproximadament des del centre de les dues càmeres. Això fa que no puguem fer un càlcul exacte de la posició dels objectes, de totes maneres, aquest error de calibratge no ens condicionarà massa els resultats obtinguts, ja que pel guiatge del robot és important saber la distància aproximada dels objectes però el més important és la rapidesa.

Un cop tenim calculada la distància a la que es troba cada objecte cal representar-la gràficament. Ho farem representant-ho en un gràfic en una escala de grisos on, com més clars són els píxels de la imatge, més proper és l'objecte que hi ha en aquella posició. Per eliminar part del soroll de la imatge

indicarem amb el fons negre els blocs que la correlació té un desplaçament màxim de 2 píxels. D'aquest mètode obtindrem els resultats que mostrem a les següents figures. Observem unes imatges en escala de grisos en que es poden veure la posició aproximada dels objectes en colors més clars si es troben prop de la càmera i més foscos si estan allunyats. A l'igual que amb el càlcul dels màxims i dels mínims els resultats no són òptims però observem que amb els blocs més petits hi ha més definició dels objectes (figures 51 i 53) que amb els blocs més grans (figures 50 i 52), però que en tots casos apareix soroll en zones amb canvis de color degudes a la mala il·luminació.

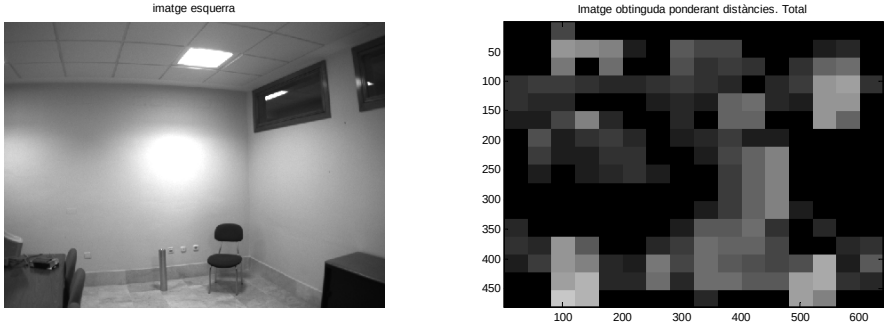


Figura 50. Imatge esquerra i mapa de correlació per bloc gran

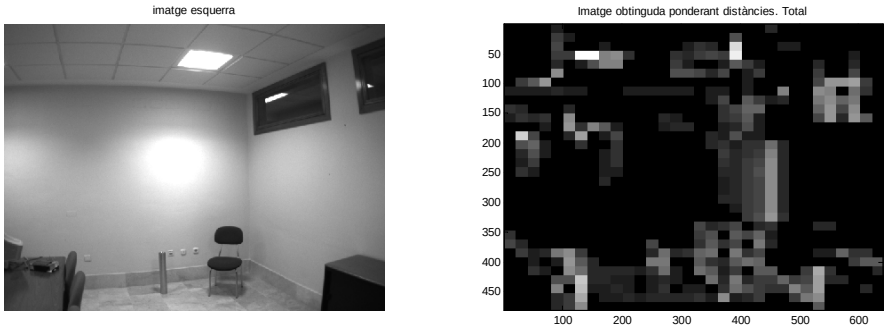


Figura 51. Imatge esquerra i mapa de correlació per bloc petit

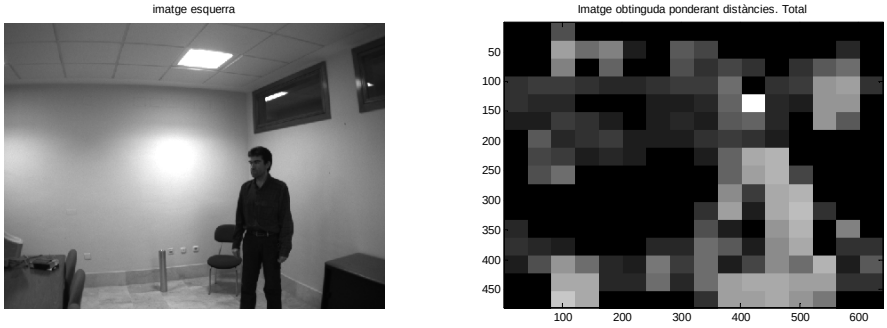


Figura 52. Imatge esquerra i mapa de correlació per bloc gran

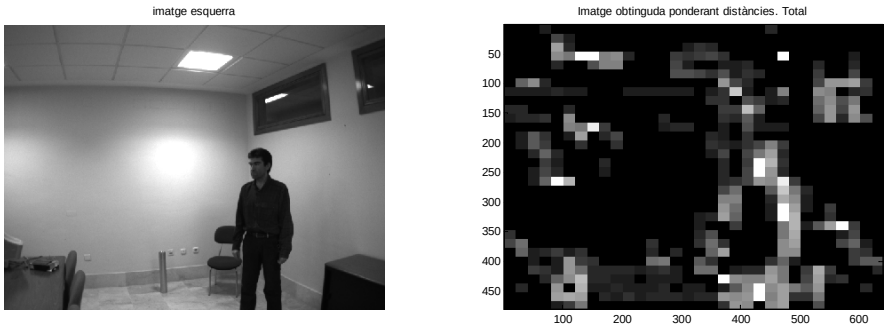


Figura 53. Imatge esquerra i mapa de correlació per bloc petit

5.5. Combinació dels mètodes anteriors

Amb el càlcul dels màxims i mínims no hem aconseguit obtenir les correspondències entre ells però hem aconseguit de localitzar-los amb certa exactitud. El càlcul de la correlació ens és molt útil per trobar la profunditat dels objectes però com hem observat hi ha zones on no hi ha objectes que probablement degut a la deficient il·luminació ens dona errors.

Si combinem els dos mètodes ens permetrà trobar la distància mitjançant la correlació, limitada en els blocs on hi ha un màxim o un mínim de cert nivell. Per tant, en les vores dels objectes calcularem la distància. Observem els resultats en algunes de les imatges (figures 54-56).

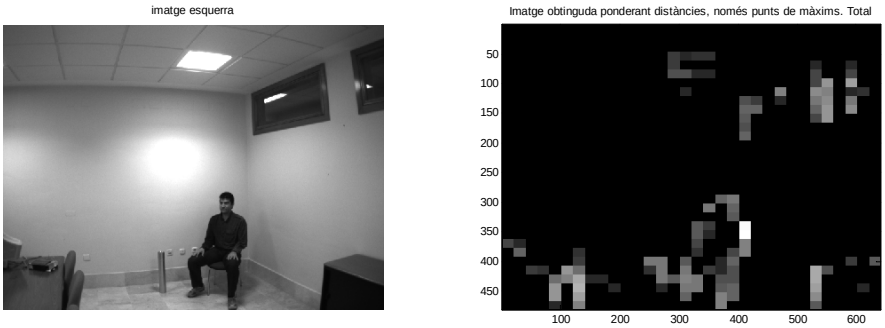


Figura 54. Imatge esquerra i mapa de correlació limitat

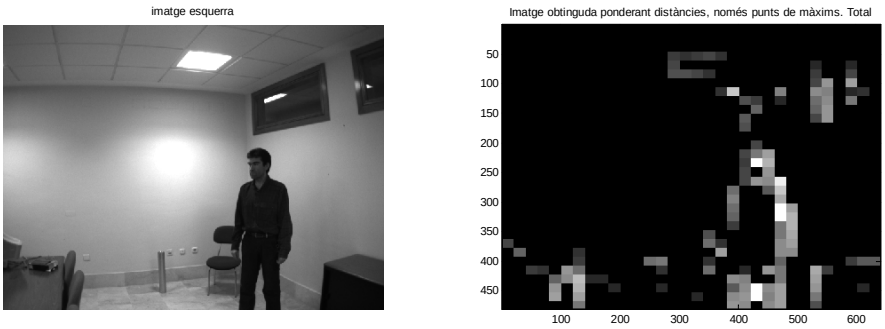


Figura 55. Imatge esquerra i mapa de correlació limitat

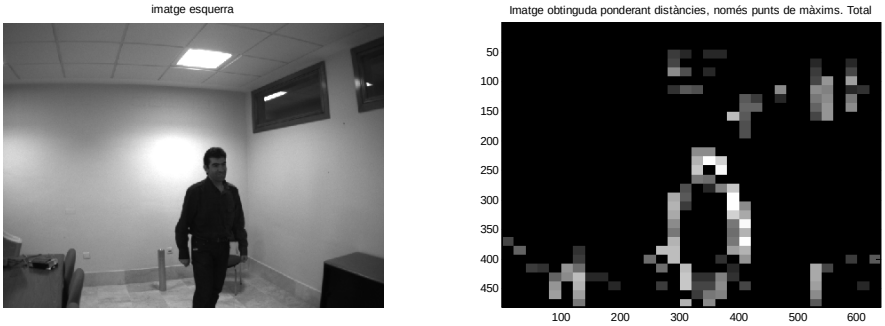


Figura 56. Imatge esquerra i mapa de correlació limitat

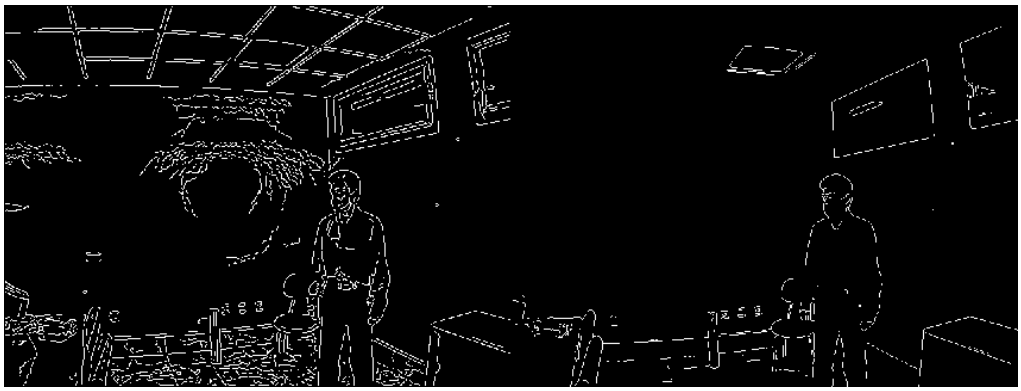
Observem que els resultats obtinguts són força correctes i a mesura que l'objecte (persona) s'apropa els seus contorns es van tornant de color més clar. També volem fer notar que d'aquesta manera també s'eliminen moltes zones sorolloses, com és el cas de la paret del fons. Aquesta notable millora es pot veure de comparar les imatges de les figures 53 i 55.

5.6. Detector de contorns.

Finalment hem desenvolupat un mètode de càlcul de la profunditat de l'escena a partir del càlcul dels contorns dels objectes. Tal i com hem explicat en l'apartat 4.2.3. hi ha bàsicament dos tipus de mètodes, els basats en gradient i els basats en creuaments per zero. Per tal d'observar-ne el comportament hem implementat en el nostre algoritme l'operador de Canny (creuaments per zero) i l'operador de Prewitt (basat en gradient).

El resultat és el que tenim a la figura 57. Observem que mitjançant Canny obtenim més detalls de l'escena que amb Prewitt. Pel nostre objectiu, que és de localitzar els objectes, l'algoritme de Canny ens detecta massa detalls, fins i tot detecta falses ombres a la paret del fons, per tant, no ens serveix. D'altra banda l'algoritme de Prewitt ens detecta exactament el que desitgem, els

contorns dels objectes, tot i que potser en algun cas no detecta els contorns sencers.



a) b)

Figura 57. Detector de contorns a) Operador de Canny, b) Operador de Prewitt

Així doncs implementarem l'algoritme de Prewitt. Per evitar les discontinuïtats utilitzarem l'operador morfològic de la dilatació, tal i com ja hem comentat a l'apartat 4.2.4. d'aquesta memòria. Aquest operador ens emfatitzarà els contorns i eliminarà les discontinuïtats que apareixien. En el nostre cas aplicarem un element estructurant quadrat, ja que l'anàlisi comparatiu que fem és per files i columnes, de tres píxels per costat, quedant com es pot veure a la figura 58.

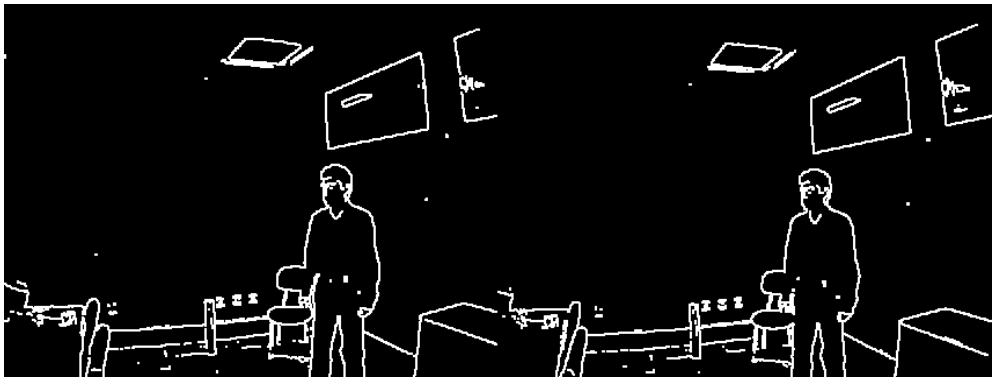


Figura 58. Detector de contorns de Prewitt amb una dilatació

Una vegada tenim la imatge formada pels contorns apliquem el mètode que hem descrit anteriorment. El resultat obtingut el mostrem a les figures 59 -61.

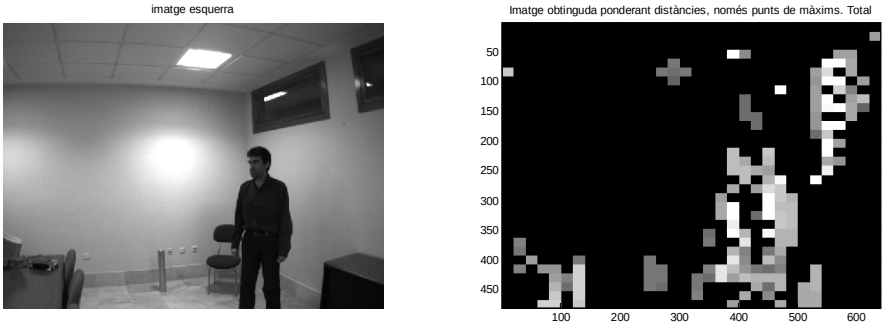


Figura 59. Imatge esquerra i mapa de correlació limitat aplicant Prewitt

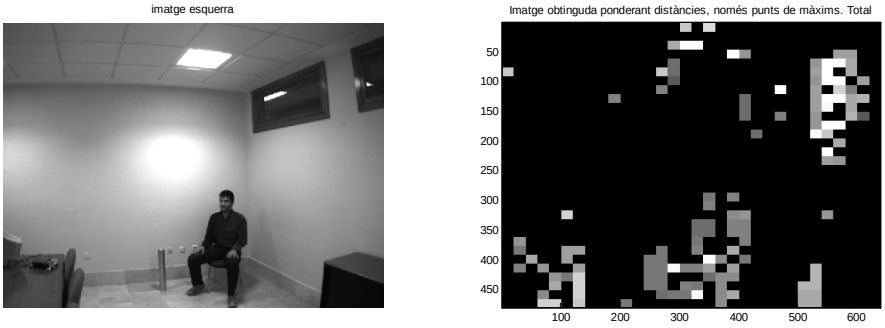


Figura 60. Imatge esquerra i mapa de correlació limitat aplicant Prewitt

Es pot observar que els resultats, tot i que amb més quantitat de càlculs, empitjoren els resultats obtinguts amb el mètode combinat.

6. Conclusions i línies de futur

En aquest treball hem aplicat un mètode totalment nou i sense cap relació amb els mètodes que s'utilitzen actualment per a calcular la profunditat dels objectes d'una escena a partir d'imatges estereoscòpiques. Un mètode amb una quantitat de càlculs molt baix respecte els mètodes més comuns.

S'ha aplicat el mètode que consisteix en l'obtenció de les funcions suma dels nivells de grisos dels píxels d'una imatge dividint-la en blocs petits. Alhora s'ha complementat amb diferents formes de detectar la profunditat a partir de les funcions suma obtingudes.

Mitjançant la combinació de la detecció de màxims i mínims, i de la correlació hem obtingut uns resultats força satisfactoris. Millors fins i tot que utilitzant un mètode més complex computacionalment com la detecció de contorns. Aquests resultats ens fan pensar és una bona base de partida per a nous experiments en aquest sentit. Observem que en general els objectes més grans es detecten amb una millor qualitat que els petits.

Una de les majors dificultats que hem tingut són les imatges de partida. La seva qualitat no era massa bona i alhora no disposàvem de la informació suficient de la posició dels objectes per poder calibrar el sistema. És possible que amb imatges de millor qualitat es puguin millorar els resultats.

Hi ha altres tècniques que es poden aplicar per intentar una millora dels resultats a partir de l'aplicació de tècniques com són el *Dynamic time warping (DTW)* o el filtratge morfològic. El *Dynamic time warping* [16] és un algoritme que ens permet mesurar la similitud entre dues seqüències que tenen variacions en el temps i en l'espai com és el nostre cas. El que fa és alinear les seqüències per tal de poder calcular la distància entre patrons. El filtratge morfològic és una tècnica basada en les propietats geomètriques que mitjançant operacions d'*opening* permet eliminar zones amb soroll i crestes a la funció suma i amb operacions de *closing* eliminar valls.

Una altra qüestió a tenir en compte seria tenir l'objectiu d'aplicació ben definit abans de capturar les imatges, ja que per exemple, si es tracta de poder fer el guiatge d'un robot que es desplaci pel terra només ens interessarà la imatge dels objectes que poden interferir amb ell, és a dir, la imatge dels objectes que estan al terra.

Una possible línia de treball podria ser la implementació en un microcontrolador perquè controli el moviment d'un dispositiu mòbil amb una càmera a la seva part superior. Hi ha altres aplicacions interessants, com indica [6] que consisteix en poder supervisar gent amb dependència dins d'una habitació (detectar moviment, posició) per controlar el seu estat de salut.

7. Bibliografia

[1] D. Marr, T. Poggio. *A computational theory of human stereo vision*. Proceedings of the Royal Society of London, 1979.

[2] F. Lecumberry. *Cálculo de disparidad en imágenes estéreo, una comparación*. III Workshop de Computación Gráfica, Imágenes y Visualización, Uruguay, 2005.

[3] C. Buehler, S.J. Gortler, M.F. Cohen, L. McMillan. *Minimal Surfaces for Stereo*. European Conference on Computer Vision, pag. 885-899, 2002.

[4] Á. Suarez. *Análisis de métodos de procesamiento de imágenes estereoscópicas forestales.* Projecte fi de màster en investigación en informática. Universidad Complutense de Madrid, 2009.

[5] S. Rodríguez, J.M. Corchado. *Stereo-MAS: Multi-Agent System for Image Stereo Processing.* Informe Técnico DPTOIA-IT, 2008.

[6] M.P. Rubio, J.M. Corchado. *Sistema de Inteligencia Ambiental mediante visión estereoscópica y representación 3D en tiempo real para la localización de personas en entornos de dependencia.* Informe Técnico DPTOIA-IT, 2008.

[7] S. Martín, J. Suárez, R. Rubio, R. Gallego. *Aplicación de los sistemas de visión estereoscópica en las enseñanzas técnicas.* Universidad de Oviedo, Departamento de Construcción e Ingeniería de Fabricación, 2004.

[8] Point Grey Research Inc. http://www.ptgrey.com/products/bumblebee2/bumblebee2_stereo_camera.asp

[9] P.E. Anuta. Digital *registration of multispectral video imagery*. Society of Photo-Optical. Instrumentation Engineers. Dallas, Texas 1970.

[10] *Technical Application Note TAN2008005. Stereo Vision Introduction and Applications*. Point Grey Research Inc., revisat el febrer 2010.

[11] C. Platero. *Apuntes de Visión Artificial*. Dpto. Electronica, Automatica e Informatica Industrial, UPM, 2009.

[12] R. Reig. *Procesado de datos multidimensionales. Módulo 4: Análisis y reconocimiento.* Master en tecnologías aplicadas de la Información, 2010.

[13] M. Ramos. *Sistema de pre-procesamiento de imágenes electrocardiográficas en telemedicina*. Tesis, capitulo 3, licenciatura Ingeniería en Sistemas Computacionales. Universidad de las Américas, Mèxico, 2003.

[14] P. Compañ, R. Satorre, C. Villagrá, R. Rizo. *Visión estereoscópica en un modelo multirresolución*. Dpto. Ciencia de la Computación e Inteligencia Artificial, Universidad de Alicante, 2001.

[15] S. Bircheld, C. Tomasi. *Depth Discontinuities by Pixel-to-Pixel Stereo.* International Conference on Computer Vision, 1999.

[16] C.A. Glasbey, K.V. Mardia. A *review of image-warping methods*. Journal of applied statistics, Vol 25, n. 2, 1998.