

Algorismes de reconstrucció en tres dimensions

Gemma Rotger-Moll

Resum— Aquest article presenta un estudi comparatiu entre els diferents algorismes de reconstrucció en tres dimensions. Es presenten diverses tècniques agrupades en tres blocs, els mètodes de calibratge de càmeres, els mètodes de reconstrucció passius i els mètodes de reconstrucció actius. L'article presenta una revisió actualitzada d'alguns dels mètodes englobats en aquests blocs. Per a totes les tècniques utilitzades s'ha fet una revisió qualitativa i quantitativa dels resultats obtinguts per tal de comparar-los en diferents escenaris. Les conclusions i comparacions extretes d'aquest estudi es descriuen i justifiquen al final de l'article.

Paraules clau— Autocalibratge, Calibratge de càmeres, Codificació amb llum estructurada, Multi view stereo, Reconstrucció 3D.

Abstract— This paper presents a comparative study between different three dimensional reconstruction algorithms. This article presents several techniques grouped in three blocks, camera calibration methods, passive reconstruction methods and active reconstruction methods. This article illustrates an up-to-date review of qualitative and quantitative results in order to compare them in different scenarios. Comparisons and conclusions extracted from this study are described and justified at the end of the article.

Index Terms—3D reconstruction, Camera calibration, Coded structured light, Multi view stereo, Self-calibration.



1 INTRODUCCIÓ

DES dels inicis del 3D, aquest camp ha estat sempre en auge doncs moltes són les aplicacions que fan ús de les seves tècniques de representació, reconstrucció, visualització, etc. Avui dia, ja no només s'utilitza en el cinema i en les aplicacions científiques sinó que qualsevol persona té accés a pantalles tridimensionals. Aquest fet fa que el mercat s'ampliï i neixin noves necessitats.

Una d'aquestes recents necessitats és la reconstrucció en tres dimensions, ja sigui per obtenir un escenari o un personatge d'un videojoc com per aconseguir una reconstrucció acurada d'un objecte o espai interior.

En aquest article es presenta un estudi d'alguns dels algorismes existents per la reconstrucció en tres dimensions de diferents espais i també d'objectes pertanyents a aquests espais. L'estudi s'estructura en els tres grans blocs que continuen.

El calibratge és una tasca fonamental per tal d'establir els paràmetres de la càmera abans de la reconstrucció, però el calibratge tradicional de càmeres que es realitza a partir de la fotogrametria requereix un objecte de calibratge molt específic i un software molt elaborat, en aquest cas l'elaborat per Bouguet [1]. En aquest treball es tractarà aquest tema comparant l'eficàcia d'aquests algorismes en comparació amb les noves propostes d'autocalibratge [2].

Entrant pròpiament en el que és la reconstrucció 3D es pot fer una prèvia classificació entre els mètodes actius,

que són aquells que interfereixen amb l'objecte ja sigui mitjançant un làser, un patró de llum, o fins hi tot per contacte, els i passius, que són aquells que només necessiten un conjunt d'imatges de l'objecte per reconstruir-lo.

En els mètodes actius, la codificació de llum estructurada permet obtenir una tercera dimensió directa de la imatge, ja que els patrons es deformen sobre l'objecte i permeten extreure'n la geometria fàcilment [3]. Aquests mètodes permeten treballar amb càmeres preparades per aquesta tasca, com per exemple la Kinect, o amb càmeres corrents que treballin conjuntament amb un projector.

Tot i que la llum estructurada sembla ser el mètode decisiu, no és apte pel seu ús en exteriors i té nombroses limitacions en interiors. Per tal de cobrir aquestes mancances, també s'ha realitzat un estudi de les tècniques passives com multi view stereo, carving o visual hull per tal de comparar-les amb les anteriors i comprovar en quins àmbits guanya terreny cadascuna [4].

Un dels principals problemes del 3D és, en general, el copió nombre d'errors. Atès que la càmera ja comet errors en prendre la fotografia, que la il·luminació no sempre és tan bona com hauria de ser i que en els mateixos algorismes de reconstrucció també se'n cometem, el resultat sol ser un 3D amb errors de desfasament, forats, etc. Durant el transcurs d'aquest article també es descriuran mètodes per tractar de reduir el nombre d'errors en què deriven alguns dels mètodes existents.

-
- E-mail de contacte: Gemma.Rotger@e-campus.uab.cat
 - Menció realitzada: Computació
 - Treball tutoritzat per: Felipe Lumbreras Ruiz (Ciències de la computació)
 - Curs 2013/14

2 OBJECTIUS

2.1 Crítics

1. Estudiar un conjunt d'algorismes de reconstrucció automàtica en 3D que englobi mètodes de cadascun dels blocs abans esmentats.
2. Realitzar un estudi de les fortaleces i debilitats de cada un dels mètodes que s'esmenten a continuació per tal de trobar el que més s'ajusta a cada aspecte de la reconstrucció 3D.

2.2 Principals

3. Aprendre la importància del calibratge i els diferents mètodes de calibratge que existeixen.
4. Analitzar mètodes de reconstrucció a partir de la codificació de llum estructurada. Patrons binaris, patrons de Bruijn i mapes de punts aleatoris.
5. Estudiar l'algorisme ICP per a la posada en correspondència de dos núvols de punts.
6. Investigar l'eficiència de l'ICP mitjançant l'ús combinat amb algorismes de detecció i correspondència de característiques. SIFT, SURF, MSER.
7. Explorar la possibilitat de combinar l'ICP amb algorismes per extreure la transformació entre dos conjunts de característiques posades en correspondència en imatges en color. RANSAC.
8. Indagar sobre l'algorisme Kinect Fusion que fa ús de patrons de punts aleatoris per reconstruir en tres dimensions en temps real.
9. Estudiar i analitzar mètodes de reconstrucció a partir d'imatges en color. Multi view stereo, Carving i Visual Hull.

2.3 Secundaris

10. Explorar les diferents combinacions possibles entre els algorismes per tractar de millorar els resultats obtinguts.
11. Incloure totes les funcionalitats anteriors en un Toolbox de Matlab.

3 ESTAT DE L'ART

La reconstrucció en tres dimensions és un tema important en la visió per computador, és per això que en els darrers anys un gran nombre d'algorismes d'excel·lent qualitat han estat desenvolupats. L'estat de l'art en aquest camp evoluciona ràpidament i molts són els algorismes que apareixen i els que van quedant-se obsolets.

Per parlar amb propietat és imprescindible parlar de l'estat de l'art fent divisió entre els grups de mètodes que es tracten a l'article, ja que en l'actualitat existeixen un gran nombre de tècniques i no totes han estat analitzades. Per un costat les tècniques actives, on es necessita un projector, emissor i receptor de IR o làser; i les tècniques passives, on l'ús d'una sola càmera en diferents posicions és suficient.

Les tècniques actives consisteixen en la projecció d'un patró, normalment de punts o ralles, sobre una superfície plana. Qualsevol objecte que s'interposi entre el projector i

la superfície plana causa la deformitat del patró, és a partir d'aquesta desfiguració que es pot obtenir la tercera dimensió dels objectes presents a l'escena.

Diversos grups d'algorismes s'engloben dins aquest conjunt de tècniques [3], se'n poden trobar de domini discret o continu. Dins el domini discret es comprenen les tècniques que multiplexen les dades de la imatge en l'espai, aquestes són els patrons de Bruijn [5][6][7], que consisteixen en la projecció d'un sol patró de línies de colors de manera que, la combinació de cada grup de m colors es repeteix un sol cop en tot el patró de longitud n . Els procediments amb M -Arrays consisteixen en patrons de punts on cada submatriu de mida $m \cdot m$ es repeteix tan sols un cop dins la matriu de mida $n \cdot n$. Finalment les tècniques no formals són aquelles que consisteixen en modificacions de les tècniques anteriors, en la majoria de casos d'aquest grup de tècniques es treballa amb l'aleatorietat del patró, per exemple, el cas de la Kinect, que treballa amb punts aleatoris. La multiplexació temporal consta de dos tipus de patrons, els binaris [8] que són una seqüència de ratlles horitzontals o verticals blanques i negres que es projecten diferents cops alterant el nombre de ratlles en cada projecció a a 2_i on $1 \leq i \leq n_{patrons}$, i els patrons N -aris, que tenen el mateix procediment però amb un patró de N colors.

Canviant al domini continu també es troben la multiplexació en l'espai amb els patrons de degradat, que consisteixen en uns patrons on es degrada el color inicial (normalment negre) fins a arribar al blanc. També existeixen els patrons de fases, simple (SPS) i múltiple (MPS) que consisteixen en un patró de degradat determinat per una ona sinusoidal. Finalment els mètodes continus que multiplexen en freqüència es basen en la transformada de Fourier (FTP)[9].

En termes de multiview, es pot reconstruir a partir de des de dues càmeres calibrades [10][11] o sense calibrar (self-calibration) multiview stereo [3] on es posen en correspondència punts comuns a les dues imatges per tal d'obtenir la transformació d'una a l'altra; a partir de les siluetes de les diferents vistes d'un objecte com el Carving [12] on s'esculpeix sobre un bloc de vòxels les siluetes de l'objecte, i el Visual Hull [13] on els cons generats per les diferents siluetes es creuen; a partir de la il·luminació i les ombres com el photometric stereo [14], on s'observa un objecte sota diferents condicions d'il·luminació o mitjançant la factorització del moviment com el 3D Factorization [15].

Totes les tècniques esmentades a continuació són les utilitzades en l'actualitat per a la tasca de reconstrucció en 3D. A més en un futur no gaire llunyà, es preveu que els telèfons intel·ligents puguin portar a terme aquesta tasca en temps real [16].

4 METODOLOGIA

La metodologia del treball es divideix en tres blocs, calibratge, reconstrucció mitjançant mètodes actius (Coded Structured Light) i mitjançant mètodes passius. A cada bloc s'especificuen els procediments realitzats.

4.1 Calibratge

El calibratge de la càmera és un pas important en la visió 3D per tal d'evitar reconstruccions deformes i mal posicionades. Encara que aquesta passa no sigui necessària gràcies a l'autocalibratge, que sembla ser la tendència, aquest procediment permet reduir la complexitat i el temps de reconstrucció. El procés de calibratge permet extreure la mètrica de l'espai a partir d'imatges 2D. D'acord amb la naturalesa d'aquest projecte, aquest pas és fonamental per obtenir resultats el més acurats possibles. Per aquesta raó s'han estudiat alguns algorismes que es troben entre el calibratge per fotogrametria, que ofereix una precisió excel·lent a expenses d'un software elaborat, i l'autocalibratge, que tot i ser menys robust permet calibrar directament sobre la seqüència de fotografies preses i sense més interacció humana que el fet d'indicar una distància real sobre una de les imatges.

Els mètodes de calibratge que s'han estudiat han estat, el calibratge per fotogrametria proposat per Zhang [17] i implementat per Bouguet al toolbox de Matlab [1], el calibratge amb objectes dels quals en coneixem la geometria [18] i l'autocalibratge amb un Tsai grid que conté imatges de dibuixos animats en lloc de taulers d'escacs.

El primer experiment ha consistit en un calibratge de la càmera fent ús del software de Jean-Yves Bouguet i d'un tauler pla. Aquest procés de calibratge s'ha repetit fins a un total de quatre cops variant el nombre d'imatges per a cada prova, en 5, 10, 15 i 20 de les que es poden observar a la figura 1. S'ha començat per 20 imatges i en cada pas s'han eliminat les que contenen els píxels erronis més elevats (outliers).

L'objectiu d'aquest experiment ha estat veure quin és el nombre d'imatges mínim d'imatges per obtenir resultats precisos en les circumstàncies específiques de l'experiment, així com veure com es redueix l'error a escala de píxel mitjançant l'eliminació d'outliers.

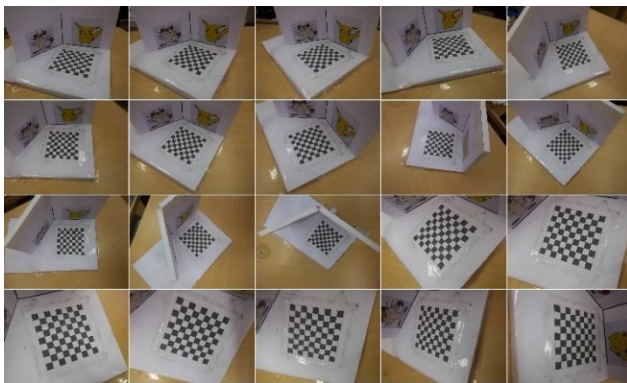


Figura 1. Les 20 imatges utilitzades per calibrar la càmera mitjançant el calibratge per fotogrametria.

El segon experiment ha consistit a comparar els resultats dels tres mètodes de calibratge prenent com a referència els paràmetres obtinguts pel millor cas de l'experiment anterior i com a fotografies les que es poden observar a la figura 2. S'ha de tenir en compte que el cub i el Tsai grid són

capaços de calibrar amb una sola imatge, aquest últim fins i tot, gairebé sense interacció humana.



Figura 2. Tsai grid i cub utilitzats per fer les comparacions entre els diferents mètodes de calibratge analitzats.

Per tant amb aquest experiment es busca quantificar com de bons són aquests mètodes respecte a la referència que considerem totalment fiable.

Per tal de calibrar una càmera en una posició utilitzant una sola imatge d'un cub es necessiten tres dades.

En primer lloc les coordenades 2D del cub a la imatge, en el cas experimentat, els punts de canvi de peça d'un cub de Rubik. En segon lloc es necessita saber la mida del costat de cub i per últim el número de divisions que té el cub per costat.

Mitjançant aquesta informació podem obtenir les coordenades 2D de la imatge seleccionats els punts esmentats. Els punts 3D es poden calcular generant un cub 3D teòric amb les dades del cub introduïdes, mida i nombre de divisions per costat.

Un cop obtinguts tots dos conjunts de coordenades necessitem saber la matriu de la càmera per conèixer els seus paràmetres. Per extreure aquesta matriu necessitem, en primer lloc, posar en correspondència aquests punts mitjançant la matriu que es pot observar a la equació 1,

$$\lambda x_i = P X_i \quad P = K[R | t] \quad [\text{Eq 1}],$$

on x_i es correspon amb els punts bidimensionals, X_i amb els punts tridimensionals, K amb la matriu de calibratge, R amb la matriu de rotació i t amb el vector de translació. [16]. Aquest problema de conversió entre espais de coordenades s'ha resolt utilitzant l'algorisme DLT (Direct Linear Transformation).

El vector que resulta de l'última columna de V en la descomposició SVD del resultat obtingut en el pas anterior es correspon amb els vectors fila que componen la matriu de la càmera p_1, p_2, p_3 i amb els valors propis $\lambda_1, \lambda_2 \dots$ de la matriu.

Un cop obtinguda la matriu de la càmera es pot descompondre en una matriu de rotació R i una matriu triangular Q equivalent a la matriu K de calibratge mitjançant una descomposició RQ .

Finalment obtenim la translació t multiplicant la inversa de K per la última columna de la matriu de la càmera P .

Per altra banda i per tal de desenvolupar un sistema quasi automàtic de calibratge s'han enganxat dos dibuixos, un a cada costat, al Tsai grid (veure figura 2). L'únic paràmetre d'entrada que es demana a l'usuari és l'angle en què es troben els dos plans l'un respecte a l'altre, tot i que si la imatge es pren amb un angle prefixat no caldria demanar cap paràmetre.

A continuació es posen en correspondència les imatges digitals amb la imatge presa del Tsai grid mitjançant el detector de característiques SURF (Speeded Up Robust Features). S'ha utilitzat el mateix algorisme com a descriptor. Seguidament s'han aplicat dites característiques al RANSAC (RANdom SAmple Consensus) [19] per tal de pesar-les, posar-les en correspondència i trobar-ne l'homografia. Per tal d'evitar l'aleatorietat del RANSAC s'han guardat les correspondències per executar en experiments posteriors amb les mateixes condicions i per tal d'obtenir els mateixos resultats.

Un cop trobada l'homografia s'ha establert un Tsai grid virtual amb una determinada escala s . A continuació s'han mapejat els punts característics de les imatges digitals sobre els plans virtuals fent ús de l'escala establerta i amb la rotació indicada per l'angle entre els plans. Un cop desenvolupades aquestes passes, s'han obtingut els punts 2D i els punts 3D de la imatge i es pot procedir amb el procés explicat per calibrar un cub.

Amb aquests experiments s'ha pretès demostrar la validesa dels resultats de mètodes alternatius de calibratge menys laboriosos que el calibratge amb fotogrametria, per tal de poder decidir, segons la precisió necessària i el temps del qual es disposa, quin és el millor mètode per calibrar en cada situació.

4.2 Mètodes actius

Ja dins la part de reconstrucció en 3D pròpiament dita, s'ha treballat, primerament, amb llum estructurada, fent ús de patrons projectats binaris i De Bruijn, i també de la Kinect.

4.2.1 Patrons binaris

Per tal de recuperar l'estructura 3D d'un objecte mitjançant patrons binaris [8] s'ha de col·locar l'objecte front un fons pla i projectar una seqüència de patrons binaris (blanc i negre) on per a cada patró es projecten 2^i línies on $1 \leq i \leq n_{patrons}$ línies tal com es mostra a la figura 3. El sistema càmera-projector ha d'estar prèviament calibrat [7][20].

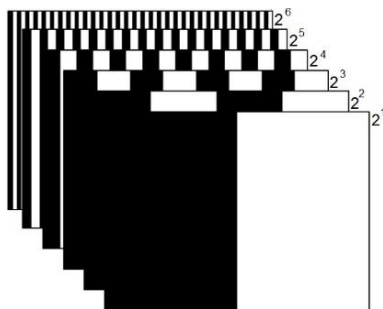


Figura 3. Patrons binaris projectats per reconstruir l'escena.

La projecció dels patrons i la posterior recuperació de les imatges mitjançant la càmera, fa que posteriorment cada píxel pugui ser codificat amb un codi de longitud igual al nombre de patrons utilitzats per generar-lo.

Per decidir si un píxel d'un determinat patró està il·luminat o no s'ha utilitzat l'algorisme de binarització d'Otsu

[21] per decidir el llindar global per a tota la imatge mitjançant el qual es pren la decisió de considerar un píxel il·luminat o fosc. Aquesta decisió s'ha presa ja que els algorismes locals poden fallar en regions on hi ha un píxel a la frontera de decisió i el color predominant de la regió sigui el contrari.

Per codificar cada ratlla del patró recuperat s'ha multiplicat el valor del píxel (0 o 1) pel valor del bit que ocupa, donant més importància als patrons amb menys ratlles ja que la distorsió de la càmera provoca menys curvatura en aquests patrons. Així per a cada píxel obtenim el següent codi de l'equació 2.

$$CodiPixel = Px_0 \cdot 2^0 + Px_1 \cdot 2^1 + \dots + Px_n \cdot 2^n \text{ [Eq 2].}$$

Un cop codificats es prenen iterativament tots els possibles valors de *CodiPixel* per tal de buscar tots els píxels de la imatge codificats amb aquell determinat codi [8].

A partir d'una línia de referència que creua tot el patró passant per totes les línies en una zona de la imatge no deformada es calcula el punt mitjà de l'etiqueta que marca tots els punts corresponents a un codi. Per evitar píxels d'error derivats del color de l'objecte (que dona com a fosc píxels que haurien de ser clars) i la reflectància de l'objecte (que dona com a il·luminats píxels que haurien de ser foscs) s'ha fet un opening (erosion + dilatation) que permet eliminar aquests errors.

Un cop definides les etiquetes es procedeix a calcular la distància entre cada píxel pertanyent a l'etiqueta i el centre de la mateixa, la component Z del punt es calcula amb la fórmula de l'equació 3.

$$Z_{i,j} = Z_{cam} - d(\text{punt}, \text{CentreEtiqueta}) * \text{resol}_x \cdot 1/f_x \text{ [Eq 3].}$$

L'experiment realitzat per aquest tipus de reconstrucció ha estat el procediment complet i cronometrat de reconstrucció per tal de comparar el resultat visual amb l'eficiència en temps de còmput amb els altres mètodes que s'estudien.

4.2.2 Patrons De Bruijn

Per recuperar l'estructura 3D d'un objecte mitjançant patrons De Bruijn [5][6][7] el procés és una mica més complicat que fent ús dels binaris, ja que els patrons de Bruijn (figura 4) són de color i com a tal comporten una dificultat afegida, calibrar el color.

Amb el sistema prèviament calibrat geomètricament s'ha de calibrar el color de les imatges, això és degut al fet que el color es deforma quan és projectat, ja que el projector utilitza filtres de colors. Per resoldre aquest problema s'han portat a terme dues passes.

La primera consisteix a seleccionar la millor imatge entre un conjunt d'imatges candidates seleccionades manualment. Aquest procés calcula la distància entre els colors i selecciona la imatge on es maximitza aquesta distància. Per tant s'obté la imatge on millor es distingeixen els colors.

La segona consisteix a decidir quins píxels pertanyen a cada color. Per portar a terme aquest pas es va plantejar en un primer pas prendre un conjunt de mostres de cada color i calcular-ne la mitjana, i seguidament seleccionar el conjunt de píxels compresos en un rang proper a la mitjana, el

procés es va portar a terme tant amb la imatge RGB com amb la component H de la imatge HSV. Aquest mètode es va descartar pel fet que necessita massa interacció humana. La següent proposta i la que finalment s'ha implementat ha estat un k -means amb tants clústers com colors tingui el patró, de tota manera el fet que els primers resultats es perceben millor que amb l'anterior mètode i no requereix cap interacció humana fa que sigui el mètode seleccionat.

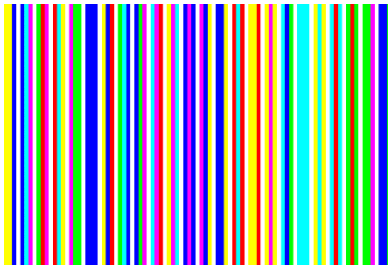


Figura 4. Patrons de Bruijn projectats per reconstruir l'escena.

Molts cops el calibratge del color és dolent, per aquest motiu s'ha decidit llençar un experiment per observar en diferents escenes sobre quina imatge és millor reconstruir el 3D calibrada o sense calibrar.

El següent pas a estat aplicar sobre cada tipus d'imatge una correlació normalitzada en una finestra lliscant que passa per tota la imatge buscant els punts de màxima correlació dins la finestra. D'aquesta manera es permet detectar, posteriorment, les ralles mitjançant les quals es reconstruirà el 3D [7].

Un cop codificades les imatges es procedeix a eliminar l'error amb un algorisme manual que permet eliminar l'error dins un determinat quadrat. En aquest cas no és possible aplicar un opening, ja que les regions d'error poden ser tan grans com la mateixa línia. El fet d'utilitzar un procés manual elimina l'automatisme de l'algorisme però en la majoria d'imatges no és necessari aplicar-lo.

Finalment, de la mateixa manera que s'ha procedit amb els patrons binaris per una determinada ratlla horitzontal que travessa tot el patró i no està deformada en cap punt del seu domini, es busca la intersecció amb les etiquetes detectades anteriorment.

S'agafa el punt mitjà d'aquesta intersecció i es compara la component x dels altres punts de l'etiqueta vertical (deformada) amb la component x de la intersecció.

Aquesta distància dóna el 3D utilitzant la mateixa fórmula que amb els patrons binaris (veure l'equació 2).

L'experiment realitzat per aquest tipus de reconstrucció ha estat el procediment complet i cronometrat de reconstrucció per tal de comparar el resultat visual amb l'eficiència en temps de còmput amb els altres mètodes que s'estudien.

4.2.3 Kinect

La Kinect és un controlador desenvolupat per Microsoft per la videoconsola xbox 360 i actualment també per PC.

Tal com es pot apreciar a la figura 5 aquest controlador disposa d'un sensor de color, un emissor d'IR i el seu corresponent sensor a més d'un array de quatre micròfons i un motor d'inclinació. En aquest projecte només s'han utilitzat el sensor de color i l'emissor i el sensor d'IR.

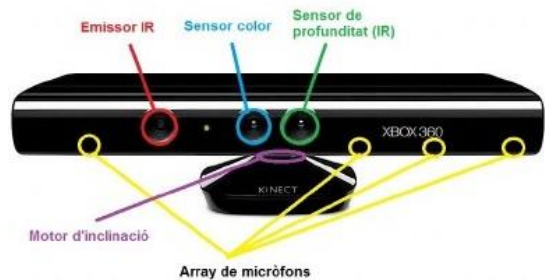


Figura 5. Els sensors de la Kinect 360.

Per procedir amb l'anàlisi de la Kinect s'han estudiat qualitativament l'algorisme Kinect Fusion per una part, i per l'altre l'Iterative Closest Point (ICP) que és un pas dels passos fonamentals Kinect Fusion, el que posa en correspondència els dos núvols de punts [22]. Aquest algorisme de correspondències també es pot aplicar als mètodes de llum estructurada explicats anteriorment.

L'algorisme ICP consisteix en un algorisme que per cada punt d'un núvol de punts calcula el punt a menor distància d'un altre núvol de punts, de manera que intenta posar tots els punts en correspondència. Per altra banda l'algorisme KinectFusion és més complicat, ja que en primer lloc calcula la posició de la càmera (tracking) tot seguit de la reconstrucció volumètrica que inclou, entre altres procediments (figura 6) l'algorisme ICP. Finalment integra les dades actuals amb el model volumètric que guarda de captures anteriors.

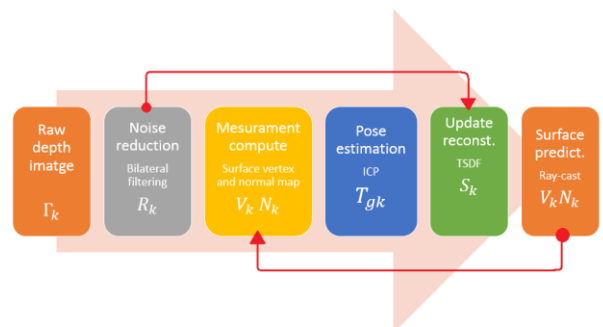


Figura 6. Estructura de la reconstrucció volumètrica utilitzada per l'algorisme Kinect Fusion

L'ICP [23][24] s'ha provat primerament sobre les escenes sintètiques que es mostren a la figura 7 per tal d'evitar, en un primer anàlisi, els nombrosos errors que deriven de la sensibilitat de la Kinect a la il·luminació, a les distàncies fora del rang òptim, etc. A la imatge A podem veure una imatge que correspon a una captura amb la càmera orbitant al voltant de l'objecte, mentre que a la imatge B podem observar que la imatge és una panoràmica, en el cas dels experiments, ambdues de 360°. En ambdós casos s'ha aplicat l'algorisme ICP entre núvols de punts corresponents a les captures i s'ha analitzat, per una part, la variabilitat del resultat segons el núvol de punts i en segon lloc, la variabilitat del resultat segons l'angle de salt entre captures. L'objectiu d'ambos experiments és reduir al màxim el temps de còmput. Per tal de reduir-lo encara més dins les possibilitats de l'experiment s'ha substituït la força bruta

en el càlcul de correspondències entre punts que implementa l'algorisme ICP per defecte, per una versió que utilitza k -d trees.



Figura 7. Imatges sintètiques utilitzades per a la reconstrucció mitjançant ICP. A, mostra de les imatges sense texturitzar utilitzades per la càmera en òrbita. B, mostra texturitzada utilitzada per la càmera moviment panoràmic.



Figura 8. Imatges RGB reals pertanyents al conjunt de dades utilitzades per a la reconstrucció mitjançant els mapes de profunditat de la Kinect i ICP. A, mostra de les imatges utilitzades per la càmera en òrbita. B, mostra utilitzada per la càmera en moviment panoràmic.

Els mateixos experiments s'ha repetit amb les escenes reals que es mostren a la figura 8, i s'han comparat amb els resultats obtinguts per les escenes sintètiques.

Finalment s'ha experimentat l'ús conjunt de ICP + SURF i RANSAC [19] per tal d'ajudar a l'algorisme ICP a aconseguir millors resultats apropant els núvols de punts prèviament abans d'aplicar l'algorisme ICP. Aquest experiment s'ha realitzat utilitzant l'angle i el nombre de punts òptims extrets dels experiments anteriors, utilitzant les imatges panoràmiques, que són les que donen pitjors resultats. S'ha utilitzat la imatge corresponent a la figura 8B.

4.3 Mètodes passius

Per portar a terme la part de multiview (diferents punts de vista) s'ha utilitzat un conjunt de fotografies [25] en color d'una determinada escena per reconstruir a partir d'aquestes imatges amb 123D Catch[26] i CMVS[27] amb VisualSFM[28]. I també Carving[29] i Visual Hull[30] amb Matlab. Hem descartat un enfocament més tradicional basat en dues vistes [31] atès que ens hem decantat per solucions de baix cost que requereixin una sola càmera.

Així doncs, els seleccionats són programes ja desenvolupats dels quals s'ha volgut provar la seva eficiència així com els resultats, tant un respecte de l'altre, com enfront als obtinguts amb els mètodes actius.



Figura 9. Imatges utilitzades per la reconstrucció amb 123D Catch i PMVS.

Com que es vol comparar el resultats obtinguts per els programes i, al no haver-se desenvolupat al laboratori no es coneixen tots els detalls d'implementació s'ha decidit utilitzar el conjunt de fotografies que es mostra a la figura 9 ja que disposa d'un ground truth que ens permet comprovar el valor dels resultats respecte alguna cosa fixa i no respecte l'un de l'altre.



Figura 10. Imatges utilitzades per la reconstrucció amb carving A, i visual hull B.

D'altra banda els algorismes Carving i Visual Hull funcionen amb imatges d'un objecte preses amb una càmera en òrbita al voltant. Per aquest motiu s'han utilitzat les imatges de la figura 10. En el cas particular de Visual Hull s'ha utilitzat la imatge d'un objecte presa amb dos miralls (figura 10B).

L'algorisme de Carving funciona esculpint sobre un bloc de vòxels la silueta de les imatges captades. És un mètode dens però la inicialització és molt costosa. El Visual Hull, en canvi, funciona encaixant totes les siluetes al punt mig de reconstrucció. Tots dos són mètodes que resulten en models 3D densos molt acurats.

5 RESULTATS

5.1 Calibratge

Experiment 1. Variació del nombre d'imatges en el calibratge fotogramètric amb patrons plans.

A la taula 1 es pot observar la incertesa en píxels generada per el calibratge variant el nombre de imatges. Les dues últimes files representen l'error en píxels que suposa en cada dimensió, x i y .

A la figura 11 es pot observar un conjunt gràfics que representa l'evolució de l'incertesa respecte el nombre d'imatges utilitzat.

	5	10	15	20
f_x	32.89	11.80	5.75	4.10
f_y	25.41	12.46	5.82	4.15
c_x	15.66	12.59	5.72	5.15
c_y	45.54	11.84	5.55	4.89
sk	0.00	0.00	0.00	0.00
k_1	0.07	0.04	0.02	0.02
k_2	0.75	0.41	0.11	0.11
E_x	0.18	0.21	0.31	0.41
E_y	0.44	0.43	0.46	0.50

Taula 1.- Variació del nombre d'imatges utilitzades per calibrar la càmera. Les columnes representen el número d'imatges utilitzades, les files els diferents paràmetres analitzats. Els resultats estan expressats en termes de píxels d'incertesa excepte en les dues últimes files que s'expressen en termes de píxels d'error.

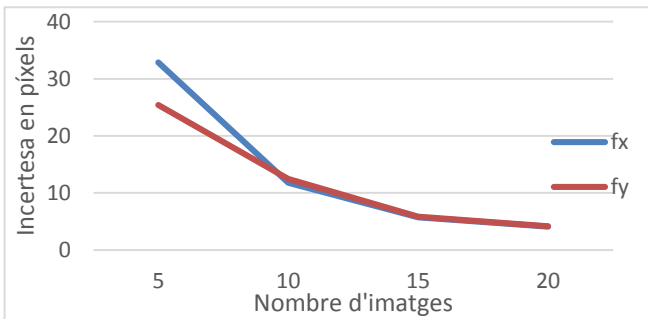


Figura 11.- Evolució de l'incertesa respecte la variació del nombre d'imatges utilitzades per calibrar la càmera.

Experiment 2. Comparació del millor model de calibratge amb fotogrametria amb el model obtingut pel cub i el Tsai grid.

A la taula 2 es pot observar la variació de cada paràmetre intrínsec dels dos models de calibratge alternatiu amb el calibratge amb fotogrametria pres com a referència.

	Fotogr.	Cub	Tsai grid
f_x	1121.70	975.92	1136.90
f_y	1125.36	983.85	1194.82
c_x	598.38	526.27	596.26
c_y	490.39	447.55	488.51

Taula 2. Variació dels paràmetres intrínsecs obtinguts amb el Cub i el Tsai grid respecte als obtinguts amb el calibratge amb fotogrametria de 15 imatges. Unitats en píxels.

5.2 Mètodes actius

Experiment 3. Comparació dels models 3D obtinguts dels patrons binaris i de Bruijn.

A la taula 3 es pot observar la mitja dels resultats obtinguts en termes de temps d'execució cadascun dels algorismes en una execució, trenta-sis execucions, que equivalen a totes les vistes de l'objecte, juntament amb el temps que es triga a fer l'ICP entre dues perspectives i també entre totes les perspectives, dues a dues, enllaçant la última amb la primera.

	Binaris	Bruijn
1 captura	19.60 s	52.10 s
36 captures	705.54 s	1875.49 s
1 ICP	3.50 s	3.50 s
36 ICP	125.83 s	125.83 s
TOTAL 36	831.38 s	2001.32 s

Taula 3. Resultats de rendiment, representats en segons per execució, obtinguts de les reconstruccions binàries i de Bruijn amb 32.265 punts per frame.

Mentre que la reconstrucció binària es desenvolupa satisfactòriament, la reconstrucció amb patrons de Bruijn no és així. La tassa no es reconstrueix degut al seu color. Es poden apreciar els resultats a la figura 12.

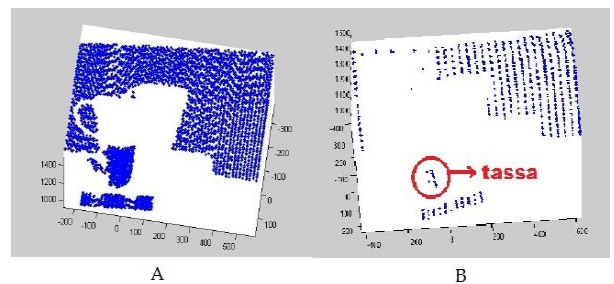


Figura 12. Resultats de les reconstruccions en 3D mitjançant llum projectada. A. Resultats amb reconstrucció binària. B. Resultats amb reconstrucció de Bruijn.

Experiment 4. Comparació qualitativa de la reconstrucció de Bruijn amb una imatge amb color calibrat i la mateixa sense color calibrat.

A la figura 13 es poden observar els resultats obtinguts que es discuteixen a la secció de conclusions.

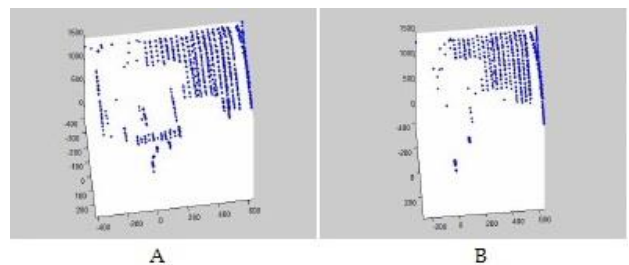


Figura 13. Resultats de les reconstruccions de Bruijn. A. Reconstrucció amb el color sense calibrar. B. Reconstrucció amb el color calibrat.

Experiment 5. Variació del nombre de punts (càmera en òrbita i dades sintètiques) en l'aplicació de l'algorisme ICP.

Per la imatge que s'observa a la figura 7A, els resultats en termes de graus d'error i segons el nombre de punts en l'execució són els que es poden veure a la taula 4. En dades reals, per la imatge que s'observa a la figura 8A.

%Pts	Error sint.	t sint.	Error real	t real
100	5.06°	12.66 s	3.95°	18.77 s
90	4.91°	15.34 s	3.93°	13.81 s
80	4.93°	9.23 s	4.04°	11.42 s
70	5.01°	8.82 s	3.70°	9.52 s
1	5.25°	1.15 s	3.81°	1.32 s

Taula 4. Resultats en termes d'error en graus i temps en segons obtinguts de la reconstrucció mitjançant la Kinect i ICP variant el nombre de punts.

Experiment 6. Variació de l'angle de salt entre dues captures (càmera en òrbita i dades sintètiques) en l'aplicació de l'algorisme ICP.

Per la imatge que s'observa a la figura 7A, els resultats en termes de graus d'error respecte l'angle entre dues imatges en l'execució són els que es poden veure a la taula 5.

Angle (°)	Error
10°	4.52 °
20°	9.54 °
30°	15.07 °
40°	19.56 °

Taula 5. Resultats en termes d'error en graus obtinguts de la reconstrucció mitjançant la Kinect i ICP variant l'angle de salt entre dues imatges a comparar i fixant el nombre de punts al 70%.

Experiment 7. Variació del nombre de punts (càmera en moviment panoràmic i dades reals) en l'aplicació de l'algorisme ICP.

Per la imatge que s'observa a la figura 8B, els resultats en termes de graus d'error i segons el nombre de punts en l'execució són els que es poden veure a la taula 6. En aquest cas l'angle entre dues imatges consecutives és de 2°.

% Punts	Error	t(s)
100	0.33°	243.80 s
90	0.30°	194.75 s
80	0.29°	161.13 s
70	0.38°	140.67 s
1	0.05°	0.36 s

Taula 6. Resultats en termes d'error en graus i temps en segons obtinguts de la reconstrucció mitjançant la Kinect i ICP variant el nombre de punts.

Experiment 8. Variació de l'angle d'error de l'ICP vs ICP+RANSAC.

A la taula 7 es poden observar els resultats obtinguts de la comparació de múltiples execucions en cada cas (ICP i ICP+RANSAC). Per a cada cas s'ha fet la mitjana de cinc execucions.

ICP	ICP+RANSAC
2.97°	0.19°

Taula 7. Resultats de la comparació del ICP vs ICP+RANSAC amb la imatge de la figura 7B i un angle de 10°

5.3 Mètodes passius

Experiment 9. Comparació resultats obtinguts 123D Catch vs PMVS.

Tal i com es pot observar a les taules 8 i 9, els resultats obtinguts en termes de rendiment no permeten utilitzar aquests mètodes per a reconstrucció en temps real. De tota manera els resultats obtinguts en termes de qualitat de la reconstrucció mitjançant un moviment de càmera panoràmic és bastant millor que l'obtingut mitjançant la llum estructurada. Els resultats qualitats es poden observar a les imatges 14A i 14B.

Activitat	Temps (min)
Updating	3
Computing	29.55
Downloading	2.4
TOTAL	34.95

Taula 8. Resultats de rendiment de 123D catch per a les condicions explicades a la metodologia.

Activitat	Temps (min)
Updating	0.2
Feature detection	0.62
Pairwise matching	9.85
3D sparse	1.37
3D dense	34
TOTAL	46.04

Taula 9. Resultats de rendiment de PMVS i CMVS catch per a les condicions explicades a la metodologia.

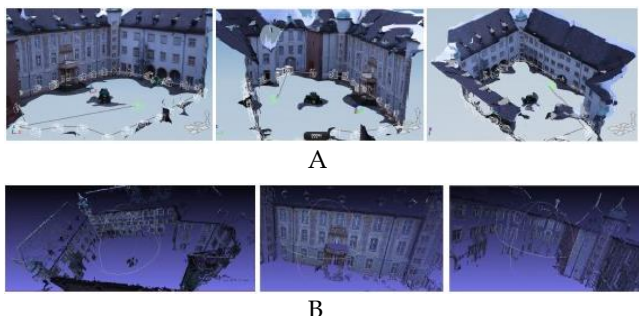


Figura 14. Resultats visuals obtinguts de la reconstrucció de la imatge de la figura 9 utilitzant 123D Catch A, i PMVS B.

Experiment 9. Comparació resultats obtinguts Carving vs Visual Hull.

Els resultats visuals obtinguts de l'execució del Carving i Visual Hull són els que es poden veure a les figures 15A i 15B respectivament. L'execució del carving ha suposat un temps total de 166.58s mentre que el del visual hull 43.95s.

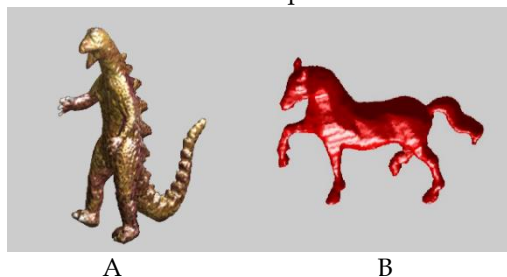


Figura 15. Resultats visuals obtinguts de la reconstrucció de les imatges de les figures 10A (carving) i 10B (visual hull).

6 CONCLUSIONS

El calibratge amb fotogrametria dona resultats amb una incertesa i errors acceptables a partir de quinze imatges. L'error creix escassament a mesura que aquest nombre s'incrementa. Això és degut al fet que el rang d'incertesa és més petit i un error comès a escala de píxel pot pesar més. De tota manera el decrement de la incertesa és considerable en el pas de cinc a quinze imatges, però aquest decrement té una forma logarítmica i no s'observa una gran millora de quinze a vint. Per aquest motiu s'han seleccionat quinze imatges com a nombre idoni per realitzar el calibratge d'una càmera en una bona proporció precisió entre nombre d'imatges.

Els resultats obtinguts amb el cub no han estat els esperats (taula 2). Aquest fet es podria arreglar adaptant l'algorisme de descens de gradient que permet estimar millor els paràmetres en el cas del calibratge amb fotogrametria a aquests models.

La reconstrucció amb llum estructurada és una bona solució per espais interiors encara que necessiti una il·luminació bona i constant, ja que permet reconstruir amb molt detall si els patrons escollits són bons. La Kinect és l'aparell de reconstrucció més nou en l'àmbit d'usuari, i segurament el que més ha revolucionat el mercat en els darrers anys, però és delicat respecte a la llum i a les distàncies i no sempre pot ser utilitzada. Per exemple, per reconstruir objectes a menys d'un metre o espais a més de tres. Per resoldre aquestes mancances podem fer ús d'altres patrons com els binaris. Amb aquests patrons no podem aconseguir una reconstrucció a temps real com amb la Kinect, però si en lloc del temps el que ens preocupa és la precisió de la reconstrucció, aquest mètode és un candidat clau a ser considerat.

La projecció de franges clares i fosques permet molta flexibilitat en la reconstrucció, ja que, no necessitem gaire il·luminació i l'únic requisit de l'escena és que els objectes siguin opacs i romanguin immòbils. Els resultats que ens ofereix són bons, en termes d'error, més bons que amb la Kinect a causa de l'opening aplicat per tal de reduir l'error.

Els patrons de Bruijn no han donat tan bons resultats en aquest experiment, segurament en unes circumstàncies més idònes també resoldrien problemes de reconstrucció satisfactòriament, però el fet que necessitin objectes sense textura i calibrar prèviament el color, ha fet que no s'hi trobi utilitat en el cas d'aquest estudi en concret.

En termes de Kinect i l'ús de l'algorisme ICP s'han aconseguit millorar els resultats fent ús de dualitat RGB - IR de l'aparell. L'algorisme RANSAC sobre imatges RGB ha estat el seleccionat per posar en correspondència les característiques detectades i descrites per el SURF. Fent ús d'aquest mètode s'ha aconseguit millorar l'algorisme tant en termes de resultat reduint l'error fins a setze vegades menys. A part s'ha reduït el temps d'execució respecte a l'execució amb el 100% de punts, ja que el procés només es calcula sobre les característiques detectades. Aquesta millora no es reflexa directament respecte al mateix nombre de punts sense aplicar el RANSAC però s'aprecia en veure que es milloren dues centèsimes tot i aplicar dos processos més sobre les dades abans de processar-les amb l'ICP quan s'esperava que el rendiment del procés empijores.

També s'ha estudiat l'efecte del moviment de la càmera i s'ha pogut comprovar com la càmera en òrbita és molt més flexible en la selecció de paràmetres (nombre de punts o angle de salt) que la càmera en moviment panoràmic. Això és degut al fet que la càmera en òrbita té moltes més correspondències a l'hora de reconstruir, ja que la superposició de les imatges és major. Per altra banda la càmera panoràmica necessita un angle molt petit de salt, ja s'ha vist que l'angle de 10° no és bo per aquesta tasca. A més, l'algorisme RANSAC ajuda a apropar els dos núvols de punts a un angle menor abans d'aplicar ICP.

Finalment, en temes de multiview, diversos mètodes s'han analitzat. El comercial 123D Catch permet reconstruccions a partir d'imatges amb un elevat nombre de característiques. Per escenes de les quals tenim poques imatges o bé tenen poques característiques resulta un mètode dolent. En termes d'eficiència és bastant lent, ja que no calcula les dades en local sinó que s'han de pujar al servidor de la companyia i a vegades és més lenta aquesta transmissió que el pròpi fet del processament.

PMVS o CMVS és encara més lent pel mateix nombre d'imatges i tot i que no conté tants errors com 123D Catch té més forats i la sensació d'un 3D poc dens.

Aquests mètodes funcionen de manera correcta tant en moviment orbital com en moviment panoràmic. Si mateix Carving i Visual Hull només ho fan en moviment orbital. Aquests algorismes tenen un cost computacional menor i els resultats obtinguts són espectaculars.

Carving permet esculpir un 3D amb molt detall i també mapejar la textura sobre el 3D ja acabat. El procés és ràpid, el pas que és lent és la inicialització, fet que podria solucionar-se preinicialitzant el sistema un cop carregades les dades o guardant la mida del bloc inicial el primer cop que es reconstrueix i utilitzar-lo la resta de vegades.

Visual Hull dona uns resultats verdaderment brillants a partir d'una sola imatge amb dos miralls. En aquest cas s'ha treballat amb quatre reflexes més la imatge original. A més, al contrari que el carving, no necessita inicialitzar el bloc, fet que fa que sigui més ràpid, gastí menys recursos de memòria i pugui ser pres amb una sola imatge.

Tasca	Mètode
Reconstrucció d'interiors	PMVS, Kinect Fusion
Reconstrucció d'exterior	123D Catch, PMVS
Reconstrucció d'objectes sense textura.	Binaris, de Bruijn, Kinect Fusion.
Reconstrucció d'objectes amb textura.	Binaris, Kinect Fusion, Carving, Visual Hull.
Reconstrucció d'objectes amb transparències.	Carving, Visual Hull.
Reconstrucció parcial amb una sola imatge.	Binaris, de Bruijn.
Reconstrucció completa amb una sola imatge (i dos miralls).	Visual Hull

Taulella 10. Relació de diferents tasques de la reconstrucció en 3D i els millors mètodes que s'ha experimentat durant aquest treball.

Finalment es proposa una relació de tasques i algorismes a la taula 10, on es pot observar un seguit de tasques de reconstrucció a realitzar i, en relació als experiments desenvolupats, quin mètode o mètodes es creuen més òptim per portar-la a terme.

Com a treball futur s'espera ampliar el conjunt d'algorismes sobretot en temes de stereo, així com millorar els existents. La major part de la feina portada a terme es pot observar al toolbox 3D Vision que s'ha desenvolupat en paral·lel amb aquest projecte.

AGRAÏMENTS

Vull donar les gràcies al meu tutor del projecte, Felipe Lumbreras, per tot el suport que m'ha donat durant aquests mesos, sense la seva ajuda aquest treball no hagués estat possible. També vull agrair al Coen Antens la paciència que ha tingut i tot el que m'ha ensenyat durant aquests mesos de treballar plegats. Per últim als meus companys de laboratori, que sempre hi han estat per donar-me un cop de mà, sobre tot l'Angès Borràs, el Joan Mas i el Jordi Abella.

BIBLIOGRAFIA

- [1] Camera Calibration Toolbox. Software. Desenvolupador: Jean-Yves Bouguet. http://www.vision.caltech.edu/bouguetj/calib_doc/
- [2] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press. Second Edition 2003.
- [3] J. Salvi, S. Fernandez, T. Pribanic and X. Llado, *A state of the art in structured light patterns for surface profilometry*. Pattern recognition 43 (2010) 2666–2680.
- [4] S.M. Seiz, B. Curless, J. Diebel, D. Scharstein and Richard Szeliski. *A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms*. Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on Volume 1 519 – 528.
- [5] H. Hügli and G. Maitre, *Generation and Use of Color Pseudo Random Sequences for Coding Structured Light in Active Ranging*. SPIE Conference Proceedings Vol. 1010 (1988).
- [6] J. Pages, J. Salvi and J. Forest, *A new optimised De Bruijn coding strategy for structured light patterns*. Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on Volume 4 (2004).
- [7] J. Pages, J. Salvi and C. Collewet, *Optimised De Bruijn patterns for one shot shape acquisition*. Image and vision computing 23 (2005) 707-720.
- [8] J.L. Posdamer and M.D. Altschuler, *Surface Measurement by Space-encoded Projected Beam Systems*. Computer graphics and image processing, 18, 1-17(1982).
- [9] M. Takeda and K. Mutoh. *Fourier transform profilometry for the automatic measurement of 3D object shapes*. Applied optics, vol.22, issue 24, pp. 3977-3982 (1983).
- [10] Y. Ma, S. Soatto, J. Kosecka and S.S. Sastry, *An invitation to 3-D vision: From images to geometric models*. Springer. 2010.
- [11] M. Kleinsteuber, *Einführung in Computer Vision* Martin Kleinsteuber. Technische Universität München, 2014.
- [12] K.N. Kutulakos, S.M. Seitz, *A Theory of Shape by Space Carving* International Journal of Computer Vision 38 (3), 199–218, 2000.
- [13] C. Guojun, S. Baoming *A Visual Hull Algorithm of 3D Reconstruction Based on Interframe Coherence*, International Conference on Virtual Reality and Visualization, 2011.
- [14] B.K.P. Horn, *Obtaining shape from shading information*. Shape from Shading, pages 121–171. MIT Press. 1989.
- [15] M. Brand, *A Direct Method for 3D Factorization of Nonrigid Motion Observed in 2D*, Mitsubishi Electric Research Laboratories, 2005.
- [16] P. Tanskanen, K. Kolev, L. Meier, F. Camposco Paulsen, O. Saurer, M. Pollefeys, *Live Metric 3D Reconstruction on Mobile Phones*, proc. ICCV13 (Int. Conf. on Computer Vision).
- [17] Z. Zhang. *Flexible Camera Calibration By Viewing a Plane From Unknown Orientations*. Microsoft Research. 1991.
- [18] C.O. Calle *Computer Vision Notes* Lunds Universitet, Sweden. 2014.
- [19] A. Hast, J. Nysjö, A. Marchetti *Optimal RANSAC - Towards a Repeatable Algorithm for Finding the Optimal Set*. Journal of WSCG, vol. 21, No.1 (2013), pp. 21-30.
- [20] S. Zhang. *Handbook of 3D Machine Vision. Optical Metrology and Imaging*. CRC Press. Taylor & Francis Group 2013.
- [21] N. Otsu, *A threshold selection method from gray-level histograms*. IEEE Transactions on Systems, Man and Cybernetics, Vol.9, No.1, (1979), P.62-66.
- [22] MSDN Library [pàgina web]. Lloc edició: Microsoft; data[14/02/2014]. Disponible a: <http://msdn.microsoft.com/>
- [23] D. Chetverikova, D. Stepanova, P. Krsek, *Robust Euclidean alignment of 3D point sets: the trimmed iterative closest point algorithm*. Image and Vision Computing 23 (2005) 299–309.
- [24] J. Mateo Prous. *Algorithms for the Multiple Variants of Registration in 3D Range Data*. Ph. D thesi. Universitat Autònoma de Barcelona, 2013.
- [25] Computer Vision Laboratory (CVLAB). Pàgina web. Lloc edició: Faculté Informatique et Communications. Ecole Polytechnique Fédérale de Lausanne (EPFL). Data[20/05/2014]. Disponible a <http://cvlabwww.epfl.ch/~strecha/multiview/denseMVS.html>
- [26] 123Dcatch. Software. Desenvolupador: Autodesk inc. Disponible a: <http://www.123dapp.com/catch>.
- [27] PMVS i CMVS. Software. Desenvolupador: Yasutaka Furukawa i Jean Ponce. Disponible a: <http://www.di.ens.fr/pmvs/>.
- [28] Visual SFM. Software. Desenvolupador: Changchang Wu. Disponible a: <http://ccwu.me/vsfm/>.
- [29] Carving. Software. Desenvolupador: Ben Tordoff. Disponible a: <http://www.mathworks.com/matlabcentral/fileexchange/26160-carving-a-dinosaur>
- [30] Visual Hull. Software. Desenvolupador: K. Forbes. Disponible a: <http://www.dip.ee.uct.ac.za/~kforbes/DoubleMirror/DoubleMirror.html>
- [31] Departament of Computer Science. Middlebury College, [pàgina web]. Lloc edició: Departament of Computer Science. Middlebury College; data[14/02/2014]. Disponible a: <http://vision.middlebury.edu/mvview/>.