

Guia Pràctica 4 – Profunditat de l'escena

Curs 2021-2022 – Dept. Ciències de la Computació, UAB.

1 Context

Per resoldre alguns problemes mitjançant visió per computador a vegades no és suficient en captar informació espectral d'elements de l'escena (el seu color, textura, ...), si no que cal extreure'n informació estructural (les seva forma tridimensional, la seva posició relativa a d'altres elements,...). En aquests casos podem emprar càmeres 3D (o de rang) per captar aquesta informació. S'han desenvolupat diferents tipus de càmeres amb aquest propòsit, essent les més prevalent les basades en temps de vol (Time of Flight), les basades en llum estructurada (projecció de patró, típicament infraroig) i les basades en visió estèreo binocular.

A classe hem treballat l'estimació de la profunditat mitjançant visió estèreo. A partir de dues imatges d'una mateixa escena, preses des de diferents punts de vista, hem mostrat que si podem posar en correspondència els elements de l'escena vistos a cada imatge, en podem estimar la seva profunditat. La situació més favorable per realitzar aquesta tasca és quan els eixos òptics d'ambdues càmeres són paral·lels, i la seva posició relativa correspon a únicament un desplaçament al llarg de l'eix X, prenent el sistema de coordenades de la càmera esquerra com a referència (Fig.1). En aquest cas, donat un punt (r, c) en una imatge, el seu punt corresponent a l'altre imatge es troba a la fila r . Això facilita molt cercar correspondències entre ambdues imatges.

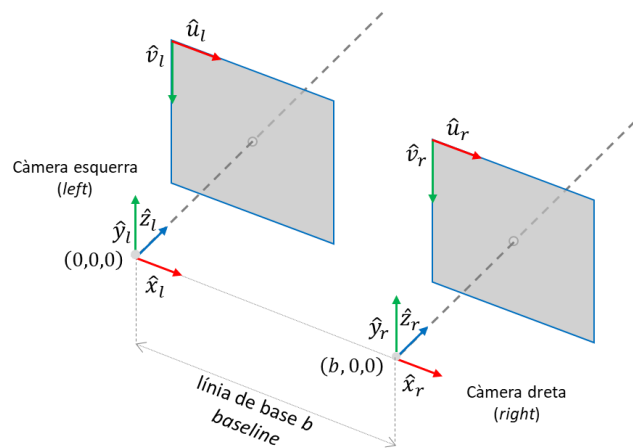


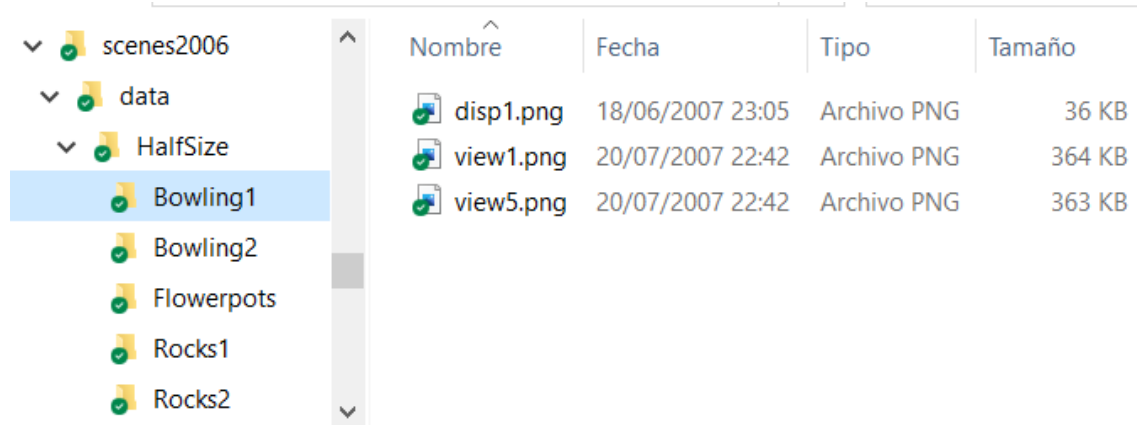
FIGURA 1 : CONFIGURACIÓ ESTÈREO FAVORABLE. REFERIT SOVINT COM SISTEMA ESTÈREO SIMPLE

Tot i que construir una càmera estèreo satisfent perfectament les condicions favorables no és factible, si calibrem la càmera per obtenir la posició i orientació relativa entre les dues càmeres, es poden transformar les imatges per obtenir les que s'haguessin capturat en les condicions favorables. Aquest procés de transformació s'anomena 'rectificació' d'un parell estèreo, i és un pre-procés que s'aplica sempre al treballar amb imatges estèreo.

En aquesta activitat es parteix d'un conjunt d'imatges estèreo rectificades, que has de processar per estimar la profunditat de l'escena a cada píxel de la imatge esquerra. El conjunt d'imatges s'ha obtingut del [Middlebury dataset](#), que és el conjunt d'imatges estèreo de referència dins la comunitat de la visió per computador. Per cada parell d'imatges, el dataset aporta la imatge de disparitat *Ground Truth* que permet quantificar el rendiment obtingut al processar les imatges.

2 Enunciat

En aquesta pràctica has de crear un Jupyter notebook per calcular la imatge de disparitat d'un parell estèreo. Les imatges a processar s'han seleccionat de la versió scenes2006 del Middlebury dataset. Aquest dataset està organitzat en carpetes (Fig.2), on cada carpeta correspon a una escena. Les imatges originals són de resolució molt gran, de manera que processarem la versió de les imatge a la meitat de la mida (HalfSize).



Nombre	Fecha	Tipo	Tamaño
disp1.png	18/06/2007 23:05	Archivo PNG	36 KB
view1.png	20/07/2007 22:42	Archivo PNG	364 KB
view5.png	20/07/2007 22:42	Archivo PNG	363 KB

FIGURA 2 : ORGANITZACIÓ DEL DATASET SCENES2006.

2.1 Repte 1: Carregar i mostrar les imatges d'una escena del dataset

Descarrega i descomprimeix l'arxiu scenes2006.zip del campus virtual. Veuràs que l'arbre de directoris que es desplega inclou 5 escenes. Dins la carpeta de cada escena hi ha els fitxers:

- view1.png: imatge esquerra del parell estèreo.
- view2.png: imatge dreta del parell estèreo.
- disp1.png: imatge de disparitat ideal (ground truth) corresponent a la imatge esquerra.

El fitxer README.txt del data et aclareix, entre d'altres coses, el següent:

The disparity images relate views 1 and 5. For the full-size images, disparities are represented "as is", i.e., intensity 60 means the disparity is 60. The exception is intensity 0, which means unknown disparity. In the half-size and third-size versions, the intensity values of the disparity maps need to be divided by 2 and 3, respectively.

En aquest repte has d'implementar una funció `get_stereo_scene`, que rebí com a paràmetre el nom d'una escena (per exemple, `Bowling1`) i retorni.

- `image_left`: la imatge esquerra en nivells de gris.
- `image_right`: la imatge dreta en nivells de gris.
- `disparity_left`: la disparitat ground truth per la imatge esquerra, tenint en compte que processem la versió half-size del data set.
- `mask_left`: imatge binària, amb valors a 1 en els píxels on la disparitat és desconeguda.

Utilitza aquesta funció per carregar les dades del parell estèreo a processar. Visualitza la informació carregada per veure com són les escenes (Fig.3).

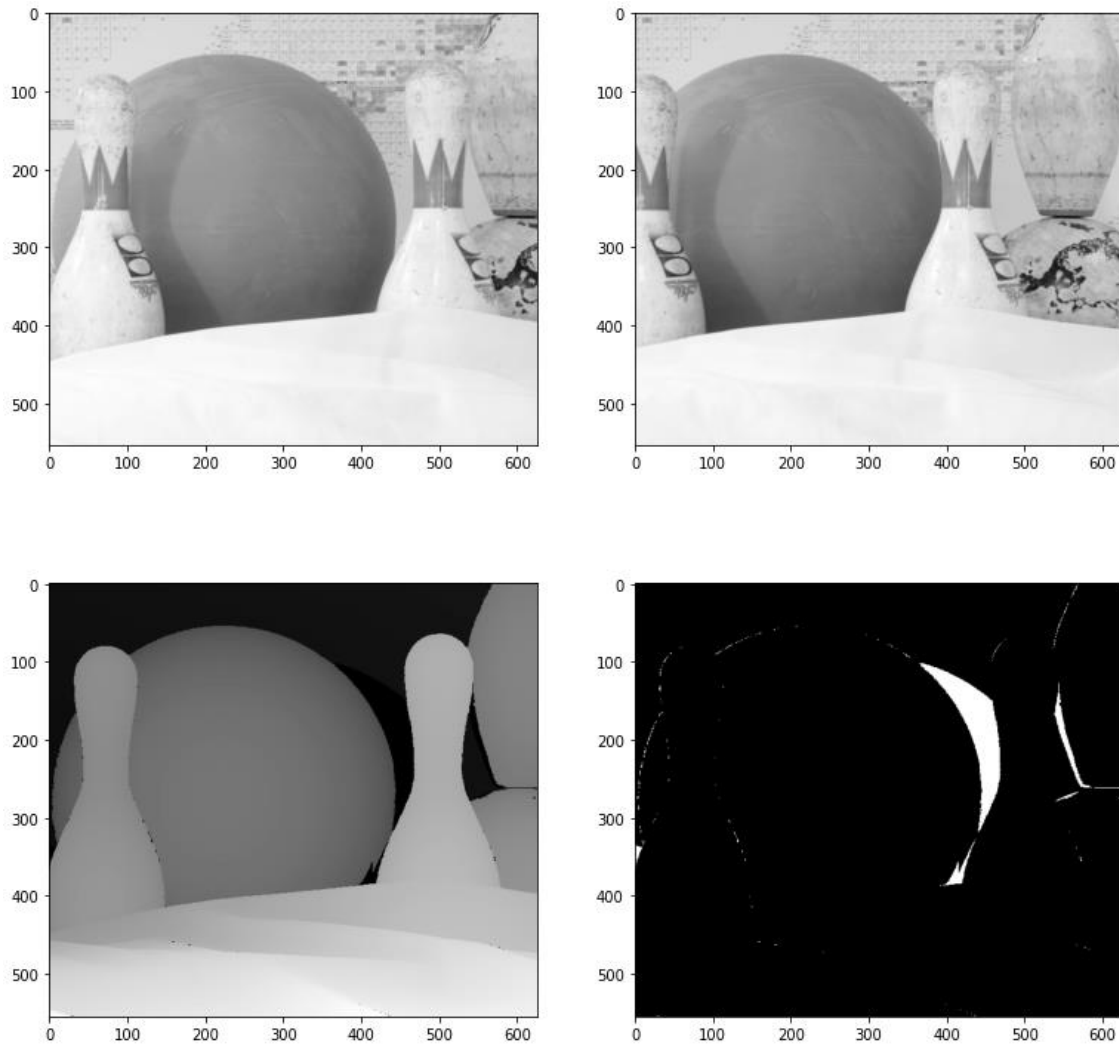


FIGURA 3: IMATGE ESQUERRA I DRETA RESPECTIVAMENT A LA FILA SUPERIOR. DISPARITAT *GROUND TRUTH* I MÀSCARA DE VALORS DE DISPARITAT NO CALCULABLES A LA FILA INFERIOR.

2.2 Repte 2: Càlcul de la disparitat esquerra aplicant *block matching*.

Per calcular la disparitat de la imatge esquerra has de trobar, per cadascun dels seus píxels, on es troba a la imatge dreta. Com que un píxel és molt poc descriptiu (un únic valor de gris), el que es fa és definir una regió quadrada centrada en el píxel (patró, block, o template), i buscar on es troba a la imatge dreta. Com que processarem imatges estèreo rectificades, la cerca només s'ha de fer a la mateixa fila de cada píxel, a la imatge dreta. D'aquest procés d'emparellar les imatges baast en patrons se'n diu *stereo* o *block matching*.

La implementació directa d'aquest procés, fent un doble-bucle per identificar cada píxel, extreure'n el seu patró, i seguidament posar-lo en correspondència a la imatge dreta és un procés molt costós. Per altra banda, hi ha una altra manera de fer aquests càlculs de manera eficient i ràpida. A continuació es descriu el procés a seguir pel cas de cercar el patró a la imatge

dreta aplicant com a mètrica de *matching* la suma de diferències absolutes (SAD), però es pot adaptar a d'altres criteris. Donat un valor de disparitat màxima `disp_max`, es realitza el següent:

- Per cada valor `d` entre 0 i `disp_max`
 - a. es genera una imatge `image_right_shifted`, resultat de desplaçar la imatge dreta '`d`' columnes cap a la dreta, completant la imatge amb zeros a la banda esquerra.
 - b. es calcula la diferència en valor absolut entre `imatge_left`, i `imatge_right_shifted`
 - c. la imatge de diferencia resultant es convoluciona amb un kernel de la mida del patró ($n*n$), amb valors iguals a $1/(n*n)$. El que s'obté com a resultat és equivalent a haver centrat un patró a cada píxel (i,j) de l'escena, i haver-lo comparat amb el patró centrat a la posició ($i,j+d$) de la imatge dreta.
 - d. El resultat es va apilant en una imatge `distance_map`, de dimensions (`#files imatge`, `# columnes imatge`, `disp_max`).

La figura 4 esquematitza el procés descrit.

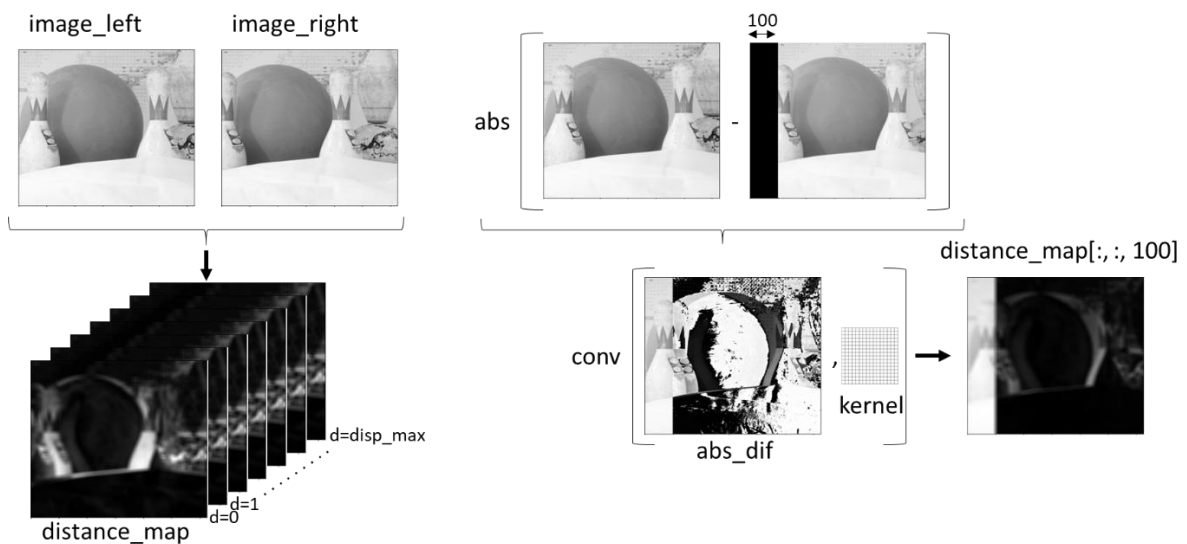


FIGURA 4 : ESQUEMA DEL PROCÉS PER CALCULAR EL MAPA DE DISTÀNCIES

En aquest repte se't demana que implementis una funció `block_matching_distance_map` que implementi aquest procés. La funció ha de rebre com a paràmetre:

- la imatge esquerra
- la imatge dreta
- una mida de patró. Controla que sigui major o igual a 3, i que sigui senar.
- un valor de disparitat màxima.

Aquesta funció ha de retornar la imatge `distance_map` resultant del procés.

Per obtenir la imatge de disparitat a partir de `distance_map` només cal buscar, per cada píxel de `distance map`, a quin canal es troba el valor de disparitat mínim (veure Fig.5). La funció `argmin` de `numpy`, degudament parametritzada, retorna aquesta informació d'una tacada.

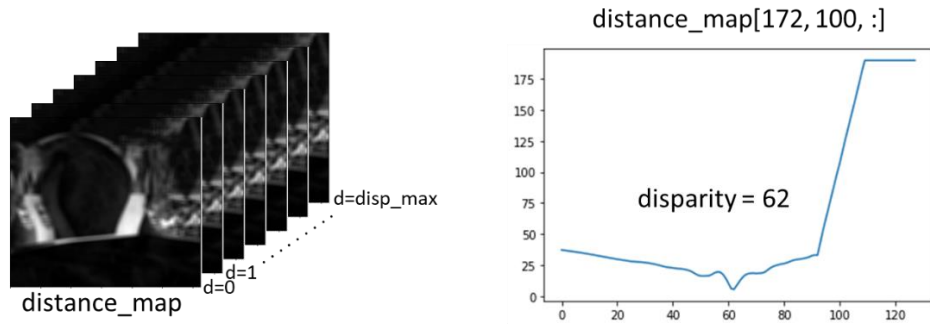


FIGURA 5 : DETERMINI DE LA DISPARITAT PEL PÍXEL DE LA IMATGE ESQUERRA (172,100)

Verifica el resultat de la implementació d'aquest mètode mostrant la disparitat ground truth de l'escena, la disparitat obtinguda amb el mètode, i la imatge del valor absolut de la seva diferència (Fig.6).

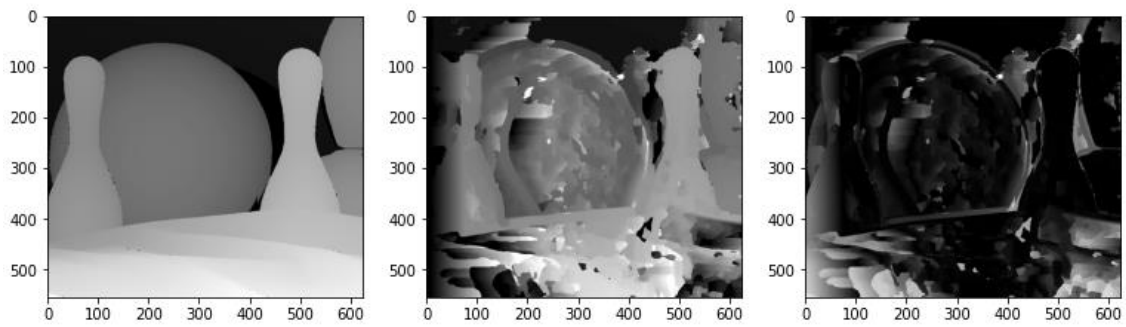


FIGURA 6: ANÀLISI QUALITATIU DEL RENDIMENT OBTINGUT EN L'ESTIMACIÓ DEL MAPA DE DISPARITAT

2.3 Repte 3: Mesura quantitativa del mètode implementat

Calcula indicadors quantitius de la qualitat de la imatge de disparitat calculada. No tinguis en compte els errors en els píxels de màscara del parell estèreo, doncs allí no es factible trobar la disparitat correcta. Mètriques que pots implementar fàcilment són la mitjana del valor absolut de la diferència (Mean Absolute Difference) entre la imatge de disparitat ground truth i l'estimada, o l'arrel quadrada de la mitjana de l'error al quadrat (Root Mean Square Error). A la [pàgina d'avaluació](#) associada al Middlebury dataset pots veure quines mètriques s'han considerat indicatives del rendiment de diferents algoritmes. Si situes el cursor sobre el nom de cada mètrica veuràs un text descriptiu de com es calcula.

2.4 Repte 4: Posar en correspondència la derivada en X de les imatges.

Afegeix un paràmetre addicional a la funció `block_matching_distance_map`, anomenat `prefilter`. Si aquest paràmetre val `True`, calcula la derivada en x de les imatges dreta i esquerra usant l'operador de sobel (convolució amb un kernel 3x3), i utilitza el resultat per calcular el mapa de distància. Analitza qualitativa i quantitativament si els resultats milloren. Intenta explicar el perquè.

2.5 Repte 5: Anàlisi de la mida del patró en el mapa de disparitat.

Un cop identificada la manera de obtenir un millor mapa de disparitat, analitza el paper que juga la mida del patró en la imatge de disparitat resultant. Intenta trobar situacions en que un patró gran és més adequat, i d'altres en què un patró més petit és més convenient.

2.6 Repte extra: Mesurar la confiança en el mapa de disparitat calculat.

El mètode per calcular la imatge de disparitat es basa en determinar-la a partir de la millor correspondència a cada píxel. A vegades però aquesta millor correspondència es pot saber que potser no serà bona. Per exemple, si la segona millor coincidència té un valor de SAD molt proper a la millor coincidència, podem tenir dubtes raonable de que haguem trobat el valor de disparitat real. Per tenir això en compte, el que es fa és determinar, per cada valor de disparitat estimat per cada píxel, quina confiança hi podem tenir. Una manera de fer-ho és calcular com de distants són, en termes percentuals, la millor i la segona millor correspondència.

Implementa aquest criteri de confiança. Un cop calculat, aplica un criteri de threshold a la imatge de confiança per separar les disparitats fiables de les dubtoses. Avalua el rendiment del mètode implementat, considerant únicament els píxels on aquest criteri determina estar segur de la disparitat estimada. Analitza-ho de manera quantitativa i també qualitativa (mostra imatges on es vegin només els valors de disparitat “de confiança”).