

Treball final de grau

Estudi: Grau en Enginyeria Informàtica

Títol: Detecció de correspondències en imatges submarines

Document: Resum

Alumne: Jan Isorna Domènech

Tutor: Rafael Garcia Campos

Departament: ATC

Àrea: ATC

Convocatòria (mes/any): Setembre 2020

Degut al tancament de centres i laboratoris de la UdG a causa de la pandèmia del COVID-19, s'ha hagut de canviar lleugerament la idea inicial del projecte, que era realitzar un estudi de la precisió i robustesa de diferents mètodes de l'estat de l'art per la cerca de correspondències entre parelles d'imatges submarines, obtingudes a partir d'una càmera estereoscòpica del Centre d'Investigació en Robòtica Submarina (CIRS) de la UdG. Segons la planificació inicial, es volia muntar un escenari submarí artificial en un tanc d'aigua, posant sorra, petxines, coralls, etc., i fixant les càmeres en un suport rígid. La intenció era omplir-lo d'aigua, adquirir dues imatges, i utilitzar aquestes imatges com a "ground truth", per a continuació anar afegint llet de 20 ml en 20 ml per simular diferents graus d'absorció de la llum per part del medi. En el seu lloc, a causa de la impossibilitat d'accedir al CIRS, i per tant d'obtenir les imatges desitjades, en lloc de realitzar l'estudi sobre imatges submarines, s'ha dut a terme utilitzant imatges "estàndard" del *Middlebury Dataset 2014*, per les quals els autors proporcionen també el *ground truth*.

1. Introducció

La detecció de correspondències en parelles d'imatges és un problema molt rellevant en el món de la visió per computador, donat que té aplicacions en la detecció de moviment, compressió de vídeo, reconstrucció 3D, etc. Tot i això, es tracta d'un problema difícil, i els diferents mètodes no presenten resultats òptims en totes les situacions, malgrat en els darrers anys han aparegut noves tècniques que ofereixen un rendiment cada vegada millor.

La motivació d'aquest treball es basa en la creença de que deixant de seguir el paradigma que regeix a la majoria dels mètodes tradicionals -els més utilitzats a dia d'avui- i investigar si els nous mètodes i tècniques basats en altres tecnologies, com Xarxes Neuronals Profundes (Deep Learning), podrien ajudar a trobar millors solucions per aquest problema. A més, les tècniques tradicionals, semblen haver-se estancat, i els seus resultats, tot i ser satisfactoris, sembla que no tenen gaire marge de millora. En canvi, canviant el punt de mira cap a altres paradigmes com el que segueixen les tecnologies comentades anteriorment, Xarxes Neuronals Convolucionals (CNN, pel seu acrònim en anglès), més enllà de validar si actualment ja es poden aconseguir resultats millors que els que ofereixen els mètodes tradicionals, també creiem que, al ser unes tecnologies que cada cop sumen més mirades i que cada vegada més persones les estudien i hi treballen, amb el temps podrien arribar a millorar molt més i permetre trobar solucions més eficients i robustes per al problema plantejat.

2. Objectius

L'objectiu principal d'aquest projecte és comparar el rendiment dels mètodes tradicionals més usats fins l'actualitat, amb mètodes basats en CNNs. Les CNN, a diferència dels mètodes tradicionals, utilitzen Xarxes Neuronals per a detectar els punts d'interès de les imatges i crear-ne descriptors. En els darrers anys les CNN han demostrat un rendiment molt superior a les tècniques tradicionals en l'àmbit de reconeixement d'objectes, però aquest extrem no ha estat demostrat per la detecció de correspondències entre parells d'imatges.

La nostra hipòtesi a validar és si les CNNs permetrien resoldre el problema de la correspondència amb un rendiment millor que les tècniques tradicionals com per exemple seria SIFT (Scale-Invariant Feature Transform), una de les millors tècniques existents en el camp de la visió per computador. Cal esmentar que l'article de SIFT publicat a l'International Journal of Computer Vision al 2004, és l'article més citat de tota la història en el camp de la visió per computador, amb més de 58.000 cites a Google Scholar.

3. Estructura de la memòria

L'estructura de la memòria d'aquest projecte està organitzada de la següent manera.

El capítol 2 descriu la metodologia que s'ha emprat per dur a terme aquest treball. A continuació, el capítol 3 presenta la planificació que hi ha hagut darrere aquest projecte per tal de poder-lo realitzar correctament i entregar-lo dins el termini especificat. Tanmateix, aquest capítol fa referència als contratemps sorgits durant el seu transcurs, i les solucions adoptades per superar-los.

El capítol 4 presenta l'estat de l'art en mètodes de correspondència entre parelles d'imatges. Aquest apartat servirà per donar a conèixer al lector el funcionament de tots els mètodes que intervindran en l'estudi realitzat en aquest treball, tant dels mètodes tradicionals (SIFT, SURF, KAZE, ORB...), com dels mètodes més rellevants basats en CNNs (ContextDesc, SuperPoint, D2-Net).

Seguidament, el capítol 5 detalla els diferents requisits, tant software com hardware, que haurà de tenir una màquina per poder executar sense problemes el codi d'aquest treball i obtenir-ne els resultats òptims.

El capítol 6 presenta a continuació l'anàlisi del sistema. Aquest capítol servirà per comentar quina és la visió del problema a solucionar i de quina forma s'afrontarà.

El capítol 7 s'encarrega del disseny i implementació del sistema. En aquest apartat es comentarà la solució adoptada després de fer l'anàlisi de problema, la forma en la qual s'ha implementat i què s'ha utilitzat per fer-ho, i els problemes trobats durant la implementació. Fruit d'aquesta implementació, el capítol 8 mostra finalment els resultats obtinguts pels diferents mètodes, fent un especial esforç per jerarquitzar-los davant diferents situacions entre les parelles d'imatges (parallax, rotacions i canvis d'escala). Aquest capítol servirà per entendre millor les potencialitats i limitacions de cadascun dels mètodes, i ens portarà de forma natural al darrer capítol, on es plasmen totes les conclusions extretes dels resultats obtinguts.

4. Software realitzat

Per aconseguir els objectius proposats, s'ha implementat un software escrit en Python, pensat per córrer sobre el sistema operatiu Ubuntu 18.04.4 LTS, el qual, utilitzant diverses llibreries i components. Aquest software té com a finalitat, l'obtenció de diferents gràfics que mostraran la precisió dels diversos mètodes de detecció de correspondències utilitzats, i la seva robustesa a canvis d'escala i rotació.

Com que aquest software s'utilitzarà aplicant-hi un dataset de parells d'imatges força extens, i això comportarà una gran generació de d'informació, mitjançant la implementació d'un altre codi en Python i les dades guardades de cada execució, es podran generar diversos dashboard per simplificar-ne la seva lectura i comprensió.