

Relazione del progetto: HDR

Francesco Rossini Sibilla Merlo

December 2020

1 Introduzione

Come spiegato in [5], le immagini scattate da una fotocamera digitale spesso risentono di aree sovra-esposte o sotto-esposte, con una conseguente perdita di dettaglio. La soluzione fornita dalla tecnica *High dynamic range imaging* è quella di fondere varie immagini della medesima scena, ma con diversi livelli di esposizione, al fine di ottenere un'immagine con un intervallo dinamico maggiore, ovvero con un intervallo più ampio (HDR) tra le aree visibili più chiare e più scure [3].

Tuttavia, i display sono LDR, quindi il risultato finale dev'essere un'immagine LDR che sia "HDR-like" [4]. Come ben descritto in [6], questo risultato si può ottenere in due modi:

- *HDR-Reconstruction + Tone Mapping*: si fondono le immagini LDR in un'immagine HDR e si usa una funzione di *Tone Mapping* per mappare i toni in un range dinamico più basso adatto ai display.
- *Image Fusion*: si fondono le immagini direttamente in un'immagine LDR, che appaia "HDR-like".

Nel presente lavoro, si descrivono tre tecniche di *Image Fusion* sviluppate dagli autori, di complessità ed efficacia via via crescente:

1. **Br(ightness)**
2. **En(tropy)Sh(arpening)Br(ightness)**
3. **Sigmoid**

Le idee tratte dalla letteratura sono state combinate con elementi nuovi, al fine di ottenere un lavoro unico e originale. Si è utilizzata l'assunzione, lecita e presente in letteratura, che i soggetti della scena rimangano immobili tra un'immagine e l'altra, in modo che le differenze siano esclusivamente nell'esposizione. Non sono state effettuate assunzioni circa le modalità di scatto della fotocamera.

Le metodologie sono state implementate in funzioni MatLab.

In questo lavoro, si fornisce la formulazione matematica precisa dei metodi, nonché un commento e un'analisi sui risultati che potrà essere verificata dal lettore nell'ultima Sezione, che mostra le immagini ottenute.

2 Notazioni

L'input del problema *Image Fusion* è un vettore di immagini caratterizzate da uguali dimensioni (M, N) , rispettivamente altezza e larghezza. L'output è una singola immagine di dimensione (M, N) . K è la lunghezza del vettore, ovvero il numero di immagini considerate per la fusione. Si utilizzano, per il pixel alla locazione (x, y) , le notazioni descritte in tabella:

Immagine Ingresso	Immagine Output
$I^{(k)}(x, y)$	$I^*(x, y)$

Grayscale	Entropia	Filtrata Passa Alto
$I_g^{(k)}(x, y)$	$I_{En}^{(k)}(x, y)$	$I_{HP}^{(k)}(x, y)$

Peso normalizzato	Peso non normalizzato
$w^{(k)}(x, y)$	$W^{(k)}(x, y)$

Laddove non sia specificato diversamente (e.g.: Grayscale), l'immagine $I^{(k)}$ è una matrice tridimensionale in cui la terza dimensione rappresenta i tre canali RGB (pertanto $I^{(k)}(x, y)$ è la tripletta di valori RGB corrispondenti al pixel locato in (x, y)).

2.1 Formula generale di *Image Fusion*

I pixel $I^*(x, y)$ sono ottenuti con una somma dei K pixel in ingresso $I^{(k)}(x, y)$, ciascuno pesato con $w^{(k)}(x, y)$. La formula generale è pertanto:

$$I^*(x, y) = \sum_{k=1}^K I^{(k)}(x, y) \cdot w^{(k)}(x, y) \quad (1)$$

Ogni tecnica si caratterizza e si distingue dalle altre per la propria peculiare definizione dei pesi $w^{(k)}(x, y)$.

3 Il metodo Br

Il primo metodo sviluppato, **Br(ightness)**, si basa sull'osservazione che è sufficiente considerare la luminanza in scala di grigio delle immagini di input per ottenere dei buoni pesi $w^{(k)}(x, y)$ da inserire nella (1).

Nello specifico, dato un pixel in (x, y) , per ogni immagine $k \in K$ si calcola la sua luminanza in scala di grigio in (x, y) e la si normalizza, dividendola per la somma delle luminanze di tutte le K immagini in (x, y) :

$$w^{(k)}(x, y) = \frac{I_g^{(k)}(x, y)}{\sum_{h=1}^K I_g^{(h)}(x, y)}$$

Normalizzando, si associa ad ogni immagine un peso in $[0, 1]$ che rappresenta la quantità della sua luminanza in relazione alle altre $K - 1$ immagini. Ogni immagine dà quindi il suo contributo, direttamente proporzionale alla sua luminanza relativa nel pixel in questione.

3.1 Analisi del metodo

Il metodo è molto semplice dal punto di vista computazionale (i pesi si calcolano con una sola operazione, la divisione, ripetuta per $(M \cdot N \cdot K)$ volte).

Inoltre, garantisce un buon risultato per i seguenti motivi:

- L'immagine risultante non ha ovviamente zone sotto-esposte, dal momento che i pesi $w^{(k)}(x, y)$ sono maggiori per immagini con luminanza elevata, rispetto ad altre immagini con luminanza inferiore nel medesimo punto (x, y) .
- Meno trivialmente viene evitata anche la sovra-esposizione: tutte le immagini, anche le più scure, danno un contributo, benché minimo, alla definizione di I^* . In tal modo immagini più scure di quelle sovra-esposte, quindi con più dettaglio, non perdono completamente la propria informazione sugli *edge* e le variazioni di intensità nel risultato finale.

Il secondo aspetto evidenziato risulta interessante: si riesce a mantenere un'informazione sul dettaglio senza utilizzare misure apposite, ma solo con il parametro luminanza.

I difetti riscontrabili in questo metodo sono un'imperfetta realisticità della luce della scena e una perdita parziale del contrasto e degli *edge*, spiegabile apparentemente dal fatto che venga "preferita" la luminosità al contrasto.

4 Il metodo EnShBr

Il metodo **En(tropy)Sh(arpening)Br(ightness)** è il risultato di un processo di raffinamento che può essere utile ripercorrere a grandi linee.

Ci si è posti l'obiettivo di utilizzare una misura di contrasto e *sharpness* degli *edge*. In [2] è presentato un metodo di fusione basato esclusivamente sull'entropia. Secondo gli autori, l'entropia è la misura statistica che, senza avere a disposizione conoscenze fisiche sullo scatto, si rivela la più imparziale: non favorisce nessun pattern, come potrebbe avvenire invece con una misura locale basata sulle distanze, e non richiede uno *smoothing* successivo dell'immagine. In generale, il metodo si basa interamente sul massimizzare il contenuto informativo dell'immagine.

Per ogni immagine in input viene calcolata l'entropia locale di ciascun pixel rispetto ai suoi vicini, utilizzando l'istogramma come indicatore della distribuzione dei livelli di grigio. Anche nei casi di immagini iniziali a colori, la misura viene effettuata sulla luminanza in scala di grigio.

L'immagine risultante è ottenuta come somma pesata dei pixel di ogni immagine $k \in K$ con la matrice dei corrispettivi valori entropia locale $I_{En}^{(k)}$, normalizzata per i valori di entropia corrispondente ad ogni immagine.

$$I^*(x, y) = \frac{\sum_{k=1}^K I_{En}^{(k)}(x, y) \cdot I^{(k)}(x, y)}{\sum_{k=1}^K I_{En}^{(k)}(x, y)}$$

Limitandosi esclusivamente a questa misura, nelle immagini risultanti non vengono rispettati il colore e la luminosità della scena; inoltre, si può notare la comparsa di artefatti in aree omogenee. Tuttavia, in zone ricche di dettaglio e di *texture*, il risultato appare migliore rispetto al metodo **Br**.

Pertanto, ci si è posti due obiettivi:

1. Rispettare la luminosità e il colore della scena originale.
2. Diminuire il rumore e gli artefatti nelle aree omogenee dovute all'imprecisione della misura entropia.

4.1 Luminosità e colore

Si è utilizzato a tal fine un parametro già descritto: la luminanza a livelli di grigio. A differenza di **Br**, non viene normalizzata sulle K immagini (la normalizzazione si effettuerà in un secondo momento sul parametro completo). Tuttavia, un problema evidente è la diversità di dominio di $I_{En}^{(k)}(x, y)$, l'immagine entropia, e $I_g^{(k)}(x, y)$: la prima varia tra 0 e 7 circa, la seconda tra 0 e 65535 (rappresentazione a 16 bit). Pertanto, entrambe le grandezze sono state scalate nel seguente modo:

$$\begin{aligned}\hat{I}_{En}^{(k)}(x, y) &= \frac{I_{En}^{(k)}(x, y)}{\max(I_{En})} \\ \hat{I}_g^{(k)}(x, y) &= \frac{I_g^{(k)}(x, y)}{\max(I_g)}\end{aligned}$$

dove $\max(I_{En})$ e $\max(I_g)$ rappresentano i massimi lungo tutte e tre le dimensioni M, N, K dei vettori di immagini entropia e livelli di grigio.

Una prima stima di $W^{(k)}(x, y)$ (non normalizzata) è fornita da:

$$\hat{I}_{En}^{(k)}(x, y) + \hat{I}_g^{(k)}(x, y)$$

Tuttavia il risultato non è ancora ottimale a causa del rumore e degli artefatti provocati dalla misura entropia.

4.2 Diminuire il rumore della misura entropia

Una tecnica che è parsa efficace per diminuire il rumore nelle zone omogenee e scarsamente popolate da *edge* è stata l'introduzione di una seconda misura di contrasto e sharpness che affiancasse l'entropia. Per questa, ci si è riferiti al metodo del filtraggio passa-alto in [5]. Si effettua una convoluzione del seguente filtro:

$$h = \begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}$$

su ognuna delle K immagini in scala di grigio, ottenendo il vettore di immagini I_{HP} ; poi, si costruiscono le rispettive K matrici *booleane* nel seguente modo:

$$B^{(k)}(x, y) = \begin{cases} 1 & \text{if } I_{HP}^{(k)}(x, y) = \max\{I_{HP}^{(h)}(x, y) : h \in K\} \\ 0 & \text{else} \end{cases}$$

Tali matrici rappresentano l'informazione, per un dato pixel in (x_1, y_2) , sull'immagine a più alto contrasto per lo stesso: tale k_1 -esima immagine ha $B^{(k_1)}(x_1, y_1) = 1$. Inoltre $B^{(k)}(x_1, y_1) = 0$ per $k \neq k_1$. Si ottiene finalmente una misura di contrasto:

$$B^{(k)}(x, y) \cdot \hat{I}_{En}^{(k)}(x, y)$$

in cui il prodotto con la matrice binaria B ha il significato di un AND logico tra le due misure di contrasto: l'immagine k contribuisce al contrasto in (x, y) solo se la sua matrice binaria ha valore $B^{(k)}(x, y) = 1$.

4.3 Pesi e normalizzazione

I pesi sono pertanto ottenuti come:

$$W^{(k)}(x, y) = (B^{(k)}(x, y) \cdot \hat{I}_{En}^{(k)}(x, y)) + \hat{I}_g^{(k)}(x, y)$$

e normalizzati:

$$w^{(k)}(x, y) = \frac{W^{(k)}(x, y)}{\sum_{h=1}^K W^{(h)}(x, y)}$$

5 Il metodo Sigmoid

La descrizione di questo metodo sarà breve: si è pensato di migliorare ulteriormente il metodo **EnShBr** passando in una sigmoide i valori dei pesi prodotti da quest'ultimo. I risultati sono apparsi formidabili.

C'è ragione di ritenere che l'introduzione di una sigmoide abbia l'effetto di un'applicazione, in ultima fase, di *smoothing* dei valori dei pesi. Tuttavia, gli autori sono costretti ad ammettere che si tratta, più che altro, di un'intuizione avallata dal risultato sperimentale e non dispongono ancora di una soddisfacente spiegazione scientifica.

La formula implementata è la seguente (si noti che alla sigmoide è stata sommata nuovamente $\hat{I}_g^{(k)}(x, y)$ per contro-bilanciare il fatto che le immagini risultanti apparivano scurite):

$$W^{(k)}(x, y) = \frac{1}{1 + e^{-(B^{(k)}(x, y) \cdot \hat{I}_{E_n}^{(k)}(x, y) + \hat{I}_g^{(k)}(x, y))}} + \hat{I}_g^{(k)}(x, y)$$

6 Implementazione MatLab

Quanto descritto è stato implementato in MatLab, sfruttando al massimo la compattezza offerta dal linguaggio nel manipolare matrici.

I vettori di immagini sono stati trattati come 4-dimensional (larghezza, altezza, canali RGB, numero di immagini) e le operazioni sono state principalmente eseguite *inline* come moltiplicazioni *point-wise* e somme, limitando le iterazioni esplicite esclusivamente alla lettura e conversione delle immagini (in alcuni casi di conversione, e.g. a livelli di grigio, è stato indispensabile indirizzare esplicitamente l'indice $k \in K$ lungo le immagini del vettore).

7 Risultati sperimentali

Il testing è stato effettuato su 4 dataset diversi, rispettivamente costituiti da 12, 6, 5 e 4 immagini riproducenti identica scena ma ad esposizioni diverse. Le immagini sono state generosamente messe a disposizione da Klaus Herrmann in [1]. Per le immagini originali, non inserite per brevità, si veda [1].

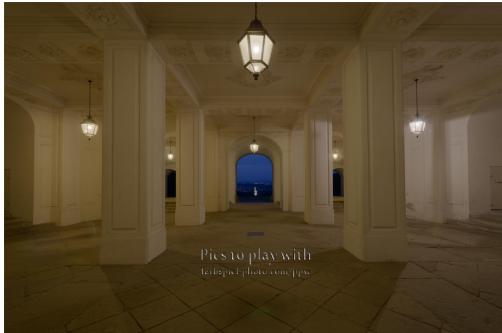
Le immagini sono incolonnate secondo i risultati ottenuti rispettivamente dal metodo **Br**, **EnShBr** e **Sigmoid**.



(a) Br



(b) EnShBr



(c) Sigmoid



(d) Br



(e) EnShBr



(f) Sigmoid



(g) Br



(h) EnShBr



(i) Sigmoid



(j) Br

(k) EnShBr

(l) Sigmoid

Riferimenti bibliografici

- [1] Klaus Herrmann. *Pics to play with*. URL: <https://farbspiel-photo.com/learn/hdr-pics-to-play-with>.
- [2] Johannes Herwig e Josef Pauli. «An information-theoretic approach to multi-exposure fusion via statistical filtering using local entropy». In: ().
- [3] *High dynamic range imaging*. URL: https://it.wikipedia.org/wiki/High_dynamic_range_imaging.
- [4] Ketan Kotwal e Subhasis Chaudhuri. «An Optimization-based Approach to Fusion of Multi-Exposure, Low Dynamic Range Images». In: *14th International Conference on Information Fusion* (2011).
- [5] Shutao Li e Kudong Kang. «Fast Multi-exposure Image Fusion with Median Filter and Recursive Filter». In: *IEEE Transactions on Consumer Electronics* 58.2 (2012). doi: 10.1109/TCE.2012.6227469.
- [6] Rui Shen et al. «Generalized Random Walks for Fusion of Multi-Exposure Images». In: *IEEE TRANSACTIONS ON IMAGE PROCESSING* 20.12 (2011). doi: 10.1109/TIP.2011.2150235.