

Anàlisi i comparativa de mètodes per a l'augment de la resolució temporal de seqüències d'imatges multi-espectre

Daniel-Pau Azemar i Carnicero



Resum– La freqüència de revisita d'un satèl·lit sobre una superfície determinada de la terra pot esdevenir restringit pel període d'aquest. Els especialitzats en imatge solen ésser multi espectrals (MS). Existeix doncs un interès en millorar la resolució temporal de seqüències d'imatges MS un cop capturades. Els mètodes d'interpolació temporal aporten la solució al problema, però aquests estan dissenyats per interpol·lar en escala de Grisos o RGB. En aquest treball es realitzarà un estudi de diferents mètodes de l'estat de l'art, s'avaluarà el seu rendiment i s'estudiarà com adaptar-los per processar imatges RGB i MS. Per fer-ho, també es genera un conjunt de dades MS per poder realitzar-hi proves d'interpolació temporal. Finalment, comprovem com fent una mitja del flux òptic en cada banda d'una imatge processada amb PCA s'obtenen els millors resultats. També observem com les noves tècniques basades en Deep Learning aporten les solucions més òptimes per a interpolació MS, tot i estar entrenades en RGB.

Paraules clau– Interpolació temporal, augmentant resolució temporal, imatges multi-espectre, Anàlisi de Components Principals (PCA), Deep Learning, Xarxa Neuronal, Flux Òptic.

Abstract– The revisit frequency of a satellite on a given surface of the earth may become restricted to its period. Satellites specialized in capturing images tend to be multi-spectrum (MS). Therefore, there is an interest in improving the temporal resolution of MS sequences once captured. Temporal interpolation methods provide the solution to the problem, but these are designed to interpolate over Gray or RGB scales. In this project, a study of different methods of the state of the art will be carried out, its performance will be evaluated and it will be studied how to adapt them to process RGB and MS images. To do this, an MS database is also generated to perform tests of temporal interpolation. Finally, we conclude that the best results are obtained when the average of the optical flow in each band of an image is processed with PCA. We also see how new techniques based on Deep Learning provide the optimal solutions for MS interpolation, despite being trained in RGB.

Keywords– Temporal interpolation, increasing temporal resolution, multi-spectral imagery, Principal Component Analysis, Deep Learning, Neural Network, Optical Flow.

1 INTRODUCCIÓ

EN els darrers anys molts telèfons intel·ligents han aparegut al mercat amb la capacitat de gravar imatges a càmera lenta. Per utilitzar aquestes capacitats s'acostuma a perdre qualitat en les imatges perquè l'obtura-

dor de la càmera ha d'obtenir més informació per unitat de temps. Amb mètodes de interpolació temporal es pot arribar a mantenir la qualitat de la imatge, sense la necessitat d'obtenir més fotogrames per segon a través de la càmera. Això és degut al fet que aquests mètodes aconsegueixen la imatge intermèdia entre dos fotogrames d'una seqüència, augmentant així els fotogrames per unitat de temps. Un exemple on s'utilitzen aquestes tècniques n'és la indústria tecnològica dels televisors, on es pot veure com alguns d'alta gamma duen per defecte aquestes tècniques amb una opció anomenada "moviment fluid" o la "interpolació del moviment". Aquests televisors augmenten doncs la freqüència d'imat-

- E-mail de contacte: dani.azemar@gmail.com
- Menció realitzada: Enginyeria de Computació
- Treball tutoritzat per: Dani Ponsa (Centre de Visió per Computació)
- Curs 2017/18

ges per segon si el senyal d'entrada de reben no és prou alta. Hi ha un gran interès d'utilitzar aquestes tècniques en aquells llocs on les dificultats per aconseguir un mostreig d'imatge més elevat es pot veure reduït per factors tant interns com externs a la càmera. Un exemple en són les imatges per satèl·lit, on la freqüència d'obtenció d'imatges es pot veure afectada per l'amplada de banda de l'aparell, el període de l'òrbita, etcètera.

S'han realitzat molts estudis sobre el comportament dels algorismes de interpolació temporal en imatges RGB o grisos, i recentment any rere any surten nous algorismes que duen a terme aquesta tasca de manera més eficient. Però, si s'han d'utilitzar les tècniques emprades per a les imatges RGB, en imatges multi espectrals (MS) com les de satèl·lit, tindrem el mateix rendiment? Què podem fer per aconseguir un rendiment més satisfactori en aquest cas? En aquest treball es proposen diferents maneres per treballar en amb les imatges MS, i s'avaluen diverses tècniques de interpolació per imatges RGB i MS.

Aquest document està dividit en diverses seccions. A la secció 2 s'hi especifiquen els objectius del TFG. A la secció 3, una visió general de l'estat de l'art en la interpolació temporal, on s'hi expliquen també els mètodes que s'utilitzen en el treball. A la secció 4, una breu descripció de com s'ha desenvolupat la creació del software utilitzat així com les eines emprades. A la secció 5 s'explica com està format la base de dades del projecte. A la secció 6 s'hi defineix l'estructura general del codi així com les mètriques avaluatives dels resultats. A la secció 6 una anàlisi dels resultats, i darrerament, les conclusions i les línies futures del treball a les seccions 7 i 8 respectivament.

2 OBJECTIUS

L'objectiu principal del projecte és *millorar la resolució temporal d'una seqüència d'imatges multi espectrals*. Per assolir l'objectiu principal, s'han identificat

- Estudiar diferents mètodes d'interpolació temporal i estat de l'art d'aquests tant clàssics com recents per poder valorar el rendiment amb imatges MS.
- Analitzar maneres de tractar dades MS en els algorismes d'interpolació temporal ja que aquests treballen amb un nombre fix de bandes (o bé una banda, o bé 3 bandes RGB).
- Implementar mètriques d'avaluació que ens permetran crear un rànquing entre mètodes.
- Establir un conjunt de dades sobre on poguéssim avaluar els diferents algorismes.
- Generació d'un Framework d'Avaluació que ens permeti realitzar experiments dinàmicament.

Podeu trobar un arbre d'objectius on s'hi pot veure l'organització amb més detall a l'Apèndix, Fig. 11.

3 ESTAT DE L'ART

La interpolació és un problema clàssic en l'àmbit de les matemàtiques i la computació que té com a objectiu estimar informació desconeguda a partir d'altra ja disponible [1]. En

referència a les imatges, quan es parla d'interpolació pot fer referència a la interpolació espacial, que tracta d'augmentar la resolució en píxels per imatge, i la interpolació temporal, que tracta d'augmentar la resolució en fotogrames per unitat de temps. Aquest darrer és al que es fa referència en el treball quan es parla d'interpolació. Matemàticament parlant i de manera general, pot definir-se la interpolació entre dues imatges com s'expressa a l'Eq. (1)

$$f(I_t, I_{t+\Delta t}) \rightarrow \tilde{I}_{t+\frac{\Delta t}{2}}, \quad (1)$$

on I_t i $I_{t+\Delta t}$ són les imatges a l'instant del temps t i $t + \Delta t$ (primera i segona imatge respectivament), i $\tilde{I}_{t+\frac{\Delta t}{2}}$ la imatge interpolada.

3.1 Mètodes de Interpolació

La interpolació temporal s'ha tractat de moltes maneres diferents, en l'estudi realitzat se n'han identificat de 3 tipus: Mètodes Aritmètics, Extractors de Flux Òptic, i de Deep Learning.

3.1.1 Aritmètics

En aquests mètodes la interpolació s'obté a partir d'interpol·lar el valor d'un píxel emprant una combinació lineal fixada dels píxels de la imatge.

- **Lineal (L):** Tècnica bàsica d'interpolació. Estableix la imatge interpolada com la mitjana de I_t i $I_{t+\Delta t}$. Aquesta tècnica s'ha implementat per establir un nivell de rendiment bàsic, que serveixi de referència per analitzar-ne el de propostes més sofisticades.

3.1.2 Basat en Flux Òptic

Es podria dir que els mètodes d'extracció del flux òptic han estat els més utilitzats per dur a terme la interpolació temporal, fet que podria canviar degut a la irrupció de les tecnologies Deep Learning, deixant aquests extractors enrere [2]. Extreuen el flux òptic de la imatge per a posteriorment sintetitzar el píxel entre els frames I_t i $I_{t+\Delta t}$. Dit d'altra manera, intenten abordar el problema a partir del moviment aparent de cada píxel de la imatge, assumint

$$I_t(x, y) = I_{t+\Delta t}(x + u, y + v), \quad (2)$$

on u, v és el vector de moviment aproximat per al píxel x, y de la imatge. Després es reconstrueix la posició d'aquest a la imatge intermèdia $I_{t+\frac{\Delta t}{2}}$ utilitzant aquest flux òptic calculat anteriorment tal que

$$\tilde{I}_{t+\frac{\Delta t}{2}}(x, y) = I_t(x + \frac{u}{2}, y + \frac{v}{2}). \quad (3)$$

Es pot extraure el flux òptic amb mètodes clàssics [3, 4], o bé utilitzant algorismes més avançats [5], o bé amb les noves tecnologies de Deep Learning [6, 7]. S'han escollit les tècniques tenint en compte els diferents tipus d'interpolació en resolució temporal que hi ha. Aquests tipus són o bé **locals** o bé **globals**. Els locals treballen tenint en compte el veïnat de píxels definits amb una finestra de mida N . Són més bons tractant el soroll de la imatge. El segon tipus treballen tenint en compte tota la imatge, permetent propagar els vectors de flux òptic cap a zones sense textura on no es pot calcular el flux òptic.

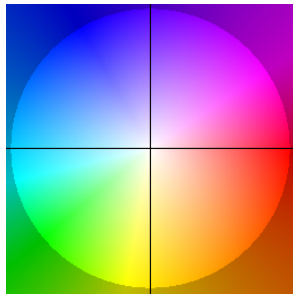


Fig. 1: Codi de colors per representar el moviment d'una escena. Moviment cap a l'esquerra es representa amb color blau cian, i cap avall amb color groc.

- **Lucas-Kanade** (LK, [3]): És una tècnica **local** que assumeix que el flux és pràcticament constant al voltant dels píxels seleccionats i soluciona les equacions del flux òptic utilitzant el mètode de mínims quadrats. Dit d'altra manera, calcula el flux òptic de cada píxel dins d'un bloc de píxels solucionant un sistema d'equacions per cada píxel. És robust contra el soroll. S'ha seleccionat per la importància i impacte que ha tingut. És considerat un dels mètodes clàssics per al càlcul del flux òptic.
- **Horn-Schunck** (HS, [4]): Mètode **global** que introdueix una restricció global a la imatge. S'ha seleccionat a causa de rellevància en el camp junt amb l'anterior (mètode local, i aquest és un mètode global), ja que també és considerat un dels mètodes clàssics per al càlcul de flux òptic.
- **Farneback** (F, [8]): Mètode **global** per a calcular el Flux Òptic (Optical Flow) entre dues imatges. Es basa a estimar el desplaçament dels píxels utilitzant tensors d'orientació.
- **TVL1** (T, [9]): Mètode **global** que es basa en la minimització de l'espai funcional que conté les dades utilitzant la norma de vector L_1 i la regularització del terme utilitzant la variació total del flux òptic. La característica principal d'aquest algorisme és que permet discontinuïtats en el mapa de flux, sent així més robust al soroll que el mètode clàssic Horn-Schunck.

Per poder visualitzar aquests mapes de flux òptic, s'hi transformen els vectors de cada píxel en un color RGB com el mostrat a la Figura 1.

3.1.3 Deep Learning

També hi ha treballs on proposen la interpolació d'aquesta imatge intermèdia utilitzant les tecnologies de Deep Learning on no es necessita calcular explícitament el flux òptic [5, 10, 11]. Per a aquest treball hem seleccionat el següent mètode:

- **Adaptive SepConvNN** (S, [10]): Mètode **global** per resoldre el problema de la interpolació amb una Xarxa Neuronal convolucional. S'ha seleccionat per comparar els mètodes anteriors amb un de nou que utilitza Deep Learning. Retorna directament la imatge interpolada.

Satèl·lit	Bandes	Resolu.	Mostreig
Meteosat [12]	12-b	1 km	15 min
WorldView-3 [13]	8-b	1.24 m	1-4 dies
SPOT 6 & 7 [14, 15]	RGB&NI	6m	1 dia
BlackBridge [16]	6-b	5m	1-5 dies

TAULA 1: Mostra d'alguns satèl·lits operatius i les seves opcions multi-banda

3.2 Imatges Multi Espectrals (MS)

Les imatges multi espectrals són imatges multibanda que recullen informació a diverses freqüències de l'espectre electromagnètic. Així doncs, les imatges RGB són un tipus d'imatge multi espectral. En aquest treball però s'utilitzarà la denominació multi espectral per a aquelles imatges que, tant incloent com si no les bandes de l'espectre RGB, inclouen altres bandes de l'espectre com ultraviolat o infraroig.

Un exemple d'imatges MS en tenim als satèl·lits que orbiten la terra captant imatges. Aquests tenen un rang de resolució temporal molt divers. Això pot ser degut a la tecnologia d'aquestes càmeres. Un altre factor n'és la distància o la resolució espacial amb la qual es prenen les imatges: Meteosat [12] és una constel·lació de quatre satèl·lits geoestacionaris (orbiten la terra tenint sempre la mateixa visió d'aquesta). Obtenen una imatge d'Europa amb una freqüència aproximada d'uns 15 minuts. Així doncs, s'aconsegueix una imatge a temps real on cada píxel acaba representant diferents kilòmetres. Per altra banda, altres satèl·lits com el WorldView-3 [13], obtenen les imatges a major resolució espacial, però la porció de la terra que captura és molt reduïda i sovint no pot tornar a capturar la mateixa zona fins passades unes hores o dies (la freqüència de mostreig d'imatges per un mateix punt dependrà del temps que trigui a tornar a passar per aquella mateixa zona). En resum, molts factors acaben afectant el temps de revisita d'una mateixa zona observada. En aquest context, la interpolació temporal permet traslladar la informació d'una fotografia capturada a un instant del temps diferent del d'adquisició. La Taula 1 disposa d'una mostra dels instruments satel·litaris especialitzats en captació d'imatge.

4 METODOLOGIA

La metodologia utilitzada per desenvolupar aquest treball està inspirada en les Àgils. Aquestes metodologies són àmpliament utilitzades en la informàtica, permeten mantenir un progrés constant del treball alhora que s'afegeixen o modifiquen tasques a realitzar. Cada setmana es realitza una reunió amb els tutors del treball que junts amb l'estudiant desenvolupador de la feina comparteixen els avenços i s'avalua el treball realitzat. Aquí s'hi poden afegir o treure tasques segons el compliment o no de la previsió, i es parlarà de la feina necessària per a la setmana vinent.

Per poder dur a terme un seguiment de les tasques, s'ha decidit utilitzar una eina anomenada Trello [17]. És un servei simple que et permet organitzar els teus recordatoris sobre un taulell. S'utilitza per veure l'estat de les tasques i tenir un flux de feina constant i revisable en el temps.

Per mantenir un registre de la feina realitzada i així comprovar l'estat del treball al llarg del temps, s'utilitza un re-

positori Git [18] on s'hi penegen els progressos del desenvolupador respecte al codi i els mateixos Informes de Control.

El programa s'ha desenvolupat amb Python [19] a causa de la facilitat d'ús de llibreries especialitzades en visió, així com OpenCV, Sklearn, Matplotlib, Scipy, etcètera; l'ètica *open-source* que defensa i la comoditat de l'estudiant per treballar-hi. El disseny del software és modular per permetre l'intercanvi entre els mètodes d'avaluació i de tractament de les imatges. Per fer-ho, la forma que prendrà el programa és de *pipeline* on a cada mòdul passarà les dades amb uns inputs i outputs prèviament especificats.

5 DESENVOLUPAMENT

5.1 Conjunt de Dades

S'ha generat una base de dades amb seqüències d'imatges diverses en relació al seu flux òptic per a l'anàlisi dels algorismes que s'implementen. Amb l'objectiu de tenir un conjunt de dades que abasteixi un ampli ventall de possibilitats sense haver de generar-los tots (infinitat de casos), s'ha establert unes guies a seguir. La primera és que les seqüències recopilades no poden limitar-se a imatges d'un nombre fix de bandes, ja que es busca trobar un mètode genèric per tractar imatges de múltiples bandes, i do pretén trobar la manera de tractar un nombre indeterminat de bandes espectrals de la millor forma possible. També s'ha tingut en compte la diversitat pel que fa al moviment d'elements a l'escena, ja que el problema no pot limitar-se a un sol cas. S'ha identificat diversos tipus de flux òptic de manera general, que va des de combinacions d'objectes tan estàtics com en moviment, sobre fons o bé estàtics o bé en moviment, així com la càmera també estàtica o en moviment. A la Taula 2 s'identifiquen aquests moviments que poden provocar flux òptic a l'escena, així com un codi identificador.

Codi	Descripció
Moviment de la càmera	
MF	Càmera fixa.
MA	Càmera en moviment sense direcció definida (gravació a mà alçada).
MT	Translació de la càmera unidireccional.
MR	Càmera en moviment de rotació.
MZ	Càmera ampliant .
Moviment dels objectes	
OC	Cap Objecte.
OU	Unidireccionals.
OM	Multidireccionals.

TAULA 2: Identificació de casos que poden provocar flux òptics diversos.

S'ha decidit treballar sobre conjunts de dades ja estandarditzades, facilitant així la creació d'un *dataset* divers i amb multitud de treball comunitari previ que ajuda a l'extracció de conclusions dels experiments d'aquest projecte. A la Taula 3 es llisten els conjunts de dades utilitzats en el TFG, amb un resum del seu contingut. Per a datasets multi espectrals tenim Surveillance, que son seqüències estil videovigilància a 7 bandes i Paintings, que son fotografies a

23 bandes d'on podrem generar un dataset sintètic. Per a imatges multispectrals manquen els flux òptics MT, MR amb OU i OM. Així doncs, s'ha creat un *dataset* propi que amplia les mancances de flux òptic en imatges *multispectrals*.

Per a la generació del **dataset propi** s'ha utilitzat la càmera multi espectral de 4 bandes (NIR, G, R, RE), Parrot Sequoia [22]. S'ha realitzat un total de 3 seqüències aprofitables que són translacions cap endavant (guijon, EE-CVC, Cantonada). Les seqüències tenen d'entre 7-12 fotogrames cadascuna.

5.2 Framework d'Avaluació

S'ha desenvolupat un entorn de procés, per permetre processar de manera automàtica els diferents datasets amb els diferents mètodes considerats en el TFG, avaluant-los amb diferents mètriques.

El disseny del programa es basa en el principi de la modularitat. Per aquesta raó, la informació flueix per diferents mòduls successivament. Aquesta estructura s'anomena *pipeline*. El pipeline general està separat en quatre mòduls, cadascun realitzant funcions específiques, amb la intenció de poder ser substituïts per canviar el comportament de la interpolació, acabant avaluant diferents *pipelines* de processament de dades. El flux d'execució del codi és el representat en la Figura 2. Una breu descripció de cada mòdul n'és:

- **Preprocessament:** Aquesta fase correspon al tractament de la imatge abans no hagi de passar per l'algorisme d'interpolació. Serveix sobretot per adequar les dades a un espai de treball on puguin ser tractades correctament per l'algorisme. Per exemple, un algorisme que interpoli a 3 bandes, no podrà treballar amb una imatge d'11 bandes, així doncs, el preprocessament haurà de transformar la imatge de diferents formes perquè puguin tenir cabuda a la interpolació.
- **Interpolació:** És la fase on es duu a terme la recreació de la imatge situada temporalment entre els fotogrames previs (I_t) i posteriors ($I_{t+\Delta t}$). És a dir, es genera la imatge interpolada ($I_{t+\frac{\Delta t}{2}}$).
- **Postprocessament:** Aquí es desfà el procés de preprocessament si és necessari per visualitzar-ho o avaluar-ho.
- **Avaluació:** Aquí s'hi duu a terme el càlcul de l'error entre la imatge recreada i la referència (*groundtruth*).

Alias	Bandes	Moviment	
		Càmera	Objectes
Middlebury-RGB [2]	RGB	MA, MT	OU, OM
Middlebury-G [2]	Gray	MA, MT	OU, OM
Surveillance [20]	7b	MF	OC, OU, OM
Paintings-Synth [21]	23b	MT, MR, MZ	OC
Meteosat	11b	MF	OM
Propi-MS	4b	MT, MR	OC

TAULA 3: Conjunt de Bases de Dades que s'ha utilitzat per al projecte. Els nombres segueixen la codificació de la Taula 2.

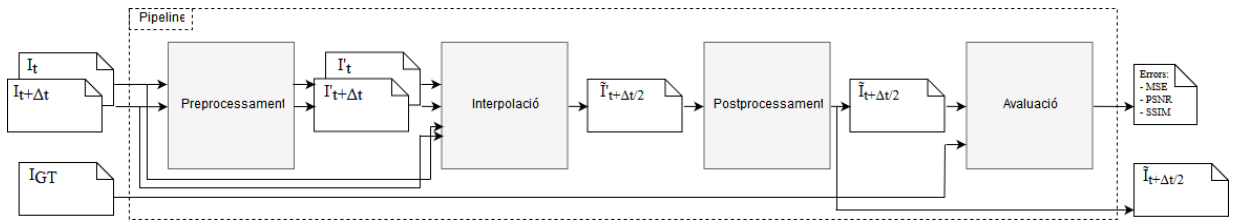


Fig. 2: Organització dels diferents mòduls del Pipeline General.

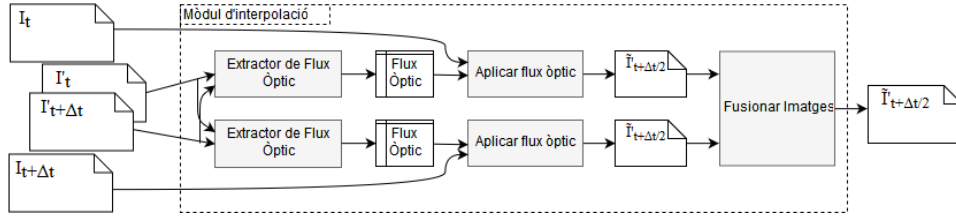


Fig. 3: Funcionament Bidireccional d'un pipeline d'interpolació basat en Flux Òptic.

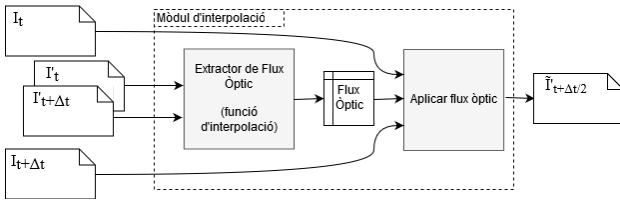


Fig. 4: Funcionament Simple (o clàssic) d'un pipeline d'interpolació basat en Flux Òptic

5.2.1 Pipeline intern de Interpolació

Respecte al mòdul d'interpolació, aquest dota d'un altre pipeline intern si l'algorisme és extractor de flux òptic. Aquest pipeline intern s'ha abordat de dues maneres diferents: la forma clàssica, representada a la Figura 4, i la forma bidireccional, representada a la Figura 3. La forma clàssica només duu a terme una extracció del flux òptic amb els mètodes d'interpolació explicats a l'apartat 3.1.2, i després reconstrueix la imatge amb aquest flux òptic. La manera en com s'aplica el flux òptic sobre les diferents bandes es pot fer de diverses maneres. Es pot extreure el flux òptic només d'una banda i aplicar-les a totes, o bé extreure'n de cada banda i fer-ne una mitjana, o bé extreure'n de cada banda i aplicar-la a cada una individualment.

El pipeline bidireccional, per altra banda, duu a terme dues interpolacions: endavant i enrere (dels frames I_t a $I_{t+\Delta t}$ i viceversa). Posteriorment, amb les dues imatges generades es realitza un procés de fusió amb el flux òptic de cadascuna.

Aquesta fusió de les dues imatges del pipeline bidireccional té en compte els píxels que s'han desplaçat a l'hora d'aplicar el flux òptic tal com s'explica en l'equació 2. Quan aquests píxels es desplacen (és a dir, un objecte de l'escena es mou), deixa un forat buit on pot no anar cap altre píxel, quan això passa diem que hi ha hagut oclusió. Aleshores, interessa emplenar aquests forats amb els valors de l'altra imatge. Els valors no són exactes (el flux òptic retornat per les funcions no es discret), sinó que a un píxel del destí pot ser emplenat per més d'un píxel. Així doncs, es realitza una matriu de votació per a les dues imatges i tot seguit se sumen tenint el pes de cada píxel respecte al total d'aquest (es

normalitza la imatge resultant). Matemàticament parlant, la imatge interpolada es genera fusionant les imatges I_t i $I_{t+\Delta t}$ tal que

$$\tilde{I}_{t+\frac{\Delta t}{2}}(x, y) = f_1 \tilde{I}_{t+\frac{\Delta t}{2}}^1(x, y) + f_2 \tilde{I}_{t+\frac{\Delta t}{2}}^2(x, y), \quad (4)$$

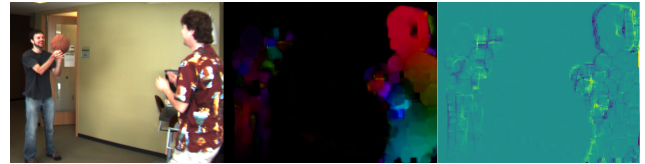
on

$$f_1 = \frac{V_{I_t \rightarrow I_{t+\Delta t}}(x, y)}{V_{I_t \rightarrow I_{t+\Delta t}}(x, y) + V_{I_{t+\Delta t} \rightarrow I_t}(x, y)} \quad (5)$$

$$f_2 = \frac{V_{I_{t+\Delta t} \rightarrow I_t}(x, y)}{V_{I_t \rightarrow I_{t+\Delta t}}(x, y) + V_{I_{t+\Delta t} \rightarrow I_t}(x, y)}$$

essent $V_{I_t \rightarrow I_{t+\Delta t}}$ la matriu de votació resultant del flux òptic de la imatge I_t a $I_{t+\Delta t}$ i $V_{I_{t+\Delta t} \rightarrow I_t}$ la matriu de votació resultant en direcció inversa. $\tilde{I}_{t+\frac{\Delta t}{2}}^1$ és la imatge interpolada de $I_t \rightarrow I_{t+\Delta t}$ i $\tilde{I}_{t+\frac{\Delta t}{2}}^2$ la imatge interpolada en direcció inversa.

A la Figura 5 veiem una un exemple de matriu de votació a partir del flux òptic extret.

Fig. 5: Exemple d'extracció de matriu de votació. D'esquerra a dreta: I_{GT} , Flux Òptic de I_t a $I_{t+\Delta t}$, $V_{I_t \rightarrow I_{t+\Delta t}}(x, y)$

Un cop s'ha realitzat tot el procés pel *pipeline*, els resultats generats es guarden en un fitxer CSV. Aquest conté la informació del Dataset utilitzat, així com el *frame* exacte interpolat, amb quin mètode s'ha realitzat la interpolació, el tipus de bandes d'aquest *frame* (RGB, 5b, 7b, etc.) i els valors de MSE, PSNR i SSIM resultants. Tot el codi generat en aquest apartat i anteriors es poden trobar al GitHub del treball [18]. Així doncs, en el següent apartat analitzarem els resultats obtinguts.

5.2.2 Adaptació dels algorismes a imatges MS

Un dels factors a tenir en compte, és com realitzar la construcció de la imatge interpolada amb els mètodes seleccionats que no treballen en un nombre de bandes indeterminat. Generalitzar el mètode lineal a N bandes és directe, però per a la resta de mètodes, cal establir com usar-los en imatges multibanda.

Per a les tècniques extractores de flux òptic, treballen sobre una banda individual i extreuen el flux òptic aparent, d'on després es reconstrueix la imatge tal com es descriu a l'equació 3. Aquesta banda pot ser una qualsevol de la imatge, però es pot arribar a seleccionar una on no es disposi de gaire informació aparent (potser la imatge és només vermella). Així doncs, és important seleccionar bé la banda, ja que aquesta ha de contenir la màxima informació aparent possible de la imatge. Podem:

- Per al cas d'imatges en color, transformar l'espai de color a un altre de ja conegut (e.g. de RGB a HLS, HSV, YCC, etc.) i seleccionar-ne la més representativa de la textura de l'escena (HLS seria L, HSV seria V, etc)
- Realitzar un Anàlisi de Components Principals (PCA) que trasllada tota la informació en un altre espai i dimensions, fent que la informació es concentri en un nombre reduït de capes.

Un cop obtingut el flux òptic, es pot aplicar aquest sobre les bandes transformades i després passar-ho un altre cop a l'espai de dades original, o bé aplicar-ho directament sobre les bandes sense transformar.

La tècnica de Deep Learning usada en aquest TFG, en canvi, treballa sobre una RGB. A l'estar entrenada en aquest espai pot ser que retorni valors inesperats en introduir altres bandes de l'espectre o d'altres representacions del color (com PCA). Això sí, s'ha de veure quin és el rendiment segons la manera d'introduir les capes. Les opcions que hem considerat són:

- Replicar (SRep) cada capa una a una per formar una imatge de 3 capes, passar-la per la xarxa, i quedar-nos amb la capa central per reconstruir la imatge banda per banda.
- Per cada banda, seleccionar les bandes immediatament anteriors i posteriors per formar una imatge de 3 capes i:
 - Reconstruir la imatge només amb la capa central de cada imatge retornada (S-Mig, SMig).
 - Reconstruir la imatge amb les tres capes alhora, havent de passar menys imatges per la xarxa neuronal (S-Continu, SCont).
- Seleccionar tres bandes el més separat possible sense repetir i reconstruir la imatge amb les tres capes alhora (S-Separat, SSep).

5.3 Mètriques de Rendiment

Per tal d'avaluar els mètodes d'interpolació s'han seleccionat diverses mètriques d'avaluació que ens proporcionen

mesures diferents de l'error de la imatge avaluada respecte a el *Ground Truth* (GT). A continuació s'hi llisten les tècniques emprades:

- **MSE** o Error Quadràtic Mig (Mean Squared Error) mesura la mitjana dels errors al quadrat, és a dir, la diferència entre l'estimador i el que s'estima, al quadrat. Per al nostre cas, seria la suma de la diferència per a cada píxel entre el GT i el fotograma interpolat [23], tal que

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I_{GT}(i, j) - \tilde{I}_{t+\frac{\Delta t}{2}}(i, j)]^2, \quad (6)$$

on \tilde{I} és la imatge reconstruïda, I_{GT} la imatge ground truth, m i n amplada i alçada de la imatge. Una variant d'aquest algorisme n'és **RMS** (Root Mean Squared Error), que no és res més que l'arrel quadrada d'aquesta mesura. Aquest algorisme, però, no té en compte la percepció humana de la imatge, i el valor resulta mecànic.

- **PSNR** o Relació Senyal de Soroll de Pic (Peak Signal-to-Noise Ratio), serveix per mesurar l'efecte que provoca el soroll (que esdevé a la imatge quan s'ha realitzat una transformació, com aplicar una translació de píxels a partir d'un flux òptic, per al nostre cas) sobre la qualitat d'una imatge respecte a l'original. El PSNR es defineix més fàcilment utilitzant el MSE, tal que

$$PSNR = 10 \times \log_{10} \left(\frac{MAX_I^2}{MSE} \right), \quad (7)$$

on MSE és l'error quadràtic mig mencionat a l'equació 6, i MAX_I el valor màxim que pot prendre un píxel en la imatge I , essent 255 si es tracta de valors de 8 bits, aconseguint un valor màxim de 50 i mínim 0 db. És una escala logarítmica.

Aquesta tècnica és una aproximació de la percepció humana sobre la reconstrucció d'una imatge. En fer la comparació de les dues imatges el soroll generat a l'hora de fer la interpolació farà perdre qualitat a la imatge. Aquesta tècnica però tendeix a donar millors resultats a algorismes que suavitzen massa les imatges.

- **SSIM** o Índex de Similitud Estructural, no busca tant la diferència numèrica sinó més aviat la perceptiva que pots obtenir si compares les dues imatges. Es defineix com

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \quad (8)$$

on (x, y) és el centre d'una finestra de mida $N \times N$ on es calcula la mitjana μ i la variància σ . c_1 i c_2 son paràmetres per estabilitzar el resultat de la divisió amb un denominador dèbil.

A diferència de PSNR, també té en compte la similitud dels contorns entre la imatge interpolada i el ground truth. Parteix de la idea que s'ha realitzat una bona interpolació. Perquè un algorisme doni els millors resultats ha de preservar la qualitat de les textures alhora que manté definits els contorns dels objectes.

6 RESULTATS

A continuació s'analitzà unes quantes proves realitzades per comprovar el rendiment dels algorismes variant els paràmetres més rellevants comentats durant aquest treball.

6.1 Estudi de rendiment del Pipeline intern de Interpolació

Començarem comparant l'ús del Pipeline Simple amb el Pipeline Bidireccional introduïts a la secció 5.2 per als mètodes que fan una anàlisi del flux òptic (LK, H, F, T). Aquests primers tests s'han realitzat sobre les imatges de Gris del dataset Middlebury. Per a cada imatge a interpolar, s'ha calculat el rendiment obtingut aplicant el flux òptic simple o bidireccional, i se n'ha calculat la diferència. La figura 6 mostra la distribució d'aquesta diferència del pipeline bidireccional menys el del simple, mostrant millor rendiment del bidireccional per a totes les mètriques. Així doncs, s'utilitzarà aquesta estructura a l'hora d'interpolació amb mètodes extractors de flux òptic.

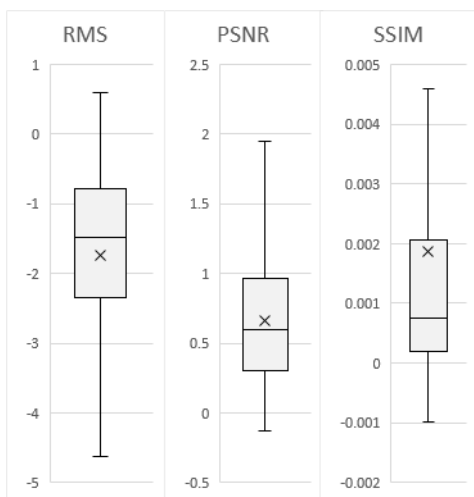


Fig. 6: Distribució de la diferència del rendiment del pipeline Bidireccional menys el rendiment del pipeline Simple.

6.2 Estudi d'entrenament del PCA

Quan volem utilitzar aquestes funcions de flux òptic amb imatges RGB s'ha de tenir en compte la capa que se selecciona abans no passar-la a la funció. El software permet transformar RGB en altres espais o representacions del color. Un d'ells n'és l'Anàlisi de Components Principals. Aquesta anàlisi es pot realitzar per tot un set de dades (global), o bé per una seqüència específica (local), o bé només pel parell d'imatges que s'hi vol interpolar (pair). Per comparar entre diferents mètodes utilitzarem una **comparació a parells matisat**. Aquesta comparació consisteix en avaluar el ràtio de vegades que un mètode és millor que un altre. S'hi genera una taula que compara els parells i que posteriorment, normalitzant els valors de la columna, s'obté quin mètode és preferible fent una mitjana dels ràtios normalitzats.

A la Taula 4 podem veure aquesta comparativa creuada per al rendiment dels algorismes segons l'extracció de components principals que fem. Com podem observar, la

diferència no és gaire notable, però tot i això utilitzar un PCA global és lleugerament preferible.

	Global	Local	Pair	Preferència
Global	1	1.0308	1.0952	34.7%
Local	0.9701	1	1	32.9%
Pair	0.9130	1	1	32.3%

TAULA 4: Comparació de parells matisat per a la comparació d'entrenament del PCA.

Deguts als recursos disponibles en realitzar aquest TFG, no ha estat possible entrenar un PCA global per als casos dels dataset MS (massa quantitat de dades) i el local fallava en la majoria de subjets. Per preservar la consistència dels resultats MS s'utilitzarà la modalitat PCA parell.

6.3 Estudi d'extracció del Flux Òptic en Imatges Multi Banda

Amb el PCA podem extreure el flux òptic de la banda més significativa i aplicar-la a totes les bandes (1D) o bé extreure-la de cada banda per separat i fer-ne una mitjana (Mean) i aplicar-la a totes les bandes, o bé tractar cada banda amb el seu flux òptic per separat (Naive). És per això que s'ha realitzat següent experiment per determinar la millor manera d'obtenir el flux òptic per a resultats de PCA en RGB. A la Taula 5 observem com, utilitzant una mitja de les bandes extretes amb el PCA, s'aconsegueix una preferència del 67% amb la mitjana.

	1D	Mitja	Naive	Preferència
1D	1	0.1462	3.2580	21.2%
Mitja	6.8387	1	3.7143	67.0%
Naive	0.3069	0.2692	1	11.8%

TAULA 5: Comparació de parells matisat per a comprovar com extreure el flux òptic utilitzant PCA en RGB (Middlebury Color)

A la Taula 6 observem com la preferència del tractament del flux òptic per als altres datasets segueix la distribució de la sorgida a la Taula 5.

	Preferència			
	Paint.	Propi	Surv.	M.RGB
1D	18.9%	20.9%	27.3%	21.2%
Mitja	67.5%	69.5%	57.5%	67.0%
Naive	13.7%	9.5%	15.2%	11.8%

TAULA 6: Comparació de preferència de com extreure el flux òptic utilitzant PCA entre tots els datasets.

Veiem com realitzant una mitjana dels canals per al cas de PCA obtenim millors resultats en tots els conjunts de dades.

6.4 Estudi interpolació aplicant flux òptic a partir d'una imatge d'intensitat

Hem vist com millorar el rendiment canviant l'espai de color RGB amb un PCA. Ara compararem aquesta tècnica

amb altres conversions de l'espai de colors. Aquests són una capa de Grisos (Gray), HLS, HSV, i YCbCr. A cada espai de color s'hi selecciona aquella banda que conté més informació, (excepte el PCA que utilitzarem la modalitat l'explicada a l'apartat 6.3). Que un espai vagi millor o pitjor per a l'extracció dependrà dels colors de l'escena. Com podem observar a la Taula 7, tots els canvis de color perjudiquen l'estat per defecte (Gray), excepte PCA. Aquest aconseguix una preferència del 25.4%.

	Gray	HLS	HSV	PCA	YCC	Pref.
Gray	1	1.40	1.05	0.80	1.12	20.9%
HLS	0.71	1	0.76	0.63	0.79	15.3%
HSV	0.95	1.31	1	0.72	1.14	19.9%
PCA	1.20	1.56	1.38	1	1.34	25.4%
YCC	0.88	1.15	0.87	0.75	1	18.3%

TAULA 7: Comparació de parells matisat per a comparar el rendiment d'extraure el flux òptic en diferents espais de color.

6.5 Estudi del rendiment de diferents tècniques d'interpolació en RGB

Arribats a aquest punt, podem comparar quin és el millor dels mètodes implementats per a la interpolació en el pipeline utilitzant les preferències més òptimes fins al moment. Com podem observar a la Figura 7, per als datasets de Color la millor de les implementades és la Xarxa Neuronal, i el pitjor de tots és el mètode Lineal.

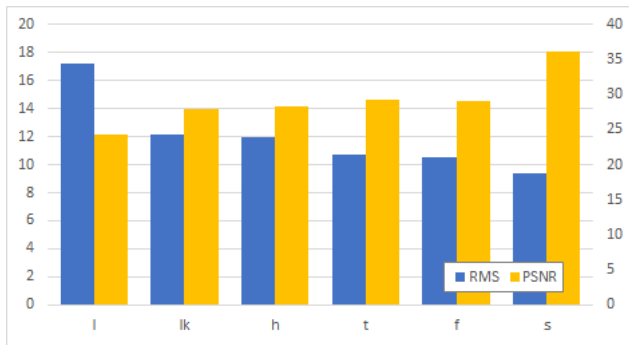


Fig. 7: Valors obtinguts de RMS i PSNR dels diferents mètodes d'interpolació amb el dataset Middlebury-Color.

6.6 Estudi del rendiment de la Xarxa Neuronal SepConv sobre imatges MS

La tècnica d'interpolació SepConv implementada treballa amb imatges RGB i per tant, si s'hi vol treballar amb imatges multi espectrals, aquestes s'hauran de tractar abans no entrar a la Xarxa Neuronal. Una forma de fer-ho és anar passant a la xarxa neuronal parts de la imatge i després reconstruir la interpolada amb les imatges que retorna la xarxa. S'han realitzat els quatre experiments mencionats a l'apartat 5.2.2 (SMig, SCont, SSep, SRep).

Comparant aquestes quatre tècniques ens trobem amb els resultats de la Taula 8.

Com veiem, la millor manera d'enviar els resultats a la Xarxa Neuronal és amb el mètode SCont. El mètode SSep,

	SMig	SCont	SSep	SRep	Preferència
SMig	1	0.57	0.70	1.03	19.7%
SCont	1.74	1	1.17	1.33	31.5%
SSep	1.42	0.85	1	1.42	28.0%
SRep	0.97	0.75	0.70	1	20.8%

TAULA 8: Comparació de parells matisat per comparar la millor entrada d'imatges MS al mètode SepConvNN.

en ser una variant del mètode SCont, també obté bons resultats. Així doncs, quan els datasets multi espectrals utilitzin aquesta xarxa neuronal (S), s'emprarà la modalitat SCont. Amb aquest ja tenim tots els mètodes adaptats a imatges multi espectrals del nostre dataset.

6.7 Estudi del rendiment de les tècniques d'interpolació en els diferents datasets

Per finalitzar, repassem els errors obtinguts per cada dataset del treball. Primerament, analitzem el rendiment de les diferents tècniques en MS. A la figura 8 observem com el mètode basat en Deep Learning (SepConvNN) és el que presenta millors resultats per la majoria de datasets aplicant la modalitat explicada a l'apartat 6.6.

A la Figura 9 s'hi mostren tots els datasets generats i recopilats en aquest treball. Veiem com es presenta una gran disparitat entre els errors de les seqüències. Això és degut a les oclusions de l'escena: Quant més es moguin els objectes de l'escena i destapin píxels desconeguts per a la interpolació, més error obtindrà la imatge degut a que és probable que s'equivoqui. En el nostre cas, els que més oclusió presenten són Grove i EE-CVC.

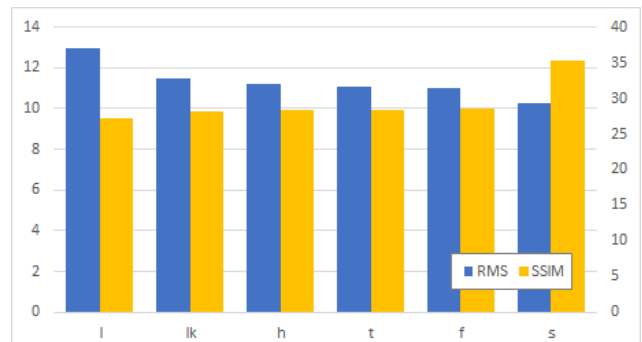


Fig. 8: Valors obtinguts de RMS i PSNR dels diferents mètodes d'interpolació en els datasets MS (Propi, Surveillance i Paintings).

7 CONCLUSIONS

El treball realitzat en aquest TFG ha permet assolir tots els objectius plantejats inicialment. Les principals contribucions del TFG són:

- Un dataset multi spectral per a l'anàlisi del Flux Òptic.
- Un entorn d'anàlisi automatitzat i modular que permet afegir i modificar noves funcions d'interpolació fàcilment.

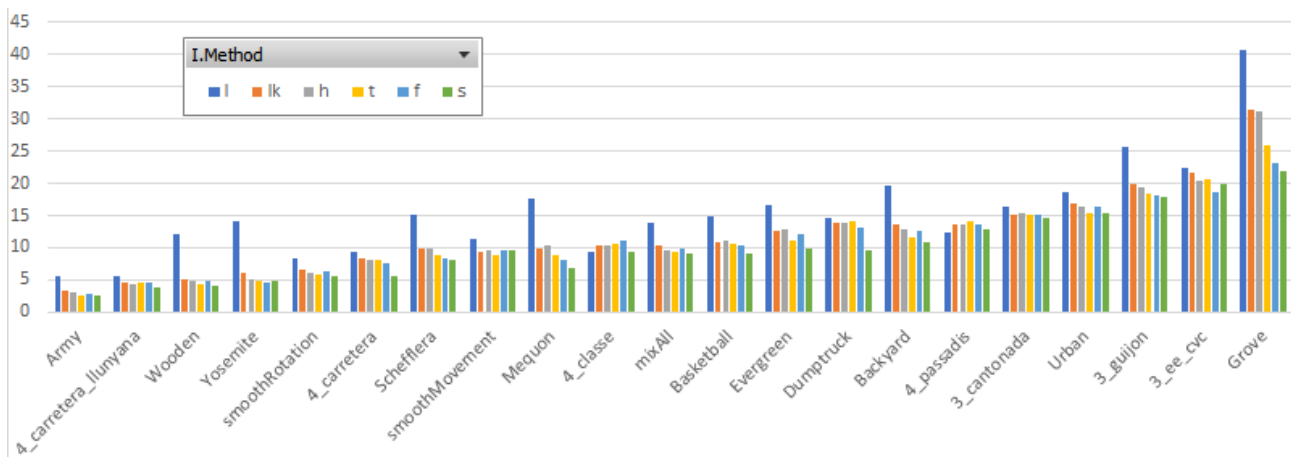


Fig. 9: Error RMS promig de cada Dataset segons el mètode interpolatiu emprat.

- Mètodes d'interpolació i mètriques de rendiment implementades en l'entorn d'avaluació.
- Diferents propostes d'adaptació pel cas d'imatges MS en una xarxa neuronal entrenada en RGB.
- Una anàlisi quantitativa de diferents opcions a tenir en compte a l'hora d'interpolat imatges RGB i MS.

Respecte els mètodes d'interpolació considerats:

- Realitzar un preprocessament PCA i fer una mitjana del flux òptic extret a cada banda aconsegueix millorar els resultats en la gran majoria dels casos.
- El mètode SepConvNN és el que obté un major rendiment per la majoria dels datasets.
- El mètode SepConvNN també es pot utilitzar per a imatges MS, essent el que aconsegueix millor rendiment per la majoria dels datasets MS.

Després de realitzar els experiments i un posterior anàlisi s'ha pogut interpolat satisfactòriament seqüències d'imatges multi espectrals utilitzant mètodes extractors de flux òptic i aplicant aquest flux a altres bandes de la imatge. S'ha realitzat una anàlisi per augmentar l'efectivitat d'aquest extractor canviant l'espai de colors de la imatge que s'utilitza per extraure el Flux Òptic, així com també utilitzar el Flux Òptic que pot extreure's de les altres bandes per realitzar una mitjana del flux. D'aquesta manera s'han obtingut els millors resultats per als tests d'imatges en color RGB. Hem comprovat com, en imatges MS, el procés per minimitzar l'error es repeteix. També s'ha pogut interpolat els resultats en una Xarxa Neuronal especialitzada per RGB, tant amb imatges RGB com en imatges multi espectrals. Hem comprovat com aquesta xarxa és la que dona millors resultats d'interpolació tot i les millores realitzades als extractors de Flux Òptic. Amb aquesta xarxa neuronal s'ha vist també les diverses formes d'interpolat imatges multi-espectrals sense la necessitat de ser RGB, de les propostes presentades, la que obté millors resultats és enviar en cadena les bandes de la imatge en subimatges de tres bandes i després reconstruir aquesta amb el resultat retornat.

Ha faltat, però, optimització general del codi. El procés general del Pipeline es relantitza per colls d'ampolla que, per manca de temps, no s'ha pogut optimitzar tot el desitjable.

8 TREBALL FUTUR

Actualment s'interpola el frame del mig, fet que complica l'obtenció d'un instant de temps determinat entre els dos frames. Podríem variar en els pipelines la velocitat dels vectors de flux òptic resultants per aconseguir altres instants de temps. A més, un cop realitzada aquesta millora, podríem realitzar la interpolació intermèdia a partir de diverses interpolacions d'un *step* reduït, aconseguint així possibles millores de predicció del flux òptic.

Quant a la xarxa neuronal, podria millorar-ne el rendiment tractar les dades realitzant una transformació de la imatge abans no passar-la per la xarxa, com per exemple, un *style transfer* que transformés les imatges d'*N*-Bandes a RGB, i viceversa. També podríem tornar a entrenar la xarxa per poder processar imatges de *N*-Bandes.

Normalment el flux òptic està lligat a un objecte de l'escena en moviment, per això, doncs, té sentit utilitzar el flux òptic per a totes les bandes. En imatges multi-espectrals, però, podria donar-se el cas en llocs molt específics que presentés moviments dispersos entre bandes de la imatge. Un exemple n'és la imatge satel·litària on es sol separar bandes de núvols del "fons", o bé imatge astronòmica on inclogués també RGB (Figura 10). Així doncs, estaria bé poder separar els diferents moviments ponderats de cada banda (utilitzant, per exemple, *k-means*) i aplicar-ho a les bandes que correspongui.

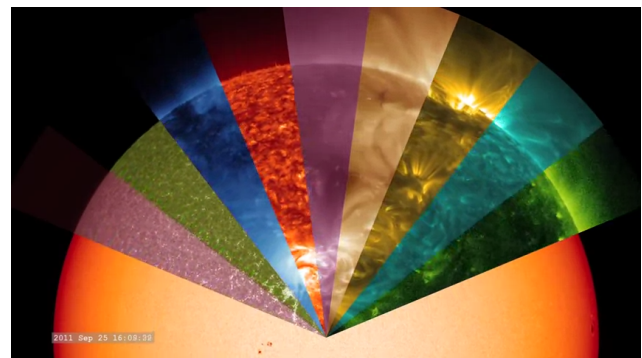


Fig. 10: Captura de Solar Dynamics Observatory de la NASA on mostra una imatge del Sol en diferents bandes de l'espectre.

REFERÈNCIES

- [1] Joan Nunes Universitat Autònoma de Barcelona. Interpolació espacial. <http://www.icgc.cat/Ciutada/Informa-t/Diccionaris/Interpolacio-espacial>, 2013. *Institut Cartogràfic i Geològic de Catalunya*, últim accés: 8 de Març de 2018.
- [2] Simon Baker, Daniel Scharstein, JP Lewis, Stefan Roth, Michael J Black, and Richard Szeliski. A database and evaluation methodology for optical flow. *International Journal of Computer Vision*, 92(1):1–31, 2011.
- [3] Article a la wikipèdia anglesa sobre el mètode d'interpolació lucas-kanade. https://en.wikipedia.org/wiki/Lucas%E2%80%93Kanade_method. Últim accés: 8 de Març de 2018.
- [4] Article a la wikipèdia anglesa sobre el mètode d'interpolació horn-schunck. https://en.wikipedia.org/wiki/Horn%E2%80%93Schunck_method. Últim accés: 8 de Març de 2018.
- [5] Pejman Rasti, Hasan Demirel, and Gholamreza Anbarjafari. Image resolution enhancement by using interpolation followed by iterative back projection. In *Signal Processing and Communications Applications Conference (SIU), 2013 21st*, pages 1–4. IEEE, 2013.
- [6] Philippe Weinzaepfel, Jerome Revaud, Zaid Harchaoui, and Cordelia Schmid. Deepflow: Large displacement optical flow with deep matching. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 1385–1392. IEEE, 2013.
- [7] Eddy Ilg, Nikolaus Mayer, Tonmoy Saikia, Margret Keuper, Alexey Dosovitskiy, and Thomas Brox. FlowNet 2.0: Evolution of optical flow estimation with deep networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, 2017.
- [8] Gunnar Farneback. Two-frame motion estimation based on polynomial expansion. In *Scandinavian conference on Image analysis*, pages 363–370. Springer, 2003.
- [9] Javier Sánchez Pérez, Enric Meinhardt-Llopis, and Gabriele Facciolo. Tv-l1 optical flow estimation. *Image Processing On Line*, 2013:137–150, 2013.
- [10] Simon Niklaus, Long Mai, and Feng Liu. Video frame interpolation via adaptive convolution. In *CVPR*, volume 2, page 6, 2017.
- [11] Simon Niklaus, Long Mai, and Feng Liu. Video frame interpolation via adaptive separable convolution. *arXiv preprint arXiv:1708.01692*, 2017.
- [12] *Meteosat second generation : the satellite development*. ESA, Noordwijk, The Netherlands, 11 1999. pag. 23, ISBN: 92-9092-634-1.
- [13] Worldview-3 satellite sensor specifications. satimagingcorp.com/satellite-sensors/worldview-3/. Satellite Imaging Corporation, últim accés: 8 de Març de 2018.
- [14] Spot-6 satellite sensor. satimagingcorp.com/satellite-sensors/spot-6/. Satellite Imaging Corporation, últim accés: 8 de Març de 2018.
- [15] Spot-7 satellite sensor. satimagingcorp.com/satellite-sensors/spot-7/. Satellite Imaging Corporation, últim accés: 8 de Març de 2018.
- [16] Rapideye satellite sensor. satimagingcorp.com/satellite-sensors/other-satellite-sensors/rapideye/. Satellite Imaging Corporation, últim accés: 8 de Març de 2018.
- [17] Pàgina web oficial de trello. trello.com.
- [18] Daniel Azemar. Repositori del projecte tfg. github.com/hamddan4/tfg. Últim accés: 8 de Març de 2018.
- [19] Python official website. python.org. Últim accés: 8 de Març de 2018.
- [20] Yannick Benezeth, Désiré Sidibé, and Jean-Baptiste Thomas. Background subtraction with multispectral video sequences. In *IEEE International Conference on Robotics and Automation workshop on Non-classical Cameras, Camera Networks and Omnidirectional Vision (OMNIVIS)*, pages 6–p, 2014.
- [21] Athanasios Zacharopoulos, Kostas Hatzigiannakis, Polykarpos Karamaoynas, Vassilis M Papadakis, Michalis Andrianakis, Kristalia Melessanaki, and Xenophon Zabulis. A method for the registration of spectral images of paintings and its evaluation. *Journal of Cultural Heritage*, 2017.
- [22] Pàgina oficial de parrot sequoia. <http://global.parrot.com/mx/companies/sequoia/>. Últim accés: 26 de Maig de 2018.
- [23] Error quadràtic mig (mse, mean squared error), viquipèdia. https://ca.wikipedia.org/wiki/Error_quadr%C3%A0tic_mig. Últim accés: 16 d'Abril de 2018.

APÈNDIX

A continuació es mostren les figures i recursos auxiliars al document.

A.1 Arbre d'objectius

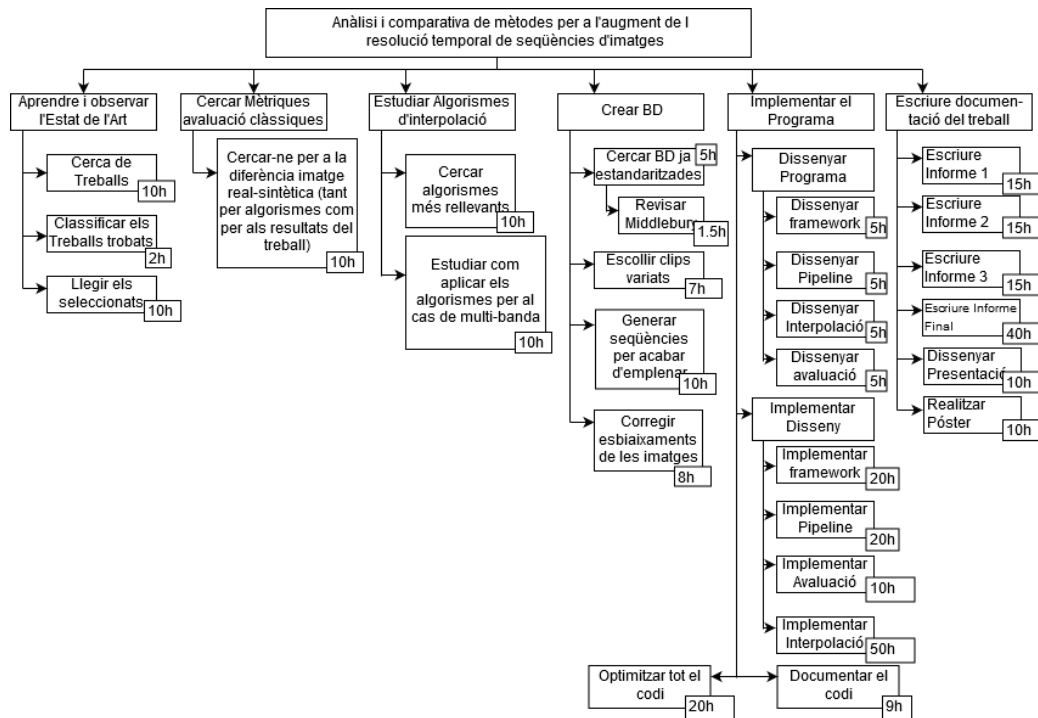


Fig. 11: Arbre d'objectius i tasques del TFG.

A.2 Interpolació Temporal - Exemples Visuals

Les següents seqüències d'imatges són una mostra exemple de les interpolacions implementades.

Respectivament cada columna és: GroundTruth, Lineal, Lucas-Kanade, Horn-Schunck, Farneback, TVL1, Sep-ConvNN.

Cada fila mostra, per ordre: Imatge interpolada, Miniatura en Detall, Error mitjà.

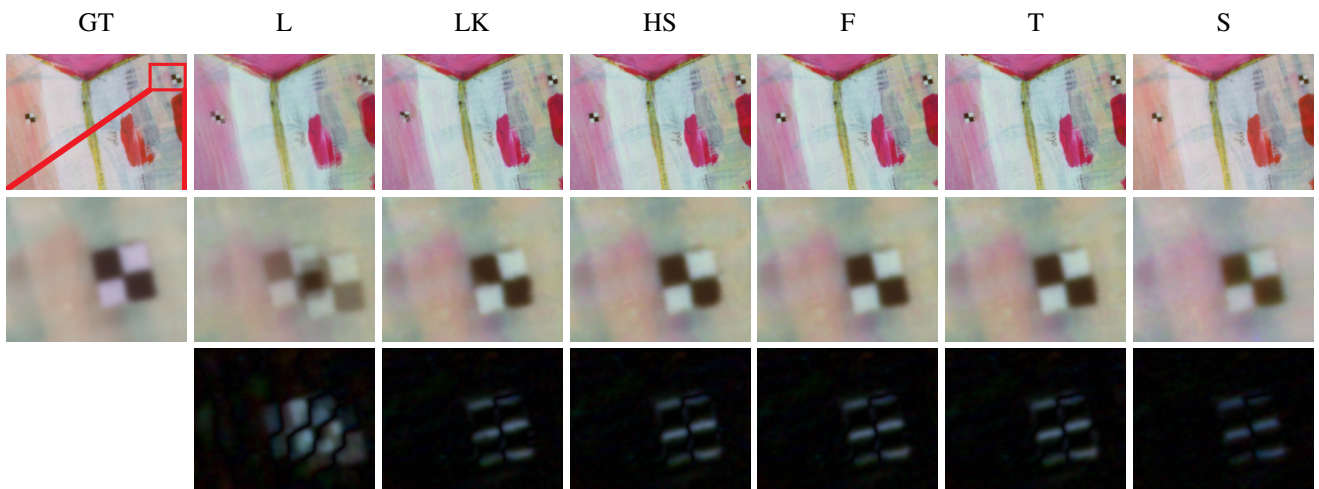


Fig. 12: Resultats per al primer frame del subset mixAll de Paintings.

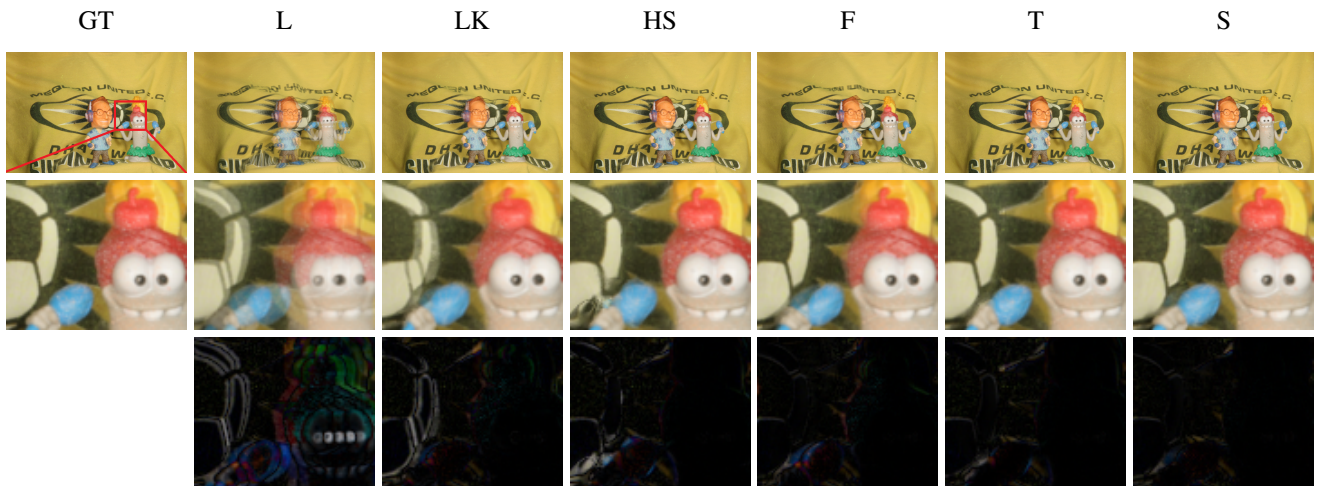


Fig. 13: Resultats per al segon frame del subset de Mequon de Middelbury-Color.

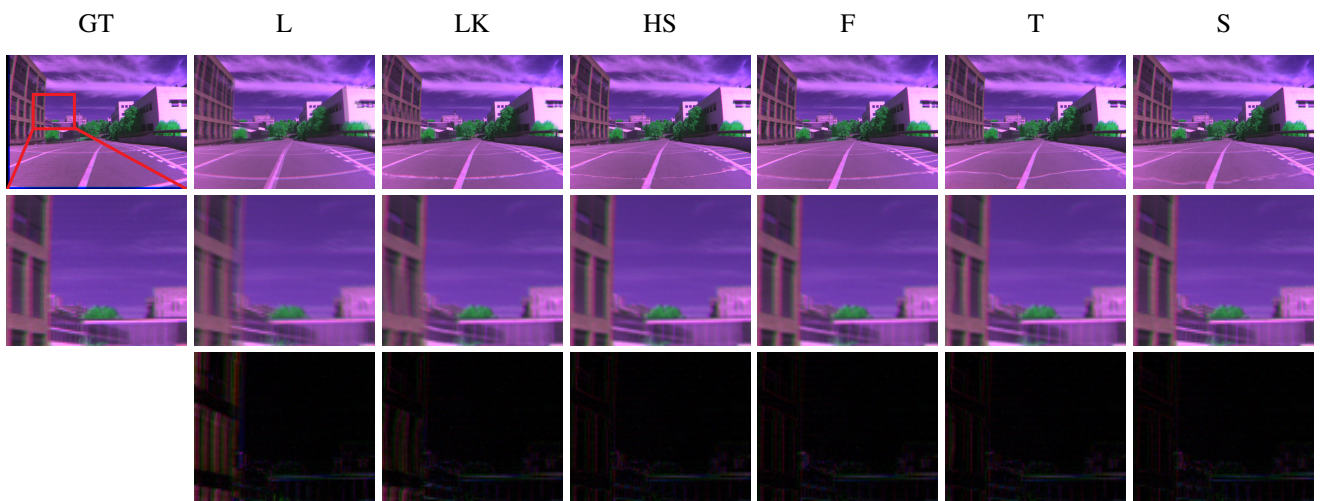


Fig. 14: Resultats per al primer frame del subset de EE-CVC del dataset Propi.

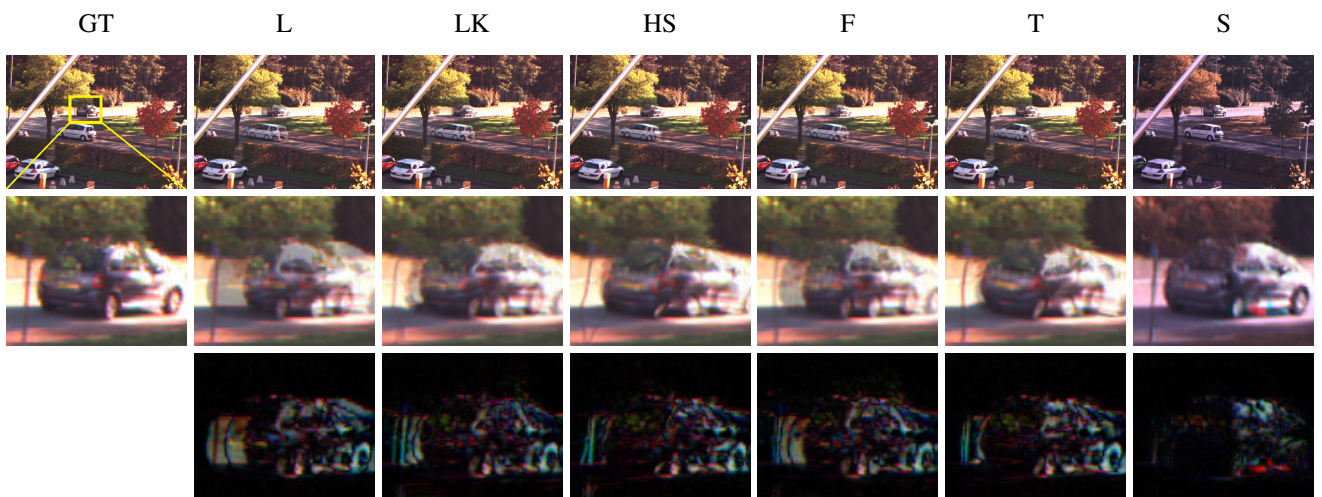


Fig. 15: Resultats per al segon frame del subset de Carretera de Surveillance.