

UNIVERSITAT POLITÈCNICA DE CATALUNYA

Departament d'Enginyeria de Sistemes Automàtica i Informàtica Industrial

**APORTACIÓ A LA DESCRIPCIÓ I
SEGUIMENT DE CAMINS
NAVEGABLES EN ENTORNS
NATURALS A PARTIR DE L'ANÀLISI
DE REGIONS EN SEQÜÈNCIES
D'IMATGES**

Autor: Josep Fernández Ruzafa

Director: Alícia Casals Gelpí

Barcelona, febrer del 1998

5. Detecció d'Obstacles i Descripció de l'Entorn a partir de l'Anàlisi de la Seqüència d'Imatges

Conforme un vehicle autònom es desplaça, arriba a noves àrees de l'entorn que fins ara, normalment, li eren desconegudes. De totes maneres, fins i tot disposant d'un mapa previ, el sistema de navegació ha d'obtenir informació de l'entorn, mitjançant el seu sistema sensorial, per detectar possibles canvis en l'entorn i/o per determinar la posició actual del vehicle a l'entorn. Amb l'objectiu de garantir la integrabilitat del vehicle i dels elements presents a l'entorn, és necessari actualitzar la informació de l'estat actual de l'entorn de forma freqüent, d'aquesta manera el sistema de navegació disposarà de la informació sobre els canvis de l'entorn, i s'estarà en condicions de detectar possibles situacions anòmales, com per exemple que el mòbil no segueixi la direcció adequada o l'aparició d'un nou obstacle, en un petit interval de temps. Evidentment, com més informació ha de ser analitzada en un interval de temps fixat, més costos serà el sistema de processament necessari.

L'actualització freqüent de la informació de l'entorn, produirà un cert solapament en la informació captada pel sistema de navegació, és a dir, la mateixa àrea de l'entorn és observada i analitzada varies vegades, normalment des de punts diferents de l'entorn, donat que el vehicle es desplaça. Aquesta redundància d'informació serà utilitzada per desfer ambigüitats, detectar els errors que s'han produït en l'anàlisi d'una única imatge, i minimitzar els errors inherents a tot sistema de mesura.

En el quart capítol s'ha presentat la primera part de la metodologia per a la descripció de camins mal o feblement estructurats proposada, en la que l'anàlisi de l'entorn es basa en una única imatge, que és segmentada en regions i aquestes són classificades en funció del seu color i la seva posició a la imatge. En aquest capítol es presenta i discuteix com es pot utilitzar la seqüència d'imatges captades mentre que el vehicle es desplaça al llarg del camí, per detectar els possibles obstacles presents a l'interior del camí, i per verificar dinàmicament la correctesa i coherència de la descripció de l'entorn generada, amb l'objectiu de proporcionar la informació adequada al planificador per navegar de forma segura pel camí.

5.1. *Detecció d'Obstacles*

Tot sistema de navegació autònoma que ha de desplaçar-se en un tipus d'espai predefinit, ha de resoldre, entre d'altres, dos problemes bàsics: la *localització i seguiment de la via* (autopista, carretera, carrer, camí, passadís d'un edifici, corredors d'una planta industrial, etc.) per la qual el mòbil ha de desplaçar-se, i la *detecció i evitació d'obstacles* que es presentin a l'interior d'aquestes vies (com ara altres vehicles, persones o objectes propis de l'entorn industrial o natural). En el nostre cas, considerem obstacle tot objecte que es trobi a l'interior del camí, i que sobresurt del terra en una alçada tal, que fa impossible o desaconsellable que el mòbil travessi aquella zona. Quan es transita per una carretera o camí ens podem trobar amb dos tipus d'obstacles: els artificials i els naturals.

Un obstacle artificial és un objecte que es troba a l'interior del camí degut a l'acció de l'home, com ara la presència d'altres vehicles, persones o barreres per impedir la circulació. Aquests objectes es caracteritzen per tenir propietats ben diferents a les del camí i a les de l'entorn, i per tant la seva distinció durant la segmentació de la imatge i la posterior classificació resulta relativament fàcil. Per contra, els obstacles naturals són objectes propis de l'entorn, com ara pedres de grans dimensions, arbres o branques d'arbres, desprendiments de terra, etc., que poden presentar característiques similars a les matèries que formen el camí, fent difícil en aquestes condicions, poder diferenciar entre el camí i els obstacles mitjançant les tècniques de visió per ordinador. En aquests casos, l'anàlisi de mapes de profunditat obtinguts mitjançant telemetria làser resulta més adequat per a la detecció d'aquests obstacles d'una forma fiable i robusta.

Atès que en el mètode que es proposa l'única informació disponible de l'entorn és una seqüència d'imatges color, només serem capaços de detectar un obstacle, si aquest és d'un color diferent al de la seva perifèria. Com hem indicat, aquest és generalment el cas dels obstacles artificials però no sempre ho és pels obstacles naturals.

Per una altra banda, no tota regió de color diferent correspon a un obstacle, ja que per definició un camí té una superfície heterogènia (un camí de sorra i herba conté regions clarament diferenciables però ambdues formen part del camí) i si un objecte que es troba a l'interior del camí té un alçada petita (o nul·la), no caldrà considerar-lo com a obstacle, ja que el vehicle pot passar sobre ell. Per tant, és necessari determinar l'alçada de totes les regions que inicialment s'han considerat com obstacle potencial, amb l'objectiu de determinar per a cada regió si es tracta d'un obstacle o no.

Per un vehicle autònom que es desplaça en un entorn natural, com és el cas que considerem, hi ha tres tipus d'obstacles:

1. Els obstacles positius. Són objectes que sobresurten de la superfície sobre la qual es desplaça el mòbil, i poden ser causa d'una col·lisió.
2. Els obstacles negatius. Són forats, sots o rases que suposen un desnivell prou important per que el vehicle es pugui veure immobilitzat o danyat al caure dintre.
3. Els obstacles penjants. Són elements de l'entorn, com ara branques d'arbres, que es troben a una alçada tal per sobre de la zona transitable, que poden causar una col·lisió amb el vehicle.

Encara que els obstacles més freqüents són els positius, si es vol garantir totalment la seguretat del vehicle i dels elements de l'entorn, cal detectar i evitar tots tres tipus d'obstacles.

5.1.1. Tècniques d'Estimació de l'Estructura Tridimensional d'una Escena

Existeixen diferents tècniques en la visió per ordinador que ens permeten determinar característiques tridimensionals dels objectes de l'escena a partir de l'anàlisi de la informació bidimensional subministrada per les imatges captades. Les tècniques més utilitzades són: *l'estereovisió, l'estructura a partir de moviment, la triangulació activa i la profunditat per enfocament*. En l'última dècada, gràcies a l'evolució tecnològica dels sistemes basats en llum làser i la informàtica, han proliferat els sistemes de navegació autònoma que utilitzen la *telemetria làser* per obtenir informació tridimensional de l'escena.

L'estereovisió es basa en l'anàlisi de 2 o més imatges adquirides en el mateix moment des de posicions pròximes i conegudes [Binford 83]. Sobre elles es busquen zones o punts de les imatges amb característiques rellevants com ara, els segments rectilinis, els vèrtexs o zones de la imatge que presenten un gran contrast. Posteriorment s'han d'aparellar les característiques de les N imatges que corresponen al mateix punt o zona característica de l'escena, els quals s'anomenen homòlegs. Finalment, es calcula la posició de la característica a l'espai tridimensional per triangulació a partir de la disparitat (diferència de posició en píxels d'un mateix punt de l'escena en les diferents

imatges). Aquesta tècnica permet calcular la posició dels objectes al món tridimensional analitzant una única parella d'imatges que han estat captades en el mateix instant, sense que sigui necessari cap acció o esdeveniment en l'escena que s'analitza. Així mateix, l'estereovisió permet calcular la posició 3D dels punts d'interès, analitzant únicament la zona d'interès de la imatge. El principal inconvenient que presenta aquesta tècnica és la no existència d'algorismes d'aparellament fiable, especialment en el cas que hi hagin varies característiques semblants a la imatge. Encara que existeixen diferents mètodes per resoldre aquest problema, aquesta és una tasca difícil i només s'obtenen solucions correctes en escenes simples. Generalment, aquest procés es veu complicat per les oclusions que provoquen que certes característiques quedin amagades, l'aparició de característiques falses i la reparició de característiques amagades. Altres limitacions de l'estereovisió són, per una banda, que només es pot calcular la posició a l'espai 3D dels punts que es troben en la zona de solapament de les imatges, i que els punts sobre els que s'obté informació han de presentar imperativament una característica rellevant respecte l'entorn per poder trobar els pixels homòlegs en cada una de les imatges (i per tant no és possible calcular la profunditat de qualsevol pixel de la imatge).

La tècnica d'*estructura a partir de moviment* (en anglès *structure from motion* o *stereo motion*) utilitza una seqüència d'imatges captades en diferent instants de temps entre els que s'ha produït un desplaçament relatiu entre la càmera i l'escena (ja sigui per un desplaçament de la càmera en una escena fixa, pel desplaçament dels objectes de l'escena quan la càmera està fixa o per un moviment simultani de càmera i objectes) [Aggarwal 88] [Jerian 91]. El desplaçament relatiu càmera-objectes es calcula a partir del desplaçament aparent d'un conjunt de característiques o patrons d'intensitat lluminosa associat als objectes, en la seqüència d'imatges. Existeixen dues alternatives a l'hora de calcular el moviment aparent dels objectes a la imatge:

Basat en característiques. S'extrau un conjunt petit, però altament discriminant, de característiques en la imatge bidimensional, corresponents a característiques dels objectes tridimensionals, com ara vèrtexs, límits de superfícies o límits indicant canvis en la reflectància de la superfície [Aggarwal 87] [Tan 92] [Wang 92]. Aquests punts, línies i/o corbes són extrets de cada imatge. Posteriorment, s'estableix la correspondència d'aquestes característiques en la seqüència d'imatges. En aquest procés es tenen en compte un conjunt de restriccions, basades en assumpcions com ara la del moviment dels cossos rígids (la distància a l'espai tridimensional entre dos característiques d'un cos rígid es manté constant independentment de la distància entre l'objecte i la càmera). Aquestes restriccions, normalment, generen un sistema d'equacions no lineals. Els desplaçaments

observats de les característiques obtingudes a la imatge, són utilitzats per resoldre aquestes equacions, obtenint finalment uns paràmetres que caracteritzen el moviment dels objectes a l'escena. Aquesta tècnica, al igual que l'estereovisió, presenta l'inconvenient de necessitar l'aparellament de les característiques extrems de dues imatges consecutives, amb els problemes i limitacions que això suposa.

Basat en el flux òptic. En aquest cas es calcula el flux òptic (o velocitat instantània dels valors d'intensitat en la imatge) [Prazdny 79] [Waxman 87]. En lloc de considerar els canvis temporals en la imatge intensitat a l'hora de calcular el flux òptic, també es pot considerar els canvis temporals en uns valors obtinguts com a resultat d'aplicar un operador local, com ara el contrast, l'entropia o les derivades espacials, a la imatge intensitat original. En ambdós casos, s'obté una imatge de flux dens, normalment un valor de flux per a cada pixel de la imatge original. Finalment, el flux òptic, juntament amb restriccions o informacions corresponents a l'escena, s'utilitzen per determinar la velocitat relativa actual a l'espai tridimensional entre la càmera i els objectes de l'escena. El fet que el flux òptic es basi en el càlcul de les derivades parcials de la imatge intensitat fa que aquesta tècnica sigui molt sensible al soroll.

La tècnica d'estructura a partir de moviment requereix, o bé de l'anàlisi d'una seqüència llarga d'imatges per conèixer l'estructura i el moviment relatiu de l'entorn, o de l'anàlisi de poques imatges (un mínim de 2) si es coneix el desplaçament de la càmera. En aquest últim cas, els errors presents en la estimació del moviment de la càmera, es reflecteixen en un error en el càlcul de l'estructura de l'escena.

La *triangulació activa* es basa en la projecció sobre l'escena d'uns patrons de llum estructurada coneguts. A partir de l'anàlisi de la deformació observada a la imatge intensitat del patró de llum emès, es calculen les característiques geomètriques de les superfícies. El patró de llum pot ser un conjunt de franges, una retícula de línies o un punt de llum que escombra la zona d'interès de l'escena.

En els últims anys, gràcies als progressos realitzats en el camp del VLSI analògic, s'han obert noves possibilitats per aquesta tècnica, en les que es combinen amb idees de la tècnica de mesura per temps de vol. S'estan desenvolupant nous sensors tipus CCD on cada element sensitiu enregistra en paral·lel el temps transcorregut des de l'emissió fins a la detecció de la llum provenint d'un punter làser que escombra l'escena a alta velocitat [Nayar 95]. La tensió de sortida de cada element del sensor està directament relacionada amb la distància entre el sensor i els objectes que es troben en el seu camp de visió. En aquesta tècnica, la informació subministrada pel

sensor permet obtenir la profunditat sense necessitat de realitzar un gran nombre de càlculs. La seva resolució espacial és, avui en dia, limitada (només s'han integrat fins 128x128 cel·les en un xip) i poden subministrar una nova imatge cada milisegon [Araki 91]. Per contra requereix de l'emissió d'una radiació, la qual cosa comporta un consum d'energia elevat, tenen una baixa resolució espacial, i el seu abast i rang d'operació amb precisió es reduït (arriben fins a pocs metres), condicions que només fan viables la utilització de la triangulació activa en entorns interior.

En la tècnica de *profunditat per enfocament*, a diferència de l'estereovisió i l'estructura per moviment, es prenen un conjunt d'imatges des del mateix punt però amb diferents configuracions de l'òptica. En el cas de profunditat per enfocament, s'adquireix un conjunt gran d'imatges mentre es desplaça en petits increments l'enfocament de la càmera [Grossman 87] [Asada 93]. Posteriorment s'utilitza un operador enfocament (generalment un operador sensible a les altes freqüència, com ara el filtre Laplaciana, que és un indicador de l'enfocament de la imatge) per detectar per a cada pixel en quina imatge del conjunt es produeix el millor enfocament. En canvi, la profunditat per desenfocament normalment utilitza dues imatges i a partir de l'estimació del desenfocament relatiu entre les imatges s'obté la profunditat [Nayar 96]. L'ús d'aquesta tècnica està limitat a objectes texturats, ja que no és possible avaluar l'enfocament/desenfocament de superfícies llises. Per superar aquesta limitació en [Nayar 95] es proposa la projecció sobre l'escena d'un patró de llum amb l'objectiu de fer aparèixer una textura, fins i tot sobre les superfícies llises. Al igual que en la triangulació activa, la projecció d'un patró de llum només es possible sobre distàncies curtes, ja que la possibilitat de percebre sobre l'escena el patró de llum projectat, es redueix amb la distància.

Encara que la profunditat per enfocament pot donar un mapa dens de profunditats (tanta resolució espacial com elements té la imatge digital original), es requereix una gran capacitat computacional per a la seva realització.

Els sistemes de *telemetria làser* permeten obtenir un mapa dens de profunditat de l'escena, on la mateixa imatge subministrada pel sistema de càlcul de distància per escombrat d'un làser (LRF), ja és una bona aproximació a l'estructura 3D de l'escena captada. Existeixen dues alternatives per al càlcul de la distància :

Càlcul del temps de vol. S'emet una seqüència de polsos de llum làser en direcció a l'objecte del qual es vol conèixer la distància a la que es troba, i es comptabilitza el temps necessari per a que la llum sigui reflectida per l'objecte i torni al receptor. La distància emissor/receptor-objecte és proporcional a aquest

temps, i es pot calcular ja que es coneix la velocitat de propagació de la llum a l'aire. Aquests sistemes permeten mesurar distàncies en el marge de 2 metres fins distàncies llargues (al voltant de 1 quilòmetre) amb una precisió, en el millor cas, de 10 centímetres [Borenstein 96].

Desplaçament de fase. En aquest cas s'emet un feix continu de llum que es modula en amplitud. Una petita part de l'energia emesa és reflectida per la superfície de l'objecte i retorna cap al detector. L'energia retornada es compara amb el senyal que es transmet ja que el desfasament entre els dos senyals és proporcional a la distància recorreguda pel senyal retornat. La relació entre desfasament i distància queda definida per $\phi = \frac{4\pi d}{\lambda}$. El principal problema d'aquest sistema és l'anomenat *interval d'ambigüitat* ja que si la distància recorreguda és més gran que la longitud d'ona de la moduladora, totes les distàncies $d=(x+n\lambda)/2$ provoquen el mateix desfasament (on n és un enter i λ és la longitud d'ona de la moduladora). Per contra l'avantatge d'aquest sistema és la possibilitat de mesurar també la direcció i velocitat relativa dels objectes mòbils gràcies a l'efecte Doppler. El marge de mesura d'aquests sistemes és més reduït, a causa de l'interval d'ambigüitat, éssent un valor típic de 1 metre fins a 30 metres, amb una precisió que arriba fins a 1 centímetre.

Si bé els LRF subministren directament informació tridimensional de l'escena amb molt bona precisió, aquests sistemes presenten inconvenients importants, com són el seu cost elevat, el gran consum d'energia i, per certes aplicacions, el fet que es tracti d'un sensor actiu.

Com s'ha pogut veure, l'única tècnica que només requereix d'una càmera per obtenir informació tridimensional de l'escena és la d'estructura a partir del moviment. Totes les altres tècniques requereixen, ordenades en funció de la complexitat dels elements addicionals, o bé de dues o més càmeres (o una estructura mecànica que desplaça a diferents posicions una càmera per prendre les imatges) i d'un sistema d'adquisició d'imatges més sofisticat en el cas de l'estereovisió, o d'una càmera amb enfocament controlable electrònicament en el cas de la profunditat per enfocament, o d'un sistema per a la projecció de patrons de llum en el cas de la triangulació activa, o d'un sistema emissor/receptor de llum làser disposat sobre una estructura mecànic-òptica molt sofisticada i precisa en el cas de la telemetria làser.

Pel que fa referència a la complexitat del procés necessari per analitzar les dades subministrades pels diferents tècniques d'estimació de l'estructura tridimensional, la

telemetria làser resulta la més simple, éssent totes les altres tècniques, que es basen en l'anàlisi de la imatge, d'un cost computacional comparable.

Les tècniques més utilitzades, i de menor cost, per a la detecció d'obstacles són l'estereovisió i les diferents versions, imposant condicions d'operació i/o de l'entorn diferents, de l'estructura a partir de moviment.

Els avantatges de l'estereovisió, quan s'utilitza per obtenir informació 3D de l'entorn des d'un vehicle en moviment, són la necessitat de només una parella d'imatges captades en el mateix instant de temps per calcular la profunditat, i la poca influència que les rotacions de les càmeres al voltant dels eixos definits pel pla imatge, tenen en el càlcul de la profunditat d'un punt. Els principals problemes que presenta l'estereovisió són [Huber 94]:

- La necessitat d'una calibració molt precisa per l'obtenció de resultats correctes. Una petita rotació del conjunt de les càmeres al voltant de l'eix òptic (que pot ser deguda a les irregularitats de la superfície per la que es desplaça el vehicle) desfà la restricció de la línia epipolar.
- La distància òptima entre càmeres depèn de la distància a la que es volen fer les mesures; si les càmeres estan molt separades es pot calcular la profunditat d'objectes distants, però el problema de la correspondència entre punts de les imatges es complica considerablement; si la distància entre càmeres és petita, l'error en les mesures és elevat. Tenint en compte que el gran marge de mesura necessari en els sistemes de navegació autònoms, no és possible fixar una configuració del sistema que permeti realitzar bones mesures en tot el marge.
- Només és possible mesurar la profunditat d'aquells punts que formen part de la zona en la que les dues imatges es solapen. El solapament entre les dues imatges depèn de la distància entre càmeres, i per tant afecta al marge de mesura del sensor.
- Els errors en l'aparellament dels punts homòlegs, generen un error relatiu important en el càlcul de la profunditat.

La tècnica d'estructura a partir de moviment permet obtenir resultats més precisos que en el cas de l'estereovisió [Huber 94], en canvi el principal problema que presenta quan s'utilitza per obtenir informació 3D de l'entorn des d'un vehicle en moviment, és la necessitat de considerar un gran nombre de punts extrets d'una seqüència d'imatges i resoldre un sistema complexa d'equacions no lineals. La complexitat d'aquest procés es pot reduir considerablement si es disposa d'informació complementària com és el

moviment de la càmera al llarg de la seqüència, que el moviment aparent a la imatge és únicament degut al moviment de la càmera, o que els objectes es troben sobre el pla de terra.

Un altre aspecte interessant de la tècnica d'estructura a partir de moviment, en el cas que es considera -la detecció d'obstacles en un sistema de navegació en camins feblement estructurats-, és la possibilitat de que una part del processament de la imatge, la segmentació i posterior classificació de les regions en funció del seu color i posició a la imatge, és comú per les dues tasques bàsiques que cal realitzar: la detecció i seguiment del camí, i la detecció d'obstacles. En la imatge classificada, s'indica quines són les regions que són susceptibles de ser obstacle i quina àrea defineixen a la imatge. En aquest cas, no és necessari realitzar una nova segmentació de la imatge per a la detecció dels obstacles, la qual cosa ens permet reduir el temps de cicle i/o el cost computacional.

Tenint en compte que la tècnica d'estructura a partir de moviment permet realitzar, en les condicions d'operació considerades, la tasca de detecció d'obstacles amb una eficiència i robustesa similar a les altres tècniques presentades, i a un cost computacional inferior, ens proposem basar-nos en aquesta tècnica.

5.1.2. Càlcul de l'Alçada dels Objectes a partir de l'Anàlisi d'una Seqüència d'Imatges

Per a que un vehicle autònom pugui detectar i evitar un obstacle no és necessari obtenir-ne una descripció tridimensional molt detallada de l'entorn. Coneixent aproximadament les dimensions dels objectes i la seva posició, és suficient per determinar si un objecte donat és un obstacle i evitar-lo. El nostre objectiu és el disseny d'un mètode que permeti obtenir aquesta informació amb els mínims requeriments sensorials, potència de càlcul i temps de procés.

Una vegada detectat un obstacle, es pot deduir per anàlisi geomètric, analitzant una única imatge, la seva posició a l'entorn i la amplada de l'obstacle a partir del coneixement que es té del procés de formació de la imatge (veure annex A), si s'assumeix que el pendent del terra de l'àrea del camí analitzada és constant i si es coneixen els paràmetres intrínsecs i extrínsecs del sistema d'adquisició d'imatges.

Ens queda obert el problema del càlcul de l'alçada dels objectes. En el nostre cas, es considera com obstacle, tot objecte d'alçada superior a un cert llindar que es troba a l'interior del camí. Així doncs, el problema de la detecció d'obstacles queda centrat en el càlcul de l'alçada dels objectes tridimensionals de l'escena a partir de la informació continguda en les imatges bidimensionals que n'obtenim.

El mètode que es proposa per al càlcul de l'alçada dels objectes es pot englobar dintre de la tècnica d'*alçada a partir de moviment*, és a dir, es determina l'alçada d'un objecte a partir de l'anàlisi de l'alçada de la regió que defineix l'objecte en una seqüència d'imatges. Aquest mètode és una adaptació del d'estructura a partir de moviment, ja que s'utilitza informació addicional, com és el coneixement del desplaçament relatiu del vehicle entre l'adquisició de les imatges -obtingut a partir dels sensors odomètrics-, i a més, no es pretén obtenir tota la informació que ens permet aquest mètode, com són les característiques tridimensionals dels objectes i/o determinar el moviment relatiu entre càmera-objecte, sinó únicament es calcula una aproximació de l'alçada dels objectes.

El mètode que es proposa, basat en el principi d'**hipòtesi i test**, està orientat a estimar de forma aproximada l'alçada dels objectes de l'entorn que han generat una regió en la seqüència d'imatges. Per això, en cada imatge de la seqüència cal obtenir un únic paràmetre de cada regió analitzada, l'alçada de la regió. Les tres restriccions i condicions d'entorn que s'imposen amb l'objectiu de reduir el temps de procés són:

1. Només es consideren un conjunt reduït i discret de possibles alçades pels objectes. La resolució amb que es defineix aquest conjunt de possibles alçades, determinarà l'error mínim a l'hora d'estimar l'alçada dels objectes.
2. Els objectes dels quals s'estima l'alçada es troben al mateix pla sobre el que avança el mòbil. Aquesta restricció és coherent amb la que també s'ha inclòs en el model genèric de camí (el pendent del camí és localment constant).
3. El vehicle disposa de sensors odomètrics que ens permeten conèixer el desplaçament del vehicle.

El mètode que es proposa, partint d'una regió **R** que ha estat classificada com obstacle potencial, es realitza en dos passos:

1. Definició de les hipòtesis sobre l'alçada de l'objecte **O**, a partir de l'alçada de la regió **R** de la imatge, en la primera imatge d'una seqüència, associada a l'objecte **O**.

2. Verificació de les hipòtesis formulades, a partir de l'evolució de l'alçada de les regions associades a l'objecte **O** en la seqüència d'imatges.

Per a la regió **R** que s'analitza en la imatge inicial d'una seqüència, s'estableixen **N** hipòtesis diferents sobre l'alçada que té l'objecte **O** de l'escena associat a la regió (Figura 5.1). Anomenem **objectes-hipotètics** al conjunt d'objectes imaginaris que es troben a la mateixa posició, tenen alçades diferents, però quan es projecten a la imatge, generen la mateixa regió **R**. A la figura 5.1 podem veure com la regió **R** de la imatge pot ser generada per un dels dos objectes-hipotètics considerats. Aquests objectes imaginaris seran utilitzats per calcular l'alçada de l'objecte **O** de l'entorn associat a la regió **R**. Per definició, tots els objectes-hipotètics defineixen la mateixa regió a la imatge inicial.

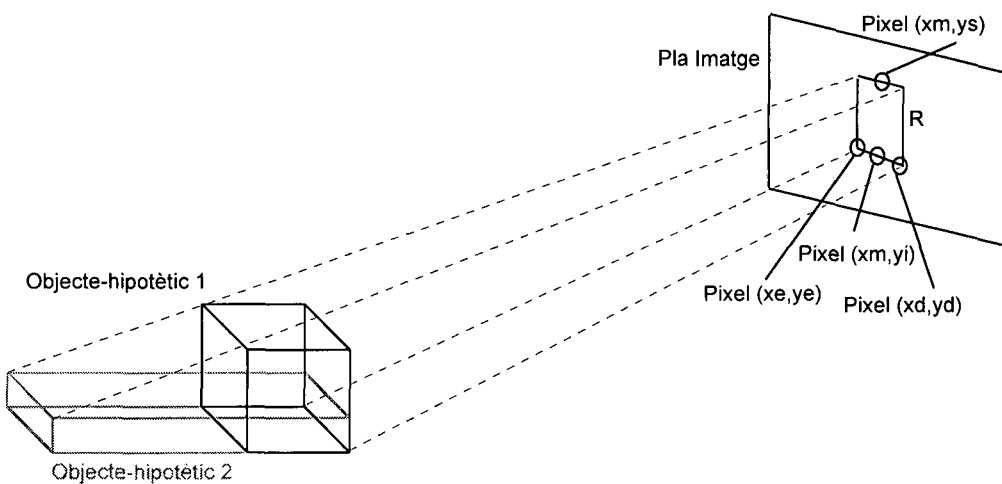


Figura 5.1. Relació entre una regió de la imatge i el conjunt d'objectes-hipotètics. Els dos objectes-hipotètics es troben en la mateixa posició però tenen dimensions diferents.

Si la distància entre la càmera i el conjunt d'objectes-hipotètics varia, cadascun d'ells generarà una regió d'alçada diferent en les posteriors imatges. Anomenem **regió-hipotètica** a la regió imaginària que defineix un objecte-hipotètic a la imatge.

Donada una regió R_0 de la imatge inicial, anomenem **regions homòlogues de R_0** , a les regions R_1, R_2, \dots, R_M , cada una d'elles provinents d'una imatge diferent de la seqüència, que han estat generades pel mateix objecte **O** de l'entorn.

En el segon pas, s'avalua en les posteriors imatges de la seqüència, quina de les hipòtesis inicials s'ajusta més a la realitat, mitjançant l'anàlisi de l'evolució de l'alçada de les regions homòlogues de R_0 . En concret, es calcula en cada imatge i de la seqüència, les diferències entre l'alçada de la regió homòloga R_i i l'alçada de cada una de les **N** regions-hipotètiques associades als **N** objectes-hipotètics definits al primer

pas. Les diferències entre la regió homòloga R_i i cada una de les N regions-hipotètiques, es va acumulant al llarg de la seqüència d'imatges.

Es selecciona com l'alçada estimada de l'objecte, l'alçada de l'objecte-hipotètic k que en l'anàlisi de la seqüència d'imatges, ha generat una diferència d'alçada acumulada mínima entre les regions-hipotètiques k i les regions homòlogues R_i .

En aquesta etapa d'estimació de l'alçada, s'associa el concepte d'alçada d'una regió, al nombre de pixels de la regió en cada columna de la imatge. Sigui $NP_c(R)$ el nombre de pixels de la regió R a la columna c de la imatge. Es defineix l'alçada de la regió R com el valor màxim de la mitjana del nombre de pixels de R en $2 \cdot C + 1$ columnes consecutives (figura 5.2), on C és un nombre enter natural que indica el nombre de columnes que cal considerar per determinar l'alçada de la regió.

$$\text{Alçada}(R) = \text{màxim} \left(\frac{\sum_{c=i-C}^{i+C} NP_c(R)}{2 \cdot C + 1} \right)$$

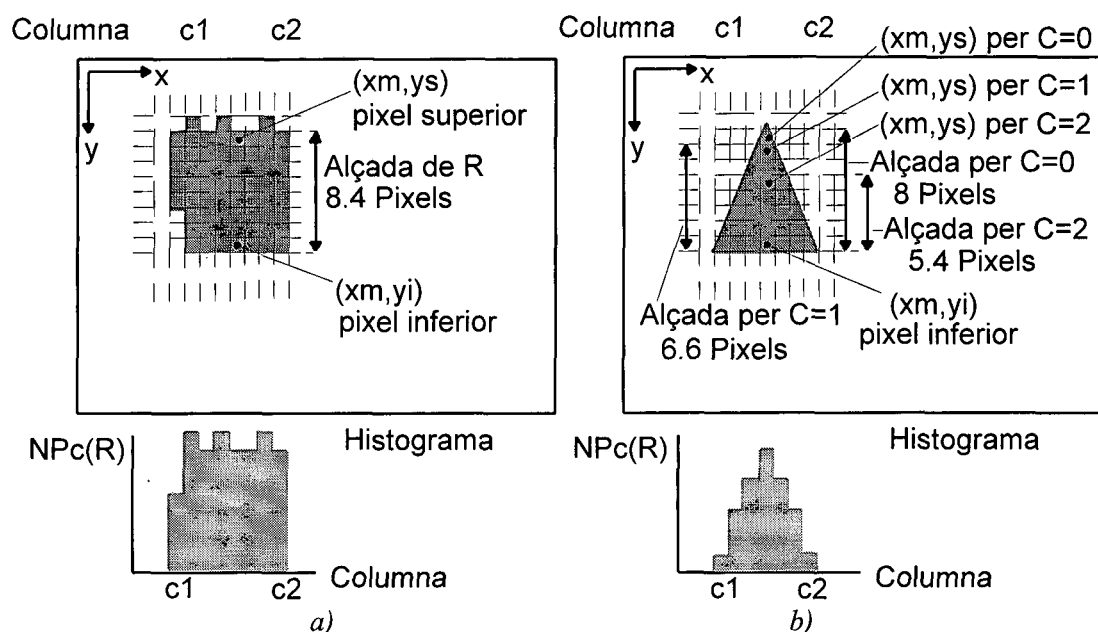


Figura 5.2. Dos exemples de determinació de l'alçada d'una regió de la imatge.

Com l'alçada de la regió està associada a la seva forma a la imatge, els errors degut a la discretització de la imatge i els generats per la segmentació, que fan variar lleugerament les formes de les regions homòlogues, afecten a l'hora de determinar l'alçada de la regió. La figura 5.2.a mostra aquesta situació, on l'alçada de l'objecte és constant, però els errors de discretització i segmentació fan que el contorn superior de la regió no sigui horitzontal. És per aquesta raó, que per determinar l'alçada de la regió,

es consideren $2 \cdot C + 1$ columnes en comptes d'una única columna. D'aquesta forma es detectarà una característica de les regions homòlogues, analitzades al llarg de la seqüència d'imatges, que depèn d'una zona de la regió ($2 \cdot C + 1$ columnes) en comptes d'una característica puntual, fet que ens permet ser menys sensible als errors que modifiquen la forma de les regions. Aquesta mesura de l'alçada de les regions s'ha mostrat més estable que la utilització del valor màxim de $NP_c(R)$ o l'alçada del rectangle envoltant de la regió R a la imatge (situacions que són equivalents a calcular $Alçada(R)$ fixant $C=0$).

Com més gran sigui el valor de C , menys influència tindran els errors de segmentació i discretització en el càlcul de l'alçada de la regió. Per contra si s'opta per valors de C grans (>2), en algunes situacions es pot obtenir un valor d' $Alçada(R)$ bastant diferent al que està associat amb l'alçada de l'objecte, i que permet calcular de forma correcta l'alçada màxima de l'objecte associat a R . En la figura 5.2.b es pot veure la regió generada per un con (un objecte d'alçada no constant). En aquest cas on no hi ha errors de segmentació, el valor d' $Alçada(R)$ que permet estimar correctament l'alçada màxima del con és 8 (valor màxim de l'histograma que s'obtindria amb $C=0$). Per $C=1$ (mitjana en 3 columnes) s'obté un valor d' $Alçada(R)=6.6$, i per $C=2$ (mitjana en 5 columnes) s'obté un valor d' $Alçada(R)=5.4$. Per tant, la determinació del valor de C que permet reduir l'efecte dels errors de discretització i segmentació, sense que això afecti significativament a l'hora de calcular l'alçada màxima, depèn de la forma de la regió. A partir de l'anàlisi de l'histograma, es pot deduir si la regió R presenta una alçada constant (els valors de $NP_c(R)$ són constants o quasi constants) o variable al llarg de les columnes (els valors de $NP_c(R)$ varien significativament en les diferents columnes). El nombre de columnes utilitzades per calcular l'alçada de la regió en la columna i , es determina de la següent forma:

$$C = \begin{cases} 3 & \text{si Màxim}(NP_c(R)) - \text{Mínim}(NP_c(R)) \leq N \\ 1 & \text{si Màxim}(NP_c(R)) - \text{Mínim}(NP_c(R)) > N \end{cases} \quad \forall c = (i-2, i+2)$$

El procediment seguit per a l'estimació de l'alçada d'un objecte, que ha generat una regió R_0 a la imatge inicial d'una seqüència, consta dels següents passos:

1. Calcular l'histograma de les columnes c corresponents als píxels que pertanyen a la regió R_0 a la imatge inicial -nombre de píxels de la regió en cada columna de la imatge- (figura 5.2). D'aquesta forma s'obtenen els valors $NP_c(R_0)$.
2. Determinar la posició x_m a l'histograma per a la qual la mitjana del nombre de píxels de $2 \cdot C + 1$ columnes consecutives és màxim.

$$x_m = i \quad \text{on } i \text{ és tal que } \frac{\sum_{c=i-C}^{i+C} NP_c(R)}{2 * C + 1} \text{ és màxim.}$$

3. Determinar els valors y_i del pixel inferior i y_s del pixel superior de la regió R_0 a la columna x_m (Figura 5.2). Per tant, de la regió R_0 s'obtenen dues coordenades: (x_m, y_s) i (x_m, y_i) . Al pixel (x_m, y_s) l'anomenarem pixel superior de la regió i al pixel (x_m, y_i) l'anomenarem pixel inferior de la regió $-y_s = y_i + \text{Alçada}(R)$.

4. Definir una família F de N objectes-hipotètics, tal que tots ells generen la mateixa regió quan es projecten a la imatge, i que és similar a la regió R_0 que s'analitza -en el sentit que defineixen els mateixos valors (x_m, y_s) i (x_m, y_i) a la imatge-. Els N objectes de la família F es caracteritzen per que les seves alçades es diferencien en un valor *increment_alçada*. L'alçada de l'objecte-hipotètic k (h_k) de la família F és:

$$h_k = k * \text{increment_alçada} \quad \text{per } 0 \leq k < N$$

En aquest mètode l'alçada estimada de l'objecte coincidirà amb un dels valors h_k considerats.

5. Repetir els passos 1 a 3 per a la regió R_j , homòloga de la regió R_0 , en cada nova imatge j de la seqüència. Per tal d'establir la correspondència entre les regions de dues imatges consecutives d'una seqüència, es té en compte el desplaçament del mòbil entre les imatges, que es coneix gràcies als sensors odomètrics; i s'imposa la restricció de que el color de les regions homòlogues és semblant.

6. Calcular els valors (x'_m, y'_s) i (x'_m, y'_i) dels pixels superior i inferior per la regió R_j , calcular l'error entre la regió R_j i cada una de les N regions-hipotètiques. L'error s'obté:

$$\text{Error}_{k_j} = |x'_m - x'_{mk}| + |y'_s - y'_{sk}|$$

on x'_{mk} i y'_{sk} són els valors del pixel superior de la regió-hipotètica k a la imatge j . Per a cada regió-hipotètica k de la família, s'acumula l'error que genera respecte la regió real R_j en la seqüència d'imatges.

$$\text{Error}_k = \sum_{j=0}^n \text{Error}_{k_j} \quad n: \text{ nombre d'imatges analitzades}$$

7. Es pren com estimació de l'alçada de l'objecte que defineix la regió R_0 (i les seves regions homòlogues R_j en la seqüència d'imatges que s'analitza), aquella que correspon a l'objecte-hipotètic k de la família F que ha generat un error acumulat mínim.

$$\text{Alçada_estimada}(R) = \text{mínim}(\text{Error}_k) \quad \forall k \in F$$

8. Anar al pas 5.

Amb l'objectiu de minimitzar l'error generat per la discretització de la imatge, les característiques tridimensionals de la família d'objectes hipòtesis F es van actualitzant a mida que el mòbil es desplaça cap a l'objecte real. D'aquesta forma es pot aprofitar el fet que donat que la distància càmera-objecte ha disminuït, els errors en l'estimació de les característiques tridimensionals dels objectes, deguts a la discretització de la imatge, es redueixen (veure annex A).

Aquest mètode ens permet estimar, per a cada regió de la imatge, l'alçada de l'objecte de l'escena que, en projectar-se a la imatge, genera la regió. Aquesta informació és la que finalment s'utilitza per determinar quines regions de la imatge corresponen a obstacles dintre del camí.

La figura 5.3 il·lustra el funcionament d'aquesta tècnica. Com és ben conegut, hi ha diferents objectes, amb diferents alçades, que generen la mateixa projecció a la imatge digital. Donada una regió R_0 a la imatge inicial (que en la figura 5.3 està generada per l'objecte de traç negre), es consideren N objectes hipotètics d'alçades diferents (a la figura 5.3.a hi ha quatre objectes-hipotètics amb alçada h_i i llargada p_i) que tindrien com a projecció a la imatge, la mateixa regió R_0 . En les imatges posteriors, figura 5.3.b i c, obtingudes després que el mòbil s'ha desplaçat una distància D , es compararà l'alçada de la regió R_j de la imatge, amb l'alçada de les quatre regions-hipotètiques, i es calcula l'error e_{kj} per a cada una de les regions-hipotètiques (e_{01} , e_{11} , e_{21} , e_{31} en la figura 5.3.b, i e_{0j} , e_{1j} , e_{2j} , e_{3j} en la figura 5.3.c). Aquest procés es repeteix successivament (figura 5.3.c) i l'error calculat per a cada objecte-hipotètic considerat es va acumulant. En cada moment es pren com alçada aproximada de l'objecte, aquella que ha generat un error acumulat mínim (en el cas de l'exemple de la figura 5.3. seria l'objecte amb alçada h_2).

Com es pot veure aquesta tècnica és incremental, atès que partint d'una estimació inicial de l'alçada de l'objecte, normalment poc correcte, aquesta és millorada amb cada nova imatge de la seqüència analitzada, fins arribar a determinar, amb un cert error, l'alçada de l'objecte. La tècnica presentada té com a característica important, que justifica la pèrdua de resolució a l'hora de calcular l'alçada dels objectes, un baix cost computacional. Això és possible gràcies a que els elements bàsics de la imatge que són utilitzats, les regions, són subministrades per l'etapa de segmentació i classificació (que es realitzen per a la determinació de la posició del camí a la imatge) i al reduït nombre d'operacions que cal realitzar per estimar l'alçada dels objectes en la tècnica d'hipòtesi i test (veure secció 5.1.3). En la tècnica proposada no cal formular i resoldre

un conjunt d'equacions diferencials com és el cas de les altres tècniques que estan dintre del mètode d'estructura a partir de moviment.

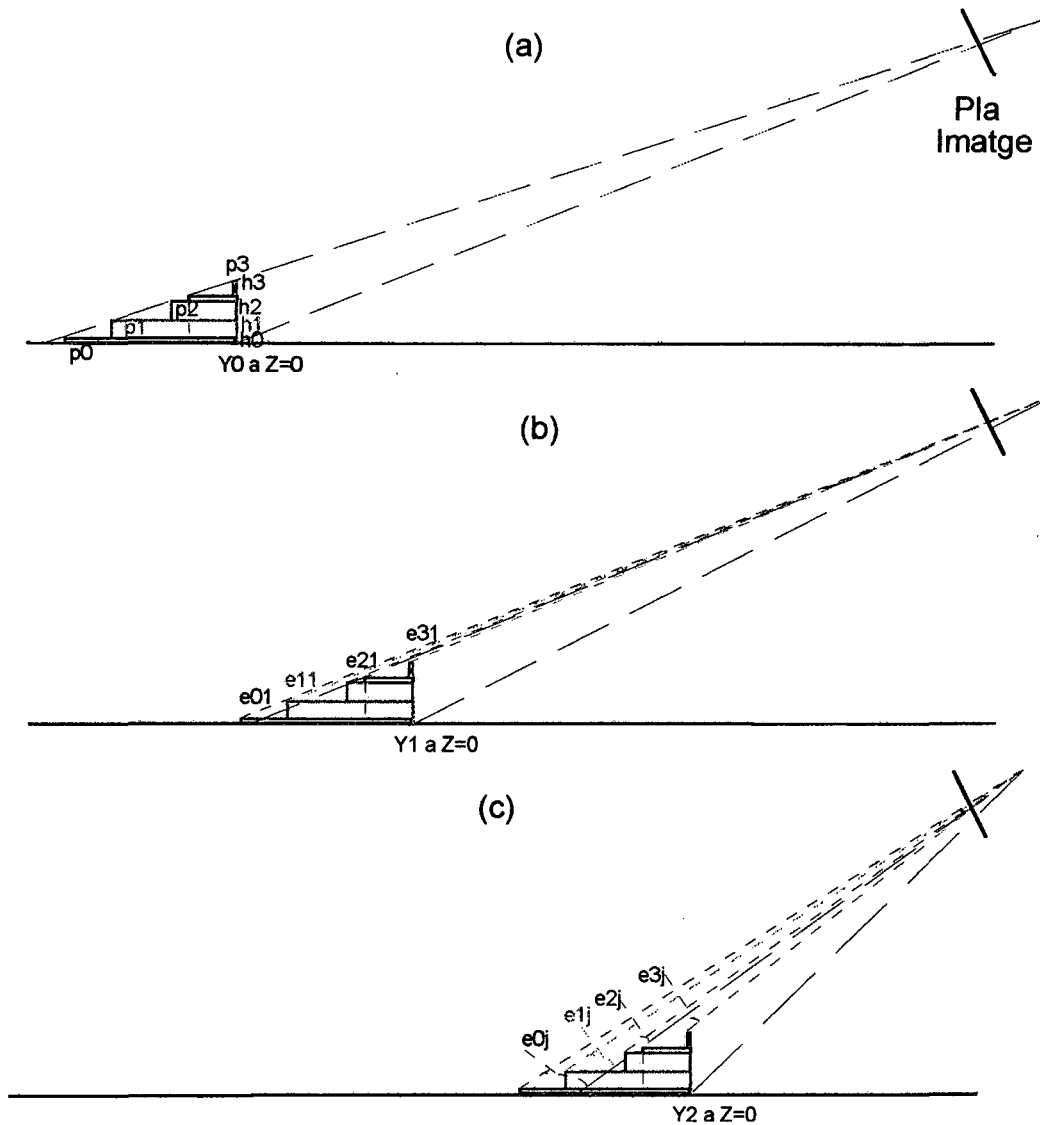


Figura 5.3. Mètode d'hipòtesi i test per al càlcul de l'alçada dels objectes.
a) Imatge inicial. b) i c) Imatges posteriors, després del desplaçament del mòbil, amb càlcul de l'error.

Una primera forma d'analitzar l'error en el càlcul de l'alçada del mètode proposat, és l'anàlisi d'imatges sintètiques. La figura 5.4 mostra una seqüència de 9 imatges sintètiques on apareix un camí i 6 objectes (O_1, O_2, O_3, O_4, O_5 i O_6) amb la mateixa base (30 per 30 centímetres), i diferents alçades (d'esquerra a dreta a les imatges, alçades de 0, 10, 20, 40, 60 i 80 centímetres respectivament). En la imatge 5.4.a els objectes es troben a 18 metres de la càmera, i per les imatges posteriors hi ha un desplaçament relatiu de 1.5 metres.

La figura 5.5 mostra l'error mig en l'estimació de l'alçada dels 6 objectes O1 a O6 al llarg de quinze seqüència d'imatges sintètiques, similar a la de la figura 5.4, on l'alçada i la posició inicial dels objectes s'ha modificat lleugerament en cada seqüència¹. L'estimació de l'alçada dels objectes s'ha realitzat amb 20 objectes-hipotètics, que es diferencien en 5 centímetres d'alçada. Com s'observa a la figura 5.5, l'error mig en l'estimació de l'alçada dels objectes O1 a O6 esta en l'interval de 1.0 a 1.7 centímetres. Un error d'aquesta magnitud és tolerable, ja que per determinar si un objecte donat pot suposar un obstacle pel vehicle, no és necessari conèixer la seva alçada de forma exacta. Al sisè capítol 6, s'avaluarà el funcionament del mètode d'hipòtesi i test quan aquest analitza seqüències d'imatges captades en situacions reals.

En el mètode per a l'estimació de l'alçada proposat, l'error té tendència a decreixer quan augmenta el nombre d'imatges analitzades, el que equival a una disminució de la distància càmera-objecte. Això és degut a dues raons, la primera és que es disposa de més informació per estimar l'alçada, i la segona és l'augment de la resolució espacial amb la que s'observen els objectes, a mida que la distància càmera-objecte disminueix.

L'error mig i màxim en l'estimació de l'alçada depèn del valor d'*increment_alçada* utilitzat per generar la família F d'objectes hipotètics. Si la resolució és de 10 centímetres l'error mig és de 4.22 centímetres i l'error màxim és de 6.5 centímetres, amb resolució de 5 centímetres, els errors mig i màxim són de 2.21 i 3.5 centímetres, i per la resolució de 1 centímetres en la definició de les hipòtesi d'alçades, els errors mig i màxim són de 1.48 i 2.5 centímetres. Per tant, el valor d'*increment_alçada* a utilitzar per definir el conjunt d'objectes hipotètics depèn del màxim error que es vol assumir en el moment d'estimar l'alçada dels objectes. De totes formes, com no es requereix una descripció 3D detallada dels obstacles per que siguin detectats i evitats, no és necessari treballar amb valor d'*increment_alçada* petit. En la secció 6.1, s'analitza com varia l'error en l'estimació de l'alçada en funció del valor d'*increment_alçada*.

D'una altre banda, per obtenir una estimació de l'alçada de l'objecte que s'aproximi al valor real, només cal analitzar dues imatges de la seqüència. Amb aquesta informació, ja s'està en condicions de determinar si l'alçada de l'objecte és superior al lllindar H que s'utilitza per determinar si l'objecte és considerat com obstacle.

¹ De les tres seqüències analitzades una d'elles és la de la figura 5.4. Les altres dos són de les mateixes característiques que l'anterior, però en aquests cas els objectes O1 a O6 tenen alçades de 1.5, 8.5, 18.5, 38.5, 58.5 i 78.5 centímetres respectivament, en la segona seqüència i en la tercera, l'alçades dels objectes O1 a O6 són 3.5, 6.5, 16.5, 36.5, 56.5 i 76.5 centímetres respectivament.

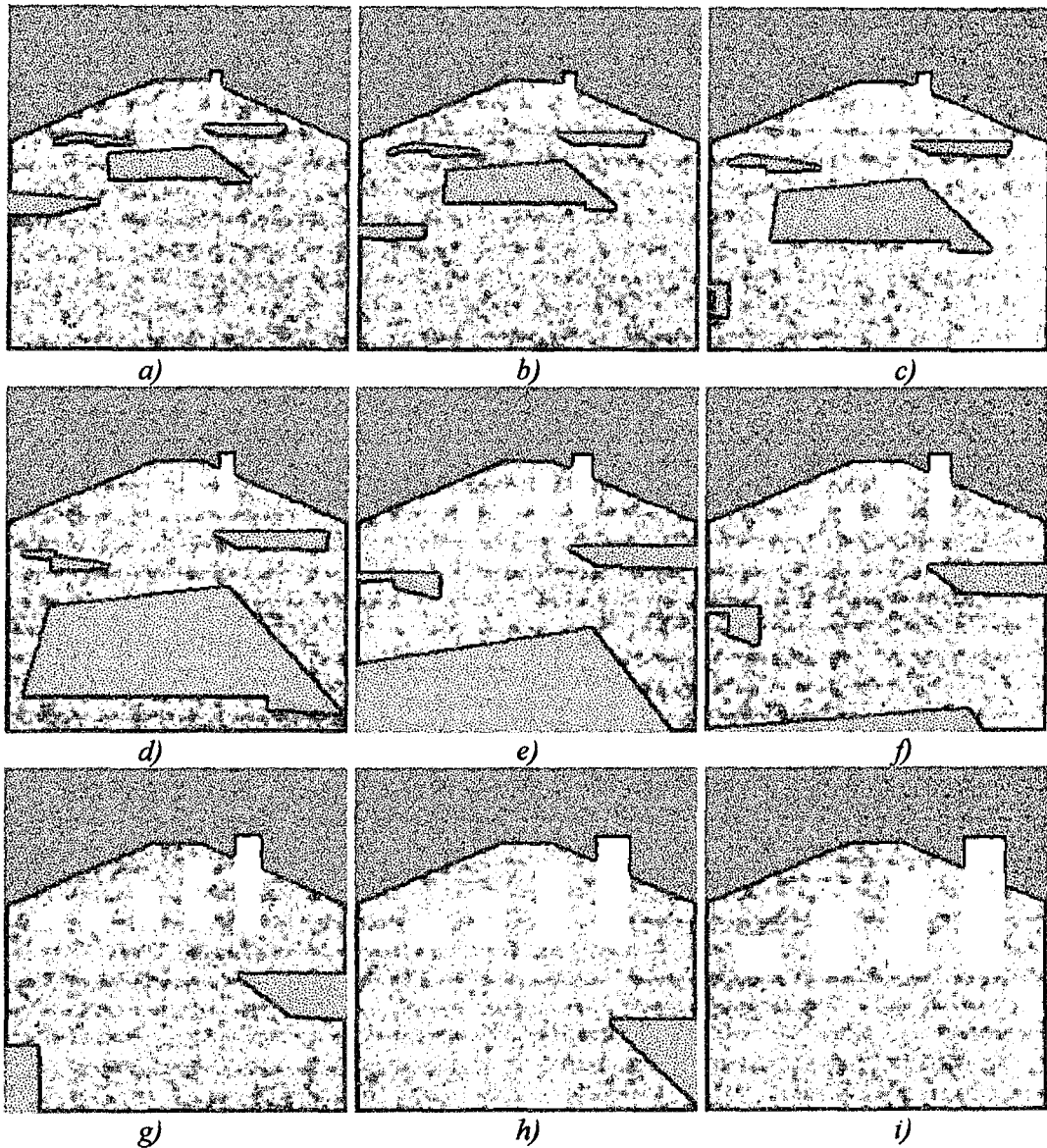


Figura 5.4. Seqüència d'imatges sintètiques on apareixen 6 obstacles d'alçades 0, 10, 20, 40, 60 i 80 centímetres.
a-i) Obstacles a una distància de la càmera de 18, 16.5, 15, 13.5, 12, 10.5, 9, 7.5 i 6 metres respectivament.

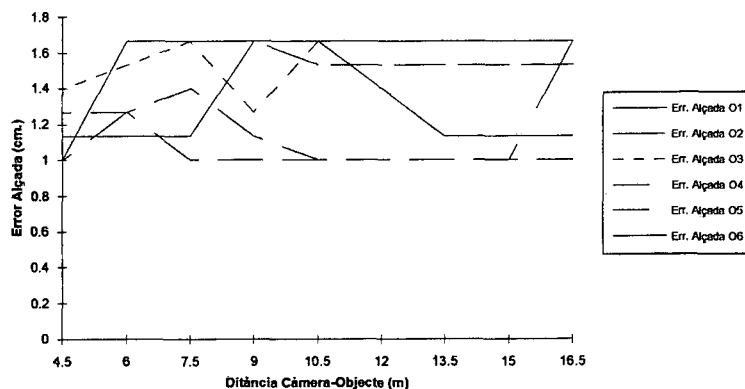


Figura 5.5. Error mig en l'estimació de l'alçada dels objectes O1 a O6 mitjançant la tècnica d'hipòtesi i test, en funció de l'alçada de l'objecte i la distància càmera-objecte.

Finalment, totes les regions que en la imatge inicial han estat classificades com obstacle potencial, i que durant l'anàlisi de les seves regions homòlogues s'ha determinat que l'alçada de l'objecte associat a la regió té una alçada superior al llindar prefixat H , es classifiquen com obstacle i s'incorporaran a la descripció de l'entorn. En canvi, si l'alçada del objecte és inferior a H , la regió es classificarà com camí si la regió està envoltada per regions classificades com camí o obstacle, i com a marge en cas contrari.

El procés d'estimació de l'alçada mitjançant la tècnica d'hipòtesi i test, es basa en disposar d'una bona estimació de la distància càmera-objecte, del que es vol calcular l'alçada. Si el sistema d'adquisició d'imatges està calibrat, es pot determinar la posició i l'amplada de l'objecte a partir de la posició de la regió R classificada com a obstacle a la imatge (veure annex A). Per a la regió R es determinen la posició dels pixels inferior esquerra (x_e, y_e) i dret (x_d, y_d) de la regió (veure figura 5.1), i mitjançant les equacions A.6 i A.7 s'estima la posició dels punts $(X_e, Y_e, 0.0)$ i $(X_d, Y_d, 0.0)$ de l'entorn associats als pixels inferior esquerra i dret. En l'estimació de la posició, s'està assumint que els obstacles es troben sobre el terra i que el pendent de l'àrea del terra, que es troba entre la càmera i l'obstacle, és constant.

En condicions reals d'operació, hi ha dues fonts d'error a l'hora de determinar, a partir de la posició de la regió a la imatge, la distància a la que es troba l'objecte. La primera és l'error de calibració, i la segona són el moviment de la càmera degut a les vibracions del vehicle i les irregularitats de la superfície del camí. Per resoldre aquests problemes, es poden utilitzar diferents tècniques, com ara les presentades a [Amat 86] o [Huber 94], que permeten calcular la distància càmera-obstacle a partir de l'anàlisi d'una seqüència d'imatges. La tècnica proposada a [Huber 94], que es presenta a l'annex A.2, permet determinar la distància a la que es troba l'objecte a partir de l'anàlisi de la variació a les imatges de dos punts de l'objecte, amb un error relatiu del 1%. Els dos punts de l'objecte es caracteritzen per no presentar un moviment relatiu entre ells, i per ser equidistants a la càmera al llarg de la seqüència.

La figura 5.6 mostra l'error mig i l'error màxim en l'estimació de l'alçada dels objectes O1 a O6 independentment de la distància a la que es trobin². L'estimació de l'alçada

² Els valor que es presenten a la figura 5.6 s'ha obtingut mitjançant l'anàlisi de 15 seqüències d'imatges similars a la figura 5.4. En aquestes 15 seqüències similars es consideren tres alçades similars pels objectes O1-O6 (O1: 0, 1.5 i 3.5 cm., O2: 10, 8.5 i 6.5 cm., O3: 20, 18.5 i 16.5 cm., O4: 40, 38.5 i 36.5 cm., O5: 60, 58.5 i 56.5 cm., i O6: 80, 78.5 i 76.5 cm. d'alçada) i 5 posicions inicials dels obstacles

dels objectes s'ha realitzat amb un valor d'*increment_alçada* de 5 centímetres. A la figura 5.7 es mostra l'error mig en l'estimació de l'alçada, independentment de l'alçada real dels objectes analitzats, en funció de la distància càmera-objecte. Com es pot observar, no existeix una relació entre l'alçada de l'objecte i l'error en l'estimació de l'alçada. En canvi, la figura 5.7 mostra com l'error en l'estimació de l'alçada té tendència a disminuir amb la disminució de la distància entre càmera i objecte.

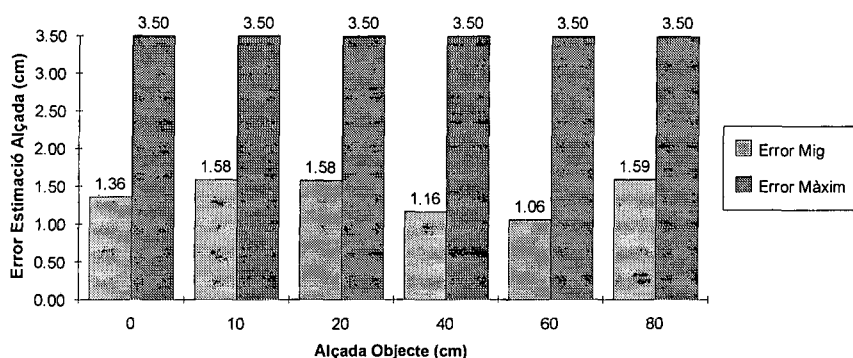


Figura 5.6. Error mig i màxim en l'estimació de l'alçada d'objectes de diferents alçades, independentment de la distància càmera-objecte.

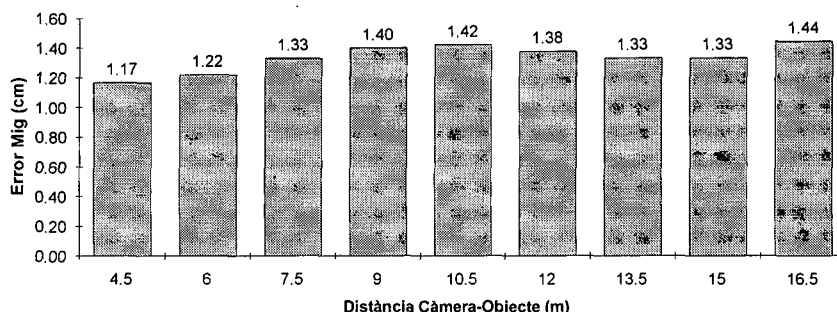


Figura 5.7. Error mig en l'estimació de l'alçada en funció de la distància càmera-objecte en les 15 seqüències d'imatges similar a la de la figura 5.4.

L'error que es comet en l'estimació de l'alçada dels objectes és degut principalment a la resolució amb la que s'observa el món (veure secció 3.1) i a la discretització dels possibles valors de l'alçada. Les característiques tridimensionals dels objectes-hipotètics que es consideren durant el procés d'estimació es modifiquen cada 3 imatges (desplaçament relatiu de 4.5 metres), amb l'objectiu de minimitzar l'error

diferents (distància obstacles-càmera en la imatge inicial de 18, 17.97, 17.94, 17.91 i 17.88 metres). L'estimació de l'alçada dels objectes en les 15 seqüències permet calcular l'error mig i l'error màxim del mètode proposat en diferents condicions (diferents alçades dels objectes respecte les hipòtesis i errors de discretització de la imatge). En totes les seqüències, el desplaçament relatiu de la càmera entre imatges és de 1.5 metres i la resolució de les imatges és de 256x256 pixels.

degut a la resolució amb la que observem l'entorn. Com es evident, una disminució del paràmetre *increment_alçada* ha d'implicar, fins arribar a un límit, una disminució de l'error, però amb un increment del temps de procés necessari per l'estimació de l'alçada dels objectes de l'entorn.

El mètode d'estimació d'alçada proposat, només permet calcular l'alçada dels objectes que defineixin, a partir del color, una regió en la seqüència d'imatges. És possible que un obstacle natural tingui el mateix aspecte (i en particular el mateix color) que el camí o el marge, no podent diferenciar-los en el procés de segmentació. Per tant, per detectar aquest tipus d'obstacle és recomanable utilitzar un segon sensor que subministri una informació complementaria a la que es pot obtenir amb un sistema de visió per ordinador. Molts autors [Crowley 89] [Curran 93] proposen l'ús de sensors d'ultrasons per a la detecció d'obstacles. Aquest sensor és una bona alternativa per a la detecció d'obstacles, independentment de si el seu aspecte visual és similar al del camí i/o del marge. Malauradament, aquest sensor també presenta importants limitacions [Grandjean 93]:

1. Només poden detectar obstacles que excedeixen de la superfície del camí dintre d'un rang de distància reduït.
2. No poden detectar obstacles negatius (com forats o rases), ni obstacles penjants (com les branques d'arbres que es troben sobre el camí a una alçada del terra inferior a l'alçada del vehicle).

Els LRF poden resoldre aquest problemes, però el seu cost és avui en dia molt elevat. Una alternativa més simple a la detecció de tot tipus d'obstacle és la utilització de la tècnica de triangulació activa. Projectant un patró de llum conegut sobre l'àrea d'interès de l'escena i analitzant l'aparença del patró a la imatge, es pot obtenir informació tridimensional del objectes presents a l'àrea d'interès. Donat els problemes que es presenten quan es vol projectar un patró de llum a una gran distància, aquesta tècnica no permet detectar obstacles llunyans. Així doncs, aquest sensor no pot reemplaçar el sistema de visió per la detecció del camí i dels obstacles que es troben lluny del mòbil, requeriment que és necessari si es vol planificar una trajectòria segura i suau que pugui ser seguida pel mòbil a velocitat constant.

Per tant, per detectar tot tipus d'obstacle és necessari la utilització de dos mòduls: un principal, que detecta els obstacles distingibles per visió, utilitzant la tècnica d'alçada per moviment, i un segon mòdul basat en la projecció d'un patró de llum estructurada, que detecta els obstacles no trobats pel sistema de visió en un rang de distància reduït i que ha de ser utilitzat per aturar el vehicle quan es detecti una col·lisió imminent.

El mètode de detecció d'obstacles proposat, és una alternativa que no requereix el càlcul de l'estructura 3D de l'escena, i que utilitza tota la informació disponible, amb l'objectiu de simplificar aquest procediment. En concret el mètode assumeix que es donen els següents fets:

1. Es coneix el desplaçament que ha realitzat el mòbil entre l'adquisició de les imatges gràcies a l'odometria.
2. El color de les regions es manté constant en la seqüència d'imatges (la il·luminació ambiental varia lentament i la sensibilitat cromàtica de la càmera és constant).
3. Es considera que la presència del punt de l'objecte que defineix el pixel superior (x_m, y_s) de la regió R a la imatge, és independent de la translació i la rotació respecte l'eix y de la càmera.

Les dues primeres assumpcions ens permeten simplificar el procediment d'aparellament de les regions homòlogues en la seqüència d'imatges i la tercera ens permet estimar l'alçada màxima d'un objecte analitzant una parella de punts de l'objecte, que s'assumeix que sempre estan presents -no hi ha oclusions- en la seqüència d'imatges. Aquestes assumpcions són normals en el món real.

En l'annex B, es presenta en detall la implementació del mètode d'hipòtesi i test.

5.1.3. Estudi comparatiu de tècniques per a la detecció d'obstacles

A. Determinació de l'alçada dels objectes a partir de la geometria del procés de formació de les imatges.

En el mètode d'hipòtesi i test, l'estimació de l'alçada dels objectes es fa de forma indirecta, ja que es basa en determinar l'alçada de l'objecte analitzant quin dels objectes hipotètics presenta l'evolució de la regió hipotètica més semblant a les regions presents en la seqüència d'imatges. La comparació i el càlcul de l'error entre l'objecte i les hipòtesis es realitza en el pla imatge, comparant la diferència d'alçades entre les regions homòlogues i les regions hipotètiques.

És possible determinar l'alçada de l'objecte de forma directa a partir del coneixement del procés de formació de les imatges, de la següent forma:

Sigui O un objecte de l'entorn observat des de la càmera i R_0 la regió que genera l'objecte O a la imatge en l'instant t_0 , quan el centre òptic de la càmera està situat en la posició C_0 .

Siguin $P=(X_P, Y_P, Z_P)$ i $Q=(X_Q, Y_Q, Z_Q)$ les coordenades a l'escena dels punts associats als píxels p_0 i q_0 que delimiten la regió R_0 .

Cal remarcar, que tots els punts de l'entorn situats a la zona P' i Q' tenen com imatge els píxels p_0 i q_0 a la imatge digital (Figura 5.8). Si un punt de l'entorn té un únic píxel a la imatge (aquest procés s'anomena projecció directa), l'invers, la projecció inversa, no és cert. Per un píxel p de la imatge, es defineix la recta rp que passa pel centre òptic de la càmera C_0 i el punt central de l'àrea delimitada al pla imatge pel píxel p . La recta rp_0 és la representant de la zona P' i la recta rq_0 és la representant de la zona Q' .

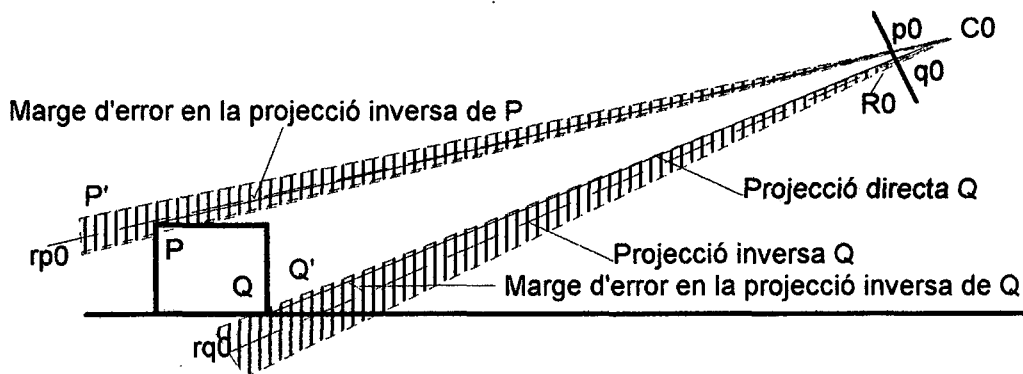


Figura 5.8. Procés de projecció directe (punt del món-píxel) i projecció inversa (píxel-zona del món associada).

Sigui R_1 la regió generada per l'objecte O en l'instant t_1 . En l'interval de temps $\Delta t=t_1-t_0$, la càmera s'ha desplaçat una distància $(\Delta X, \Delta Y, \Delta Z)$, essent la nova posició del centre òptic de la càmera $C_1=C_0+(\Delta X, \Delta Y, \Delta Z)$. Sigui p_1 i q_1 els píxels associats als punts P i Q , que delimiten la regió R_1 (Figura 5.9).

La intersecció de les rectes rp_0 i rp_1 determinen quina és la posició a l'entorn del punt P , i la intersecció de les rectes rq_0 i rq_1 determinen quina és la posició a l'entorn del punt Q (Figura 5.9). Per tant, es pot determinar l'alçada de l'objecte O com $H= Y_P-Y_Q$.

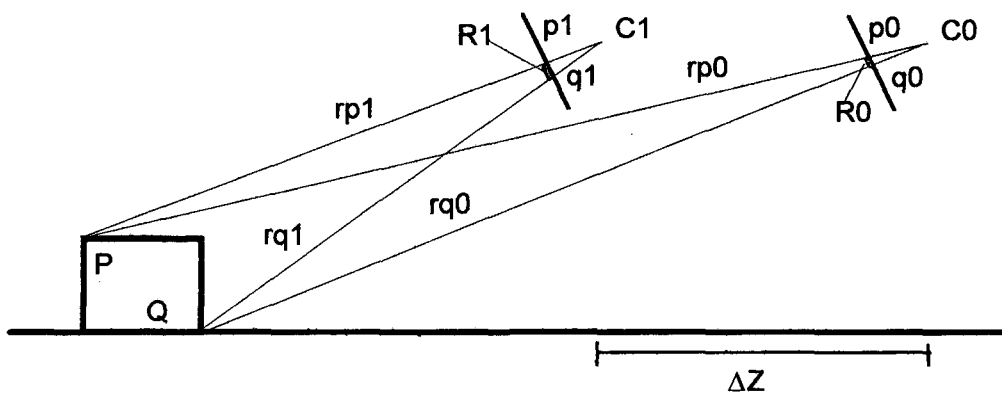


Figura 5.9. Determinació dels punts P i Q a partir de dues imatges.

L'error en el càlcul de l'alçada serà funció de l'error en la determinació del paràmetres de les rectes. Si es coneix el desplaçament relatiu de la càmera i el pla imatge té una resolució infinita (o molt gran), l'error en el càlcul de les coordenades dels punts P i Q, a partir dels 4 punts imatges considerats, tendeix a zero. Si bé es pot conèixer amb una bona precisió el valor del desplaçament, l'efecte de la discretització és clau a l'hora de determinar la direcció de les rectes, ja que la direcció de la recta rp_i i la recta que passa pels punts C_i i P no tenen per que coincidir (Figura 5.10). L'error màxim entre la direcció d'aquestes rectes dependrà de la mida del pixel al pla imatge.

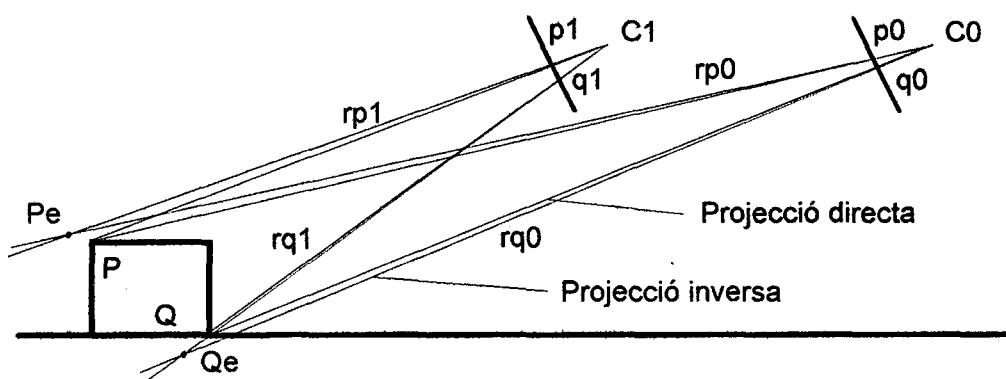


Figura 5.10. Error en el procés de determinació d'un punt P a partir d'un conjunt de punts imatge p.

Els errors en la determinació de les direccions de les rectes rp_0 , rp_1 , rq_0 i rq_1 fan que els punts on intersequen la parella de rectes (P_e i Q_e , respectivament), no coincideixi amb el punt P o Q (Figura 5.10), i per tant fan que el càlcul de l'alçada de l'objecte O sigui incorrecte ($H_e = Y_{P_e} - Y_{Q_e}$). L'error en el càlcul de l'alçada ($H - H_e$) és directament proporcional a la mida del pixel a la imatge i la distància càmera-objecte. Per minimitzar l'error degut a l'efecte de la discretització de la imatge, es pot determinar la posició dels punts P i Q no com la intersecció de dues rectes sinó com el punt on es creuen les N rectes rp_i (o rq_i), determinades a partir de N imatges de la seqüència, utilitzant el mètode de mínims quadrats. En la figura 5.11 es mostra l'error mig en el

càlcul de l'alçada dels objectes en funció de la distància³, a partir de l'estimació, amb el mètode de mínims quadrats, de la posició dels punts P i Q. Com es pot veure l'error mig decreix amb la disminució de la distància i, el conseqüent increment de rectes utilitzades per determinar la posició dels punts P i Q.

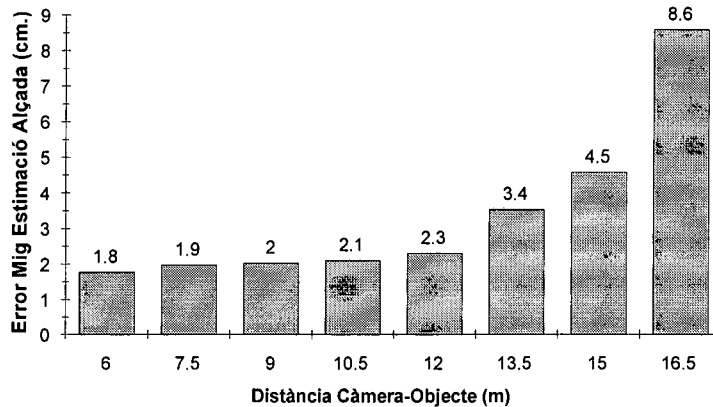


Figura 5.11. Error mig en el càlcul de l'alçada d'un objecte a partir d'una seqüència d'imatges utilitzant la geometria de la formació de la imatge.

Comparant les figures 5.11 i 5.7 es pot veure que la discretització de la imatge té un efecte més important en l'error, quan l'objecte es troba lluny de la càmera, en el cas del mètode basat en la geometria del procés de formació de les imatges, que en el mètode d'hipòtesi i test. Per contra l'error màxim en el càlcul de l'alçada, quan l'objecte es troba proper a la càmera, és inferior en el mètode basat en la geometria, que en el d'hipòtesi i test. Això és degut a que en el mètode d'hipòtesi i test, la discretització de les hipòtesis considerades sempre és causa d'error. Quan la distància càmera-objecte es redueix, l'error degut a la resolució amb la que s'han definit les hipòtesis, és més important que l'error degut a la discretització de la imatge.

B. Determinació de l'alçada dels objectes a partir de l'estereovisió.

La tècnica d'estereovisió permet calcular la distància (Z) a la que es troba un punt P a partir de la disparitat (d) entre els dos punts imatge (p_d i p_e) de $P=(X,Y,Z)$ a la imatge dreta $p_d=(x_d,y_d)$ i l'esquerra $p_e=(x_e,y_e)$, a partir de l'equació:

$$Z = \lambda \frac{b}{d} \quad \text{on } d=(x_e-x_d) \text{ i } b \text{ és la distància entre les dues càmeres.}$$

Així mateix, es poden calcular les coordenades X i Y de P a partir de la disparitat:

³ L'anàlisi de l'error que es representen a la figura s'ha fet utilitzant un sistema d'adquisició de dades i analitzant els mateixos objectes que en el cas de la figura 5.7.

$$X = \frac{x_c \cdot b}{d} \quad i \quad Y = \frac{y_c \cdot b}{d}$$

La desviació típica de l'error en el càlcul de les coordenades X, Y, Z del punt P, en funció de la distància Z, mitjançant la tècnica d'estereovisió [Aranda 97] és de:

$$\sigma\left(\frac{\varepsilon X}{Z}\right) = \sigma\left(\frac{\varepsilon Y}{Z}\right) = \sigma\left(\frac{\varepsilon Z}{Z}\right) = \frac{0.408}{d}$$

La desviació típica de l'error és inversament proporcional a la disparitat, ja que com més gran sigui aquesta, l'error degut a la discretització de la imatge té un efecte menor.

La figura 5.12 mostra quin és l'error mig en el càlcul de l'alçada d'un objecte on la posició al món dels punts P i Q s'han determinat amb la tècnica d'estereovisió⁴. L'error per una distància entre càmeres de 0.5 metres és de 4.2, 14.0, 31.6 i 56.1 centímetres a distàncies de 5, 10, 15 i 20 metres respectivament. Un augment de la distància entre càmeres incrementa linealment la disparitat, i per tant disminueix linealment l'error. En els vehicles autònoms equipats d'un sistema d'estereovisió, la distància entre càmeres està limitada per les dimensions del vehicle, pel camp de visió necessari i pel marge de mesura desitjat.

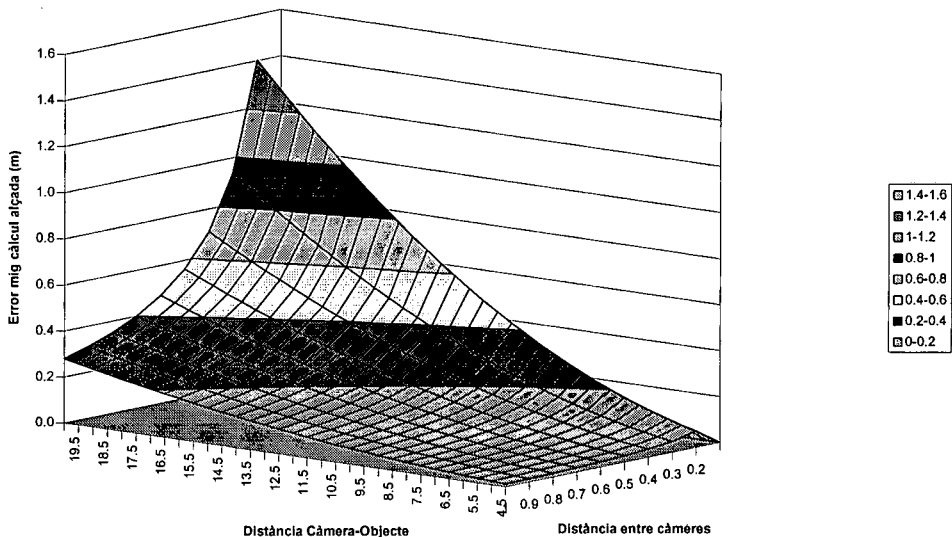


Figura 5.12. Error mig en el càlcul de l'alçada utilitzant la tècnica d'estereovisió, en funció de la distància càmeres-objecte i la distància entre càmeres.

⁴ Els valors presentats en la figura 5.12 s'han obtingut fixant una distància focal de 20 (el sistema d'adquisició d'imatge és similar al utilitzat per generar les dades de les figures 5.7 i 5.11).

C. Conclusions

La comparació de les figures 5.7, 5.11 i 5.12, ens mostra que l'error mig en el càlcul de l'alçada d'un objecte, és inferior utilitzant la tècnica d'hipòtesi i test que es proposa, que amb l'anàlisi geomètric de la imatge o la tècnica d'estereovisió. L'error mig, en el mètode d'hipòtesi i test, és particularment inferior quan l'objecte es troba a una certa distància del vehicle. Pel fet que per evitar un obstacle és necessari detectar-lo quan aquest es troba distant del vehicle, la tècnica d'hipòtesi i test és la que millor permet calcular l'alçada d'un objecte en aquestes condicions. Donat l'elevat cost de les altres tècniques per a l'anàlisi de l'estructura tridimensional de l'escena presentades a la secció 5.1.1 (triangulació activa, profunditat per enfocament i telemetria làser), no s'ha realitzat l'estudi comparatiu d'aquestes tècniques amb la de hipòtesi i test.

Com s'ha vist, l'estimació de l'alçada dels objectes, mitjançant el mètode d'hipòtesi i test, resulta més eficient que els mètodes genèrics presentats a l'inici d'aquesta secció, ja que permet calcular l'alçada únicament dels objectes que ens resulten d'interès (les regions que han estat classificades com obstacles potencials), analitzant en cada moment una única característica, l'alçada de la regió. La precisió dels resultats obtinguts són adequats per a la tasca que es vol resoldre -la detecció d'obstacles- i comparables amb les dels altres mètodes (error mig al voltant de 1 cm. i error màxim de 3.5 cm. si es treballa amb un valor d'*increment_alçada* de 5 cm.). La simplicitat - cost computacional- del mètode presentat s'assoleix gràcies al coneixement que es té del desplaçament relatiu del mòbil mitjançant l'odometria, i de l'escena que s'analitza, així com el fet que només es calcula una estimació de l'alçada dels objectes.

5.2. Descripció de l'Entorn

La modelització del món és un procés clau de tot sistema informàtic dissenyat per interaccionar de forma "intel·ligent" amb el seu entorn. El model del món ha de contenir tota la informació necessària de l'estat del món en el que es troba el sistema per poder-ne deduir les accions que cal realitzar, en vistes d'assolir l'objectiu desitjat. Per limitacions pràctiques, la informació que conté un model està molt lligada a la tasca que ha de desenvolupar el sistema i a les pròpies característiques del món, algunes de les quals s'assumeix que es coneixen a priori.

Quan la modelització es fa de forma qualitativa o poc precisa, s'acostuma a parlar de descripció. Així mateix, quan l'àmbit d'operació és il·limitat i en un moment donat el sistema només té informació d'una zona del món, que va variant conforme el mòbil es desplaça, s'acostuma a dir que el sistema manipula un model o descripció de l'entorn en el que es troba.

En el cas dels sistemes de navegació autònoma, el model o la descripció de l'entorn es realitza a partir de les dades subministrades pels sensors, que s'utilitzen per crear i/o modificar una representació interna del món o entorn. El cercle percepció/descripció/planificació/acció s'acaba tancant quan el sistema de navegació, tenint en compte l'estat actual del món, pren la decisió sobre les accions a realitzar, i controla els actuadors del vehicle per desplaçar-se en el món. Així doncs, es pot veure el model del món com l'enllaç entre les seccions de percepció del món i la de control del sistema de navegació.

La complexitat de la descripció en les sistemes de navegació autònoms, està lligada a les característiques de l'entorn (forma i propietats dels elements/objectes que formen el món) i del propi mòbil (dimensions, maniobrabilitat, velocitat, tipus de sensor, etc.), les tasques que ha de realitzar el sistema (si el vehicle o robot mòbil a més de desplaçar-se ha d'interactuar/manipular els elements presents al món) i al tipus de planificació que es realitza. Pel guiat d'un vehicle per camins mal o feblement estructurats, una descripció que indiqui on es troben els límits del camí i, si n'hi ha, la posició dels obstacles, resulta suficient.

5.2.1. Tècniques de Descripció de l'Entorn en Sistemes de Navegació Autònoms

Els aspectes més importants a tenir en compte a l'hora de dissenyar el model utilitzat per un sistema autònom són [Allen 87] :

- Facilitat de generació a partir de la informació sensorial
- Especificació implícita de les característiques del món
- Possibilitat de representar situacions complexes

Els tipus de representació més utilitzats per a la modelització del món, en els sistemes de navegació autònoma són [Angelopoulou 92]:

- **Models geomètrics.** En aquesta tècnica, la representació del món es fa mitjançant un conjunt d'objectes simples que són combinats fent servir operacions booleanes de forma similar a com es fa en la geometria constructiva de sòlids (CSG) [Goldstein 87] [De Saint Vicent 87]. Aquest tipus de representació permet descriure amb detall la forma dels objectes del món.

Una versió més simple és proposada en [Crowley 89], on el món es descriu, a partir de dades subministrades per sensors d'ultrasons, amb segments rectilinis. El model generat té en compte les imprecisions i el soroll presents en aquests tipus de dades.

- **Models basats en grafs.** En [Rao 87] s'utilitza el concepte de visibilitat per modelitzar el món. Els vèrtexs del graf són els obstacles i dos vèrtexs estan connectats amb una aresta si la visibilitat entre ells no està obstruïda. Conforme el vehicle es mou en el món, es va completant el graf de visibilitat cada vegada que arriba a un nou vèrtex.

L'anomenat model topogràfic proposat per [Kuipers 88], està inspirat en la manera en que els humans memoritzem el món. En aquest model, cada vèrtex correspon a un punt o lloc distingible del món, i els punts distingibles pròxims estan units amb una aresta. Per a cada punt distintiu, s'indica quin és el conjunt de característiques que el fan rellevant en el seu entorn.

- **Models basats en matrius.** En aquest cas la representació es basa en una matriu multidimensional, on en cada casella s'indica una certa propietat de l'àrea o volum del món associat a la casella. La informació pot ser l'ocupació [Matthies 88] [Elfes 89], les característiques bàsiques que apareixen a una zona del món (vèrtex, segments rectilini, ...) [Ron-Tabak 89] o quins elements del món són visibles des de cada punt [Kuc 87].

En [Asada 88] es presenta un model jeràrquic consistent en tres tipus de matrius (mapes): els mapes sensorial, el mapa local i el mapa global. Mentre el vehicle es mou, per a cada sensor i en un instant de temps donat, es construeix el seu mapa, que no és res més que les dades subministrades pel sensor representades en el sistema de referència de cada sensor. El mapa local es genera a partir dels diferents mapes sensorial que han estat transformats al sistema de coordenades del vehicle. Finalment, tots els mapes locals -cada un representa una zona del món diferent- són integrats per a obtenir el mapa global, que representa tota l'àrea percebuda durant el recorregut del vehicle.

La majoria de les tècniques per a la modelització del món descrites anteriorment, han estat pensades per a una aplicació concreta. Malgrat això, es poden establir uns criteris comuns amb l'objectiu de comparar-les [Angelopoulou 92], com són: la precisió de la representació, els requeriments de memòria, el temps necessari per a crear/modificar el model, l'expandibilitat i la necessitat de coneixement a priori per la seva construcció i/o utilització.

El tipus de model més precís i expandible (pot representar qualsevol tipus de situació) és el geomètric, però per contra és el que requereix més temps per ser generat. Els models basats en graf, requereixen d'un coneixement a priori, i la seva generació pot resultar molt lenta. Finalment els models basats en matrius, requereixen de molta memòria per a descriure el món amb precisió i no són expandibles -cada casella conté un tipus d'informació definida en el moment del disseny del sistema-; per contra resulten els més ràpids de generar i/o actualitzar.

5.2.2. Descripció d'un Camí Feblement Estructurat per a la Navegació Autònoma

La informació mínima que ha d'incloure una descripció que ha de ser utilitzada per navegar autònomament en un camí mal o feblement estructurat, és la posició al món dels límits del camí i la posició dels possibles obstacles. D'aquesta forma s'indica quin és l'espai navegable, ja que en aquest tipus d'entorns, encara que els marges del camí siguin plans no resulten el millor espai per desplaçar-se.

La informació que ha de contenir una descripció en el cas dels sistemes de navegació autònoms per a carreteres i autopistes, és similar. En la major part d'aquests sistemes (veure secció 2.2) s'opta per descriure la carretera mitjançant un model geomètric on s'indica bàsicament la posició i la curvatura de la carretera. Els límits d'una carretera acostumen a ser clars i estan definits seguint unes restriccions geomètriques. A més, els models geomètrics resulten eficient per realitzar el control posterior del vehicle, ja que l'angle de gir de la direcció del vehicle ha de coincidir amb la curvatura de la carretera. Per tant, aquest tipus de descripció resulta adequada per a la modelització de carreteres i autopistes (entorns exteriors estructurats).

Donat que els límits dels camins feblement estructurats són poc clars, ambigus i força irregulars, resulta bastant costós obtenir-ne una descripció geomètrica. Si a més es té

en compte, que l'estratègia a seguir davant d'un obstacle és canviar la direcció del vehicle per evitar-lo, sense seguir necessàriament la curvatura del camí, i que el resultat obtingut de l'anàlisi de la informació subministrada pel sistema sensorial són quines regions de la imatge són transitables (camí) i quines no (obstacles o marge), resulta més pràctic i ràpid d'obtenir una descripció de l'entorn basada en una matriu. Cada casella de la matriu pot ser utilitzada per indicar si una àrea donada de l'entorn és travessable o no. La modelització de l'entorn en estructures basades en matrius, és la més utilitzada pels sistemes de navegació autònoms en entorns no estructurats (veure secció 2.3).

Com s'ha vist, la segmentació de la imatge en regions i la seva posterior classificació, junt amb l'anàlisi de l'evolució de l'alçada de les regions en la seqüència d'imatges, ens permet obtenir una primera aproximació de quines àrees de l'entorn són transitables.

El fet que els límits dels camins mal o feblement estructurats siguin pocs clars i ambigus, i que el color del camí i de part del seu marge siguin semblants, fa que les etapes de segmentació i classificació no aconseguixin localitzar de forma fiable la posició del camí a la imatge (veure figura 4.11). És per aquesta raó que resulta necessari integrar la informació de la seqüència d'imatges i imposar les restriccions incloses al model genèric de camí transitable que s'ha definit (en concret les restriccions de direcció i amplada del camí localment constant), per determinar la situació del camí a l'entorn d'una forma més fiable. A partir de la integració de la informació de la seqüència d'imatges i del posicionament de camí, els marges i els obstacles respecte un sistema de coordenades al món tridimensional es generarà una **descripció de l'entorn**, en el que s'indica quin és l'espai transitable, que permetrà a una posterior etapa de planificació decidir quina trajectòria ha de seguir el mòbil.

El passos que es realitzen per generar la descripció de l'entorn són (figura 5.13):

1. Generació de la **descripció de la imatge**. En aquest pas es posicionen en l'espai 3D les regions de la imatge classificada.
2. Actualització i etiquetat de les caselles de la descripció de l'entorn, a partir de la **integració de les últimes descripcions de la imatge** de la seqüència i dels obstacles detectats.
3. Imposició en la descripció de l'entorn actual de les **restriccions del model genèric de camí mal o feblement estructurat** (amplada i direcció localment constants).

Cada un d'aquests passos requereix de la realització de varis processos que es presenten a continuació.

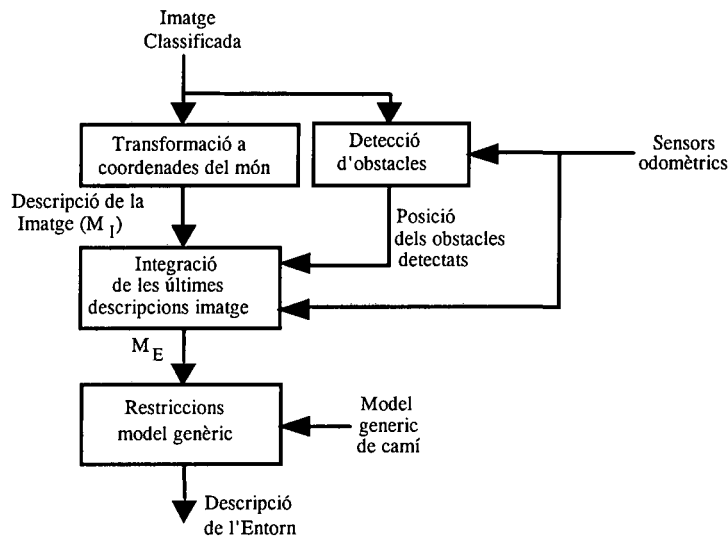


Figura 5.13. Etapes a realitzar per descriure l'entorn en la metodologia proposada.

La **descripció de la imatge** s'obté posicionant al sistema de referència del món les regions de la imatge que han estat classificades com camí, obstacle potencial o marge. Per a cada imatge de la seqüència analitzada s'en obté una descripció. Combinant la informació continguda en N descripcions de la imatge consecutives obtingudes mentre el vehicle es desplaça i els obstacles detectats, es construeix finalment la descripció de l'entorn.

Tant la descripció de la imatge com la descripció de l'entorn, es representen en una matriu bidimensional, que anomenarem M_I i M_E respectivament, on en cada casella (i,j) , s'indica la característica -camí, marge o obstacle- de l'àrea de l'entorn de mida Resolució_X*Resolució_Y que es troba a la posició $(X_{inici}+i* Resolució_X, Y_{inici}+j* Resolució_Y)$ respecte el sistema de referència del món⁵ (figura 5.14). Així doncs, cada casella (i,j) de la descripció té associada una àrea de l'entorn i una zona de la imatge que s'ha captat d'aquest entorn. A la figura 5.14 es pot veure la relació existent entre la matriu M_I (la descripció de la imatge), l'entorn i la imatge captada de l'entorn.

Si les dimensions de la matriu M_E és de M files i N columnes, la descripció de l'entorn indicarà les característiques de l'àrea de mida Resolució_X*N d'ample i Resolució_Y*M de llargada que es troba davant del punt de l'entorn on el vehicle ha pres l'última imatge.

⁵ (X_{inici}, Y_{inici}) és la posició al món de la casella (0,0) de la descripció.

Com és ben conegut, tots els punts de l'espai 3D que es troben sobre la recta que passa pel centre òptic de la càmera i un píxel donat, tenen la mateixa projecció a la imatge digital. Donat un píxel (x,y) es pot conèixer la posició del punt (X,Y,Z) associat a l'espai 3D que es projecta en el píxel, si es coneix almenys un d'aquests valors X , Y o Z (veure annex A). El mètode que s'utilitza pel càlcul de la posició al món tridimensional del píxel (x,y) de la imatge es basa en l'assumpció que el pendent del terra és constant en el tram de camí observat, i per tant el valor Z dels punts que s'han de posicionar, són coneguts -la Z del límit de camí és 0 i la dels obstacles es calcula mitjançant la tècnica d'alçada a partir de moviment-. Les equacions A.6 i A.7 permeten determinar els valors X i Y de la posició a l'espai 3D del punt associat al píxel (x,y) , si es coneix el valor de la coordenada Z del punt i els paràmetres del sistema d'adquisició. Al capítol 6 s'avaluarà els efectes que la infracció de l'assumpció de pendent constant provoca en la descripció de l'entorn.

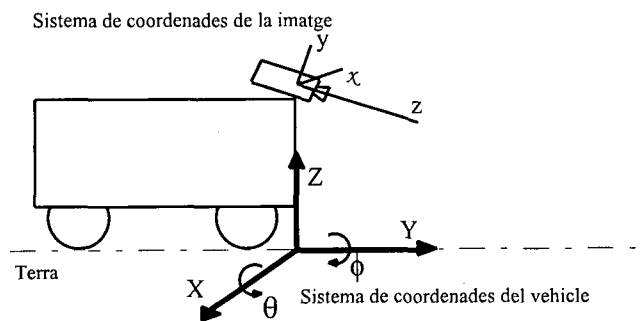


Figura 5.15. Relació entre els sistemes de coordenades del món-vehicle i el de la càmera.

Per determinar a quina casella (i,j) de la descripció està associat un píxel (x,y) de la imatge, primer s'ha de calcular la seva posició (X,Y,Z) al món -equacions A.6 i A.7- i posteriorment, tenint en compte la resolució de la descripció, la casella es pot determinar fent:

$$i = \frac{X - X_{inici}}{Resolucio_X} \quad (5.1)$$

$$j = \frac{Y - Y_{inici}}{Resolucio_Y} \quad (5.2)$$

Totes les caselles de la descripció de la imatge (M_1) tenen associat un conjunt de tres comptadors que s'utilitzen per indicar el nombre de píxels associats a la casella que tenen categoria camí/camí ampliat, obstacle potencial o marge en la imatge classificada (figura 5.16). Per a cada píxel (x,y) de la imatge classificada amb categoria v que es projecta a la casella (i,j) , s'incrementa el valor del comptador v de la casella (i,j) . Una vegada finalitzat aquest procés,

es coneix el nombre de pixels de la imatge classificada amb etiquetes camí, marge i obstacle potencial, que es projecten a la casella (i,j) de M_1 .

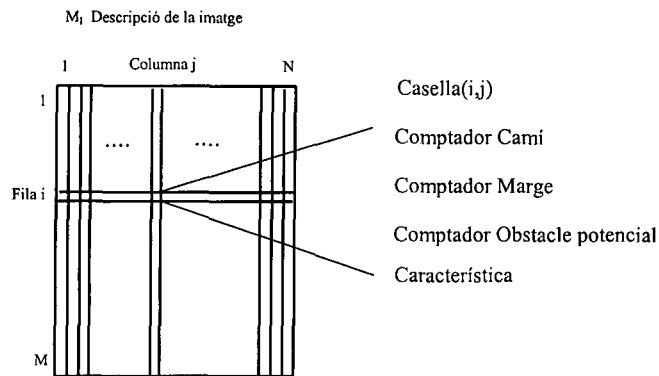


Figura 5.16. Informació continguda en la descripció de la imatge

- Etiquetar les caselles de la descripció de la imatge. Cada casella de M_1 es classificarà en **camí**, **marge**, **obstacle potencial** o **desconegut**, en funció del nombre de pixels associats a la casella de cada categoria.

Si l'àrea de l'entorn associada a la casella (i,j) és d'un mateix tipus, per exemple camí, tots els comptadors valdran zero excepte el comptador camí, i per tant la casella (i,j) es classificarà com camí. Si l'àrea de l'entorn associada a una casella (i,j) està formada per diferent tipus de zones, per exemple camí i marge, el valor d'aquests dos comptadors serà diferent de zero i el valor del comptador obstacle potencial serà zero. La classificació de la casella (i,j) de M_1 es realitza a partir del valor dels tres comptadors de la següent forma:

- Si cap pixel de la imatge es projecta sobre la casella (i,j) , tots els comptadors valen zero, la casella (i,j) de M_1 es classifica com **desconeguda**.
- Si el nombre de pixels de la casella (i,j) amb categoria marge és més gran del $V_1\%$ del total de pixels associats a la casella, la casella (i,j) de M_1 es classifica com **marge**.
- Si el nombre de pixels de la casella (i,j) amb categories marge o obstacle és més gran del $V_2\%$ del total, la casella (i,j) de M_1 es classifica com **obstacle potencial**.
- En els altres casos la casella (i,j) de M_1 es classifica com **camí**.

Per a la determinació dels percentatges $V_1\%$ i $V_2\%$ que s'utilitzen a l'hora de classificar les caselles de M_1 , s'ha de tenir en compte que una vegada classificada la casella amb l'etiqueta corresponent, s'indica que tota l'àrea de

l'entorn associada a la casella és del mateix tipus. Amb l'obtenció de la descripció de la imatge a més de realitzar un canvi de coordenades (de coordenades imatge a coordenades món) també es produeix una pèrdua de resolució espacial, ja que es passa d'una matriu (la imatge) de dimensions grans a una de dimensió $M \times N$ més petita (M_1 , la descripció de la imatge).

Per tal de garantir la seguretat durant la conducció, cal ser prudent a l'hora de realitzar la classificació de les caselles de M_1 que tenen associades zones de l'entorn de diferents tipus. En aquests casos és preferible classificar la casella com no transitable (marge o obstacle) encara que una part significativa de l'àrea de l'entorn associada a la casella sigui transitable. Per aquesta raó, si una casella té més d'un $V_2\%$ dels pixels associats del tipus marge o obstacle (on V_1 i V_2 és desitjable sigui un valor petit), aquesta és classificada com no transitable, encara que pot ser que el $(100 - V_2)\%$ de l'àrea de l'entorn associada a la casella sigui camí.

Triant l'opció contrària (prendre valors de V_1 i V_2 alts) hi ha una tendència a classificar com transitable caselles associades a àrees de l'entorn que no són completament transitables. En el pitjor cas el sistema de planificació de la trajectòria podria fer passar el vehicle per l'àrea de l'entorn associada a la casella que ha estat classificada com a camí, amb el perill que en passar per aquella zona (on hi ha part que és marge i/o obstacle), el mòbil surti del camí o col·lisió amb l'obstacle que hi pogués haver. En l'altre extrem, tampoc resulta adequat que el percentatge mínim per classificar una casella com a no transitable sigui molt petit, per exemple V_1 i $V_2=1\%$, ja que en aquest cas els marges del camí i la mida dels obstacles tindrien una clara tendència a créixer, reduint per tant la zona transitable en la descripció de l'entorn. En el pitjor cas ens podrien trobar en una situació on no fos possible, analitzant la descripció de l'entorn, definir una trajectòria segura malgrat que al camí hi hagués prou espai lliure per a que el vehicle pogués avançar.

Els valors dels percentatges V_1 i V_2 utilitzats no tenen per que ser fixes, sinó que es poden modificar en funció del risc que el sistema de navegació vulgui assumir de triar una trajectòria no segura.

En la figura 5.17 es pot veure la descripció de la imatge associada a un imatge sintètica i a una imatge captada en un entorn natural⁶.

⁶ Les regions marcades com camí color 1 i 2, són, respectivament, les regions de la imatge de color igual als dos colors més freqüent a la zona del camí més propera del vehicle. Les regions marcades com

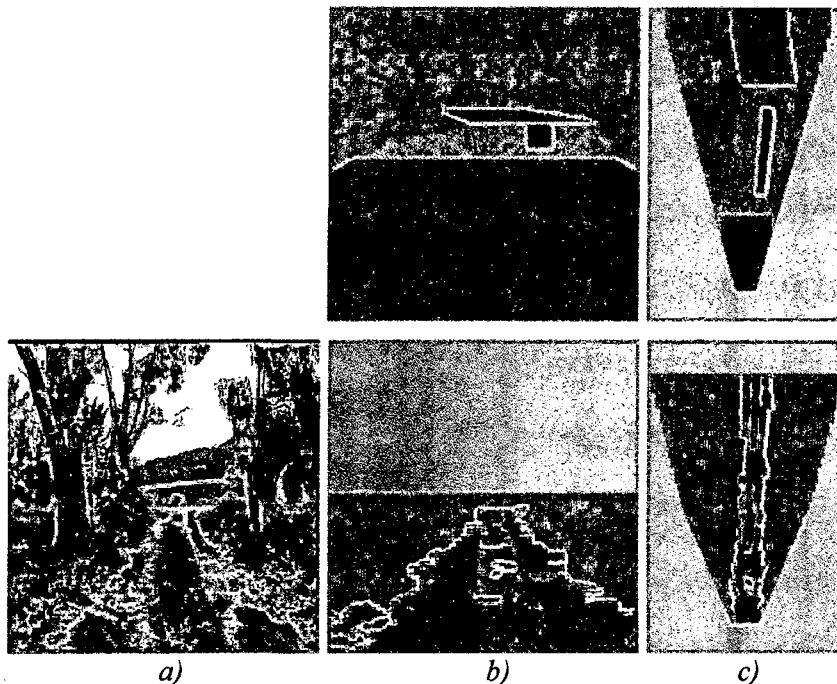
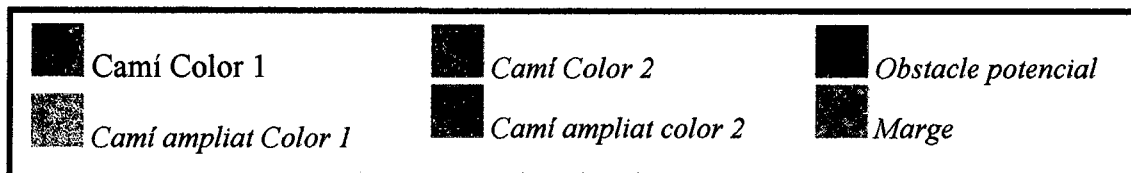


Figura 5.17. Relació entre la imatge i la descripció de la imatge
a) Imatge Original b) Imatge classificada c) Descripció de la imatge

En el segon pas s'integra la informació continguda en les últimes descripcions de la imatge i s'incorporen els obstacles detectats, tot tenint en compte, el moviment relatiu del vehicle entre els instants on s'han capturat les imatges analitzades. Els processos que es realitzen són:

1. Actualització de la posició dels elements a la descripció de l'entorn tenint en compte el desplaçament relatiu del vehicle entre el moment que s'ha capturat l'última imatge (i per tant des de que es va actualitzar per última vegada la descripció de l'entorn) i l'instant en que s'ha capturat la imatge associada a la descripció de la imatge actual. Aquest desplaçament es conegut gràcies als sensors odomètrics.

camí ampliat color 1 i 2, són, respectivament, les regions de la imatge de color similar als dos colors més freqüent a la zona del camí més propera del vehicle

Si el vehicle ha recorregut una distància ($Desp_X$, $Desp_Y$) entre l'adquisició de les dues últimes imatges, cal desplaçar totes les caselles de M_E $Desp_X/Resolució_X$ columnes i $Desp_Y/Resolució_Y$ files, per actualitzar la posició dels elements de la descripció de l'entorn a la nova posició del vehicle a l'entorn.

2. Actualitzar la descripció de l'entorn incorporant-hi la informació de la descripció de la imatge. A cada casella (i,j) de M_E s'afegeix la categoria de la casella (i,j) de M_I . En cada casella de M_E es memoritza la classificació de les caselles associades d'un nombre determinat de descripcions de la imatge (figura 5.15). Per tant, en cada casella de M_E es memoritza com s'ha classificat l'àrea de superfície $Resolució_X * Resolució_Y$ associada a la casella, que ha estat captada i analitzada a diferents distàncies conforme el vehicle s'ha desplaçat.

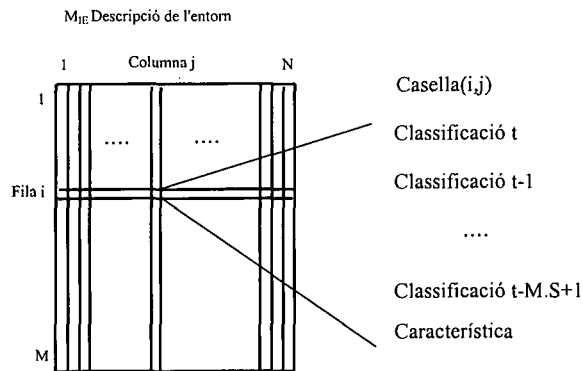


Figura 5.18. Informació continguda en la descripció de l'entorn

3. Classificar cada casella (i,j) de la descripció de l'entorn en camí, marge, obstacle o desconegut, en funció dels valors memoritzats en la casella (i,j) de M_E .

Es classifica cada casella de la descripció de l'entorn tenint en compte els valors memoritzats a la casella, i tenint en compte el fet que contra més recent és la informació, més rellevant és. Anomenem **memòria del sistema** (M.S) al nombre d'imatges utilitzades per generar la descripció de l'entorn, o el que és el mateix, el nombre de valors que s'han memoritzat en cada casella de M_E .

El nombre màxim de caselles a memoritzar està determinat pel marge de mesura del sensor (veure secció 3.1), i pel desplaçament que realitza el vehicle (en mitja) entre la captació de dues imatges consecutives:

$$M.S_{max} = \frac{Y_{max} - Y_{min}}{Desp_Y} \quad (5.3)$$

El nombre mínim per M.S és 1, és a dir, la descripció de l'entorn es genera utilitzant únicament l'última descripció de la imatge i la informació subministrada pel mòdul de detecció obstacles. Si M.S=1, la descripció de la imatge i de l'entorn són pràcticament idèntiques.

Com més gran sigui el valor de M.S, dintre dels límits possibles, es disposarà de més informació per detectar i eliminar els possibles errors que s'han generat en les anteriors etapes del processat (per exemple, que part del marge hagi estat classificat com a camí ja sigui per la similitud de color entre camí i marge, o per un error en el processament de les imatges captades). Per contra, si es té en compte moltes imatges de la seqüència a l'hora de generar la descripció de l'entorn, es requereix de més memòria i temps de procés per generar la descripció de l'entorn, així com hi haurà una tendència a reflectir amb retard en la descripció de l'entorn, els canvis que es produeixen en el món.

Si el valor de M.S és petit, s'estarà davant de la situació oposada a la descrita anteriorment, es necessita menys memòria i temps de procés per obtenir la descripció de l'entorn, que serà generada únicament amb informació recent. Per contra, els possibles errors que apareixin en una descripció de la imatge, tindran tendència a aparèixer en la descripció de l'entorn.

Per tant, l'elecció del valor de la memòria del sistema, és una decisió de disseny que s'haurà de prendre en funció de quin dels efectes es vol maximitzar o eliminar. Els valors intermitjos de M.S, són amb als que s'obté un compromís entre els efectes positius i els negatius descrits anteriorment.

La classificació en camí, obstacle potencial, marge o desconegut de les caselles de M_E es pot realitzar ponderant els M.S valors memoritzats a cada casella. A l'última descripció de la imatge (M_{It} , on t indica el temps en que s'ha generat la descripció) se li dóna pes P_0 , a la penúltima descripció de la imatge (M_{It-1}) pes P_1 , a l'antepenúltima descripció de la imatge (M_{It-2}) pes P_2 i així successivament fins la descripció de la imatge més antigues ($M_{It-M.S-1}$) que es ponderada amb pes $P_{M.S-1}$. Amb l'objectiu de donar més importància a la informació més recent, el valor dels pesos han de complir la relació $P_0 \geq P_1 \geq \dots \geq P_{M.S-1}$. Així mateix, si no es vol que un error puntual en una descripció de la imatge arribi a ser present en la descripció de l'entorn, cal que el valor d'un pes P_k sigui menor que la suma de tots els altres pesos, és a

dir, $P_k < \sum P_l \forall l \neq k$. Com resultat de la ponderació de les M.S últimes classificacions de la descripció de la imatge, per a cada casella de la descripció de l'entorn s'obtenen, a l'instant t, 4 valors:

$$ponderacio_cami(i,j) = \sum_{l=0}^{M.S-1} P_l * [M_{t-l}(i,j,caracteristica) = cami] \quad (5.4)$$

$$ponderacio_marge(i,j) = \sum_{l=0}^{M.S-1} P_l * [M_{t-l}(i,j,caracteristica) = marge] \quad (5.5)$$

$$ponderacio_obs.potencial(i,j) = \sum_{l=0}^{M.S-1} P_l * [M_{t-l}(i,j,caract.) = obs.potencial] \quad (5.6)$$

$$ponderacio_desconegut(i,j) = \sum_{l=0}^{M.S-1} P_l * [M_{t-l}(i,j,caracteristica) = desconegut] \quad (5.7)$$

$$\text{on } [M_{t-l}(i,j,caracteristica) = C] = \begin{cases} 0 & \text{Si } caracteristica \neq C \\ 1 & \text{Si } caracteristica = C \end{cases}$$

Els quatre valors són una ponderació (en la que es té en compte l'actualitat i resolució de les dades) que depenen de com ha estat classificada la casella (i,j) en les descripcions de la imatge considerades. La casella (i,j) de la descripció de l'entorn és classifica en l'instant t com a camí, obstacle, marge o desconegut, segons quin sigui el valor màxim dels 4.

$$M_E(i,j,caracteristica) = \text{Màxim}(ponderacio_C)$$

$$\text{on } C = \{\text{camí, marge, obstacle potencial, desconegut}\}$$

4. Incorporació a la descripció de l'entorn de la informació subministrada pel mòdul de detecció d'obstacles.

El mòdul de detecció d'obstacle indica la posició i alçada de l'obstacle, que independentment de la seva forma es representa com el seu paral·lel·lepid envoltant. Les caselles de característica obstacle potencial que estan associades a un objecte d'alçada inferior al llindar H, són reclassificades amb la característica camí. Les caselles de característica obstacle potencial que estan associades a un objecte d'alçada superior o igual al llindar H, són reclassificades amb la característica obstacle.

Per a les caselles associades a un objecte d'alçada igual o superior a H, cal tenir en compte que a l'hora de obtenir la descripció de la imatge, aquests objectes s'han projectat al món assumint que la seva alçada era nul·la. Aquest fet fa que els obstacles ocupin una àrea superior de la que li correspon, com es pot veure a la figura 5.16 (la ombra de l'obstacle, àrea de l'entorn no visible per quedar amagada per l'obstacle, també s'ha classificat com obstacle potencial). És en aquest moment, en el que es coneix les dimensions de l'obstacle, que es classifiquen com obstacle aquelles caselles de la descripció

de l'entorn que estan total o parcialment ocupades per l'obstacle, i les caselles associades a l'ombra de l'obstacle, que fins ara estaven classificades com obstacle potencial, són reclassificades com camí si l'obstacle es troba a l'interior del camí o com marge si l'obstacle és contigu amb al marge del camí.

El resultat del procés d'integració de la seqüència genera una matriu de dimensió $M \times N$ on en cada casella s'indica la característica (camí, marge, obstacle o desconegut) de l'àrea de l'entorn associada.

En el tercer pas, s'imposa al resultat de la classificació de M_E les restriccions d'amplada i direcció localment constants o quasi constants, tal com indica el model genèric del que ha de ser un camí transitable (veure secció 3.1). Per assolir aquest objectiu es canviaran, si cal, la característica del menor nombre de caselles de M_E .

En aquest pas, es verifica que l'estructura representada a la descripció de l'entorn presenta les característiques d'un camí transitable definides al model genèric de camí mal o feblement estructurat. Cada fila j de M_E és recorreguda per localitzar les caselles inicial i final, (i_i, j) i (i_f, j) respectivament, que han estat marcades com camí. D'aquests dos valors es pot deduir l'amplada del camí per la fila j , $Amplada_camij = i_f - i_i$. Donades dues files consecutives, j i $j+1$, verifiquem:

I. Que l'amplada del camí sigui aproximadament constant.

$$Amplada_camij \cong Amplada_camij+1$$

II. Que la direcció del camí sigui aproximadament constant i els seus marges dret i esquerra paral·lel.

Les condicions $(i_{ij} - i_{ij+1} \cong i_{if} - i_{if+1})$ i $(i_{ij} - i_{ij+1} < A)$ indiquen que es compleixen les dues restriccions.

El resultat de l'aplicació de les restriccions d'amplada i direcció localment constants és la informació que serà subministrada a la etapa de planificació de la trajectòria per a que defineixi quina és la trajectòria que ha de seguir el vehicle en funció dels criteris de navegació que s'hagin fixat.

Com es veu, la descripció de l'entorn s'obté de forma incremental a partir de la informació més recent que es coneix de l'entorn. La descripció de l'entorn inicial coincideix amb la primera descripció de la imatge (inicialment M_E és igual a M_I). Conforme s'analitzen noves imatges, la descripció de l'entorn és completada i

actualitzada. L'anàlisi de la seqüència d'imatges que es realitza, a més de garantir la coherència de la descripció de l'entorn, ens permet detectar regions mal classificades en fases anteriors.

En la construcció de la descripció de l'entorn s'utilitza el coneixement a priori que es té de l'escena que és el que s'ha utilitzat per definir el model genèric de camí transitable. L'aplicació de la restricció del model genèric de camí garanteix que el camí representat a la descripció té les característiques d'un camí transitable.

La discretització de la descripció de l'entorn en caselles, cada una de les quals representa una superfície de Resolució_X*Resolució_Y centímetres del món, permet planificar una trajectòria amb pocs requeriments de temps i memòria. Un altre avantatge d'aquesta forma de representar l'entorn és la possibilitat que ofereix de treballar dinàmicament amb una resolució espacial variable. És a dir, en cas que la descripció que es manipula no permetés trobar una trajectòria segura, es pot refinar la descripció, augmentant el nombre de caselles que la formen, aconseguint així incrementar la seva resolució espacial.

En la figura 5.19 es mostra un exemple de com es genera la descripció de l'entorn a partir d'una seqüència d'imatges. La primera columna de la figura, correspon a la seqüència d'imatges sintètiques (256x256 píxels) amb un desplaçament relatiu entre les imatges de la seqüència de 1.5 metres (la imatge superior és la primera de la seqüència). Sobre la imatge inicial, s'ha superposat les fronteres de les regions de diferent categoria generades per la etapa de classificació (en blanc frontera entre camí i marge, en vermell frontera entre camí i obstacle). La segona columna correspon a la descripció de la imatge obtinguda a partir de la imatge classificada. Mitjançant la integració de la informació de la seqüència i la detecció d'obstacles es genera la descripció de l'entorn (tercera columna de la figura 5.19, on les àrees grises corresponent a caselles classificades com desconegudes, les taronges a caselles classificades com marge, les verdes a caselles classificades com camí i les vermelles corresponen a les classificades com obstacle).

La figura 5.20 mostra un exemple de generació de la descripció de l'entorn a partir d'una seqüència d'imatges captades en un camí de muntanya amb un desplaçament relatiu entre les imatges de la seqüència de 1.5 metres. En la primera columna de la figura 5.20 apareixen les imatges RGB originals sobre les que s'ha marcat les fronteres resultants de l'etapa de classificació les regions. Com es pot veure en aquest cas real, el resultat de la classificació no sempre correspon a la superfície del camí (ja que

sovint part del marge contigu al camí té el mateix color que el camí). És gràcies a l'anàlisi de la seqüència d'imatges obtingudes durant el moviment del vehicle, i a la imposició de les restriccions del model genèric de camí (amplada i direcció localment constant) que es pot delimitar amb més exactitud l'extensió de la superfície del camí. Les àrees violetes que apareixen en les imatges de la tercera columna corresponen a caselles de la descripció de l'entorn que han estat reclassificades com camí quan s'ha imposat les restriccions d'amplada i direcció localment constants. La descripció de l'entorn (tercera columna) que s'obté d'aquesta forma, indica quina és l'àrea de l'entorn travessables.

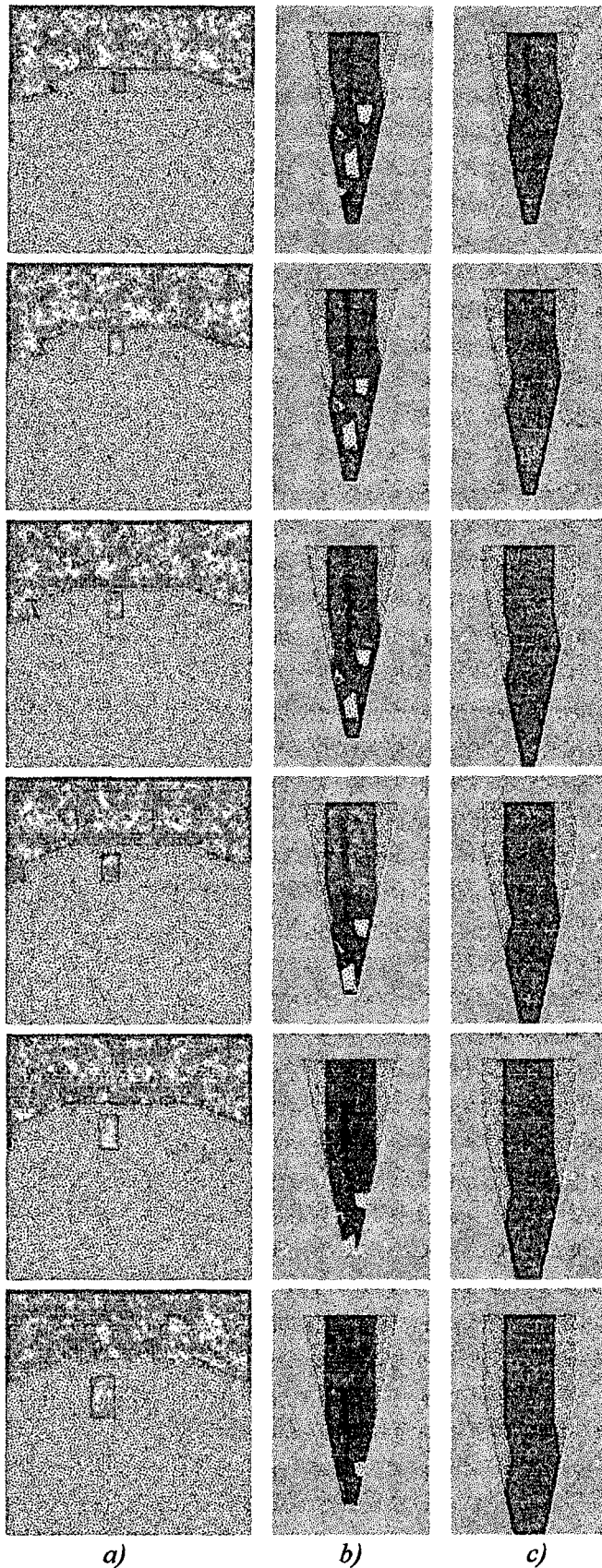
En l'annex B es presenta en detall la implementació del procés de descripció de l'entorn a partir d'una seqüència d'imatges.

5.3. Conclusions

En aquest capítol s'ha presentat la part de la metodologia que es basa en el processament de la seqüències d'imatges generades per l'etapa de classificació de regions, obtingudes mentre el vehicle es desplaça al llarg del camí. L'anàlisi de la seqüència d'imatges s'utilitza per detectar els possible obstacles que es puguin trobar a l'interior del camí, i verificar la coherència de la posició del camí i els obstacles. El resultat final d'aquests processos, és una descripció de l'entorn, que es representa mitjançant una matriu on cada casella indica la característica (entre camí, marge, obstacle o desconegut) d'una àrea de mida Resolució_X*Resolució_Y associada.

La generació de la descripció de l'entorn requereix de la realització dels següents processos:

1. Obtenció de la descripció de la imatge. Transformació de la posició de les regions al pla imatge al sistema de referència del món tridimensional (que es fa coincidir amb el sistema de referència del vehicle) a partir del coneixement de les característiques del sistema d'adquisició d'imatges. La descripció de la imatge és una matriu on cada casella està associada amb una àrea de l'entorn (i una zona de la imatge), i és classificada en funció del nombre de pixels associats a la casella amb categoria camí, obstacle potencial o marge.

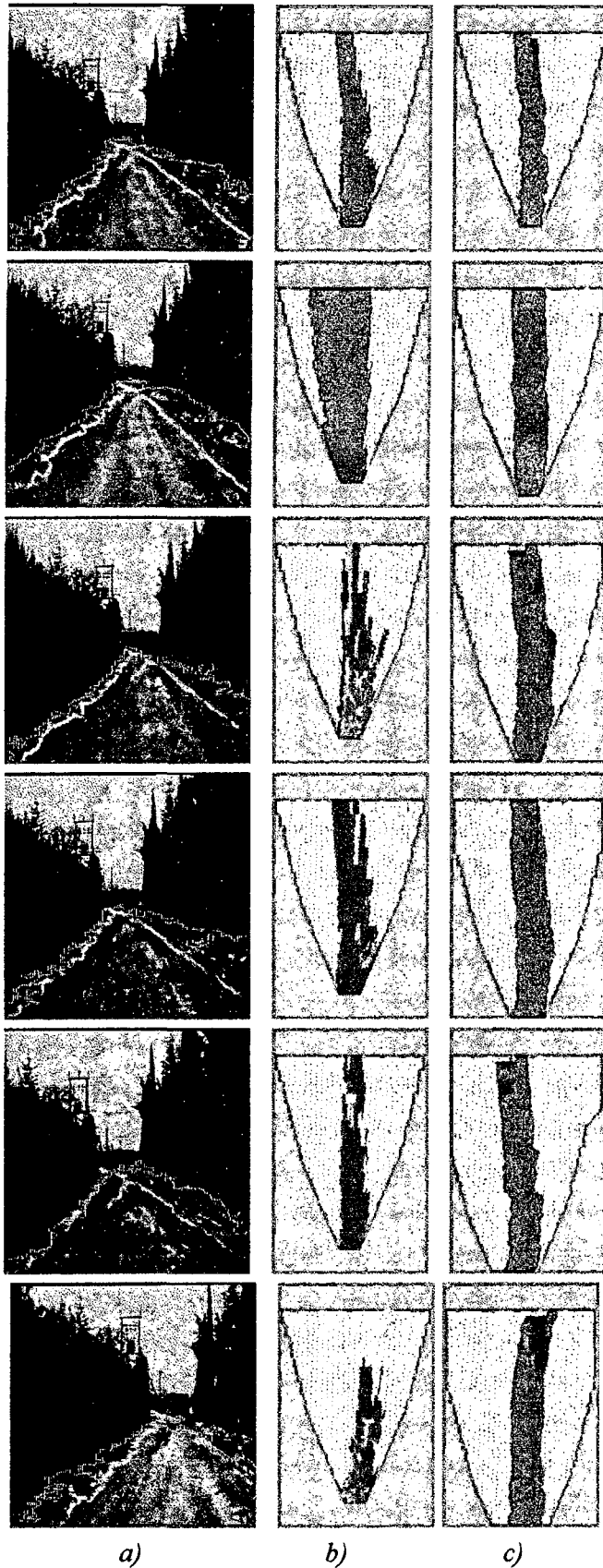


a)

b)

c)

*Figura 5.19. Exemple de generació de la descripció de l'entorn.
a) Seqüència d'imatges sintètiques amb les fronteres generades per l'etapa de classificació. b) Descripció de la imatge. c) Descripció de l'entorn després de l'anàlisi de la seqüència d'imatges fins la imatge corresponent.*



a)

b)

c)

*Figura 5.20. Exemple de generació de la descripció de l'entorn.
a) Seqüència d'imatges RGB amb les fronteres generades per l'etapa de classificació. b) Descripció de la imatge. c) Descripció de l'entorn després de l'anàlisi de la seqüència d'imatges fins la imatge corresponent.*

2. Detecció d'obstacle. Anàlisi de l'evolució de l'alçada de les regions, que han estat classificades com obstacles potencials a l'etapa de classificació de regions, al llarg de la seqüència d'imatges.
3. Integració de la informació continguda en les últimes imatges de la seqüència. Classificació de les caselles de la descripció de l'entorn, ponderant la classificació que han tingut les caselles associades de les últimes descripcions de la imatge i els obstacles detectats. En la ponderació se li dona més pes a la informació més recent.
4. Imposició de les restriccions recollides en el model genèric de camí mal o feblement estructurat, que es consideren que han de complir els camins transitables. Sobre la descripció de l'entorn generada pel pas anterior, s'imposen les restriccions d'amplada i direcció localment constants amb l'objectiu de localitzar la posició del camí en aquelles àrees on el seu límit és ambigu o esta poc definit (el color de les matèries que formen el camí i els marges són semblants).

Com es pot veure a les figura 5.19 i 5.20, la descripció de l'entorn conté la informació necessària per a que un etapa posterior de planificació i control de la trajectòria pugui definir una trajectòria segura per ser seguida pel vehicle. En el capítol 6 s'estudiaran les causes dels errors que apareixen en la descripció de l'entorn, i quin són els seus efectes en funció de les condicions en que ha de treballar la metodologia presentada. Així mateix, en el sisè capítol, s'inclouen els resultats obtinguts en la descripció de camins mal o feblement estructurats de diferents característiques.

Les principals idees i conceptes que s'han definit i utilitzat en aquesta segona part de la metodologia presentada són:

- Definició d'una tècnica de detecció d'obstacles que es basa en l'anàlisi de l'evolució de l'alçada de les regions a la seqüència d'imatges. Aquesta tècnica, que es basa en el principi d'hipòtesi i test, es troba entre les que estimen *l'alçada a partir del moviment* de la càmera, i es realitza en dos passos: definició d'un conjunt d'hipòtesis respecte l'alçada de l'objecte que ha generat la regió analitzada en la imatge inicial d'una seqüència, i verificació, en les imatges posteriors de la seqüència, de quines de les hipòtesis formulades s'ajusta més a les característiques de l'objecte analitzat. Aquesta tècnica, que estima de forma aproximada l'alçada dels objectes que defineixen una regió a la seqüència d'imatges, permet detectar els obstacles presents a l'interior del camí en un petit interval de temps. L'error en el càlcul de l'alçada dels

objectes depèn del nombre d'hipòtesis sobre l'alçada de l'objecte formulades i de la resolució amb que s'ha definit aquest conjunt d'hipòtesis.

- Integració de la informació continguda en la seqüència d'imatges, en la que s'observa l'entorn del vehicle des de diferents posicions. La integració de la informació obtinguda al observar l'entorn en moments diferents des de punts diferents i la imposició de les restriccions de model genèric de camí, permet solucionar moltes de les ambigüitats que apareixen de forma freqüent en el tipus de camí considerat, a l'hora de delimitar l'àrea que ocupa el camí a l'entorn.