

*Decomposizione ai Valori Singolari
per la compressione di un'immagine RGB*

“Numerical Methods for Computer Science”

A.A. 2018/2019

Studente:
Marco Cianiotta [701883]

Introduzione

Lo sviluppo della tecnologia digitale consente, da diversi decenni ormai, di raggiungere degli standard di qualità fotografica superiori a quelli forniti dalla fotografia su pellicola. Un'immagine digitale, può essere vista come costituita da un reticolo di $m \times n$ pixel. Nel caso di una immagine in *scala di grigi*, a ciascun pixel è attribuito un valore numero corrispondente all'intensità luminosa di quel punto nell'immagine. Una immagine a colori (salvata nel formato Portable Pix Map¹) invece, viene memorizzata con tre matrici di interi dove la prima registra le intensità di *rosso*, la seconda di *verde* e la terza di *blu*. In queste matrici lo 0 determina la minima intensità di colore mentre il 255 corrisponde alla massima. Tale modalità di rappresentazione dell'immagine a colori è conosciuta come *RGB (Red, Green and Blue)*. L'innovazione costante del settore fotografico ha dato la possibilità di avere immagini ad altissima risoluzione già attraverso lo *Smartphone*. Questo sviluppo però porta con sé un problema; proporzionalmente all'aumento della risoluzione, si ha un aumento in termini di occupazione di memoria. Si pensi, ad esempio, che una immagine (nel formato *JPG*²) di *6 megapixel* occupava circa *70Kb* mentre ora, con una di *12 megapixel* si occupano *2Mb* circa. Da questo problema nasce la necessità di *comprimere*³ le immagini. Si ricorda che il formato *.JPG* è un metodo molto utilizzato di *Compressione con perdita di dati* (in Inglese *lossy compression*). In questo progetto si dà un esempio di applicazione di un metodo per realizzare un tipo diverso di *compressione* che si basa sulla *Decomposizione ai Valori Singolari* (in Inglese *Singular Value Decomposition*) di una matrice (descritto nel 1 paragrafo). Verranno esposti gli strumenti e il metodo con la relativa implementazione del programma nel linguaggio di programmazione *Python* dove sarà effettuata la sperimentazione su una immagine nel formato *.JPG*, risoluzione 1000×575 , e peso *137Kb*. Per garantire una *compressione ottimale* sono state utilizzate indicatori qualitativi quali *SSIM (Indice di Similarità Strutturale)*, in Inglese *Structural Similarity*) e *MSE (Errore quadratico medio)*, in Inglese *Mean Square Error*) rispettivamente descritti nel 2 paragrafo. Tali indicatori ci consentiranno di fermare il processo di *compressione* dell'immagine prima che questa risulti essere percepita all'occhio umano come di bassa qualità (in termini di risoluzione).

1. DVS (Decomposizione ai valori singolari)

Ricordiamo che una matrice A di dimensioni $m \times n$ può essere rappresentata mediante la sua *decomposizione ai valori singolari* nota come

$$A = U \Sigma V^T$$

dove U è una matrice $m \times m$ ortogonale⁴, V è una matrice $n \times n$ ortogonale, Σ è una matrice $m \times n$ diagonale, cioè tale che $(\Sigma)_{ij} = 0$ se $i \neq j$, $(\sigma)_{ii} = \sigma_i \geq 0$, $i = 1, \dots, \min(m, n)$. Si assume inoltre che $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_{\min(m, n)}$. Le colonne di U denotate con u_i , $i = 1, \dots, m$ sono i vettori singolari destri di A , le colonne v_i di V , per $i = 1, \dots, n$ sono i vettori singolari sinistri di A mentre i numeri σ_i sono i valori singolari. [1]

¹ Un formato Netpbm è qualsiasi formato grafico utilizzato e definito dal progetto Netpbm. Il formato portatile pixmap, il formato portatile graymap e il formato portatile bitmap sono formati di file immagine progettati per essere facilmente scambiati tra le piattaforme.

² JPEG è l'acronimo di Joint Photographic Experts Group, un comitato di esperti ISO/CCITT che ha definito il primo standard internazionale di compressione dell'immagine digitale a tono continuo, sia a livelli di grigio sia a colori.

³ La compressione dell'immagine, in elettronica, informatica e nelle telecomunicazioni, è il nome generico sotto il quale si raggruppano gli algoritmi e le tecniche di elaborazione digitale delle immagini appartenenti alle tecniche di compressione dati, che si utilizzano per ridurre le dimensioni in byte delle immagini digitali.

⁴ Matrice ortogonale: matrice invertibile (matrice per la quale esiste un'altra matrice t.c. il prodotto matriciale tra le due restituisce la matrice identità) la cui trasposta (matrice ottenuta scambiandone le righe con le colonne) coincide con la sua inversa

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} = \underbrace{\begin{pmatrix} u_{11} & u_{12} & \dots & u_{1r} \\ u_{21} & u_{22} & \dots & u_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ u_{m1} & u_{m2} & \dots & u_{mr} \end{pmatrix}}_U \cdot \underbrace{\begin{pmatrix} s_1 & 0 & \dots & 0 \\ 0 & s_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & s_r \end{pmatrix}}_\Sigma \cdot \underbrace{\begin{pmatrix} v_{11} & v_{12} & \dots & v_{1n} \\ v_{21} & v_{22} & \dots & v_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{r1} & v_{r2} & \dots & v_{rn} \end{pmatrix}}_{V^*}$$

La decomposizione ai valori singolari può essere scritta in modo equivalente come:

$$\mathbf{A}_k = \sum_{i=1}^{\min(m,n)} \sigma_i \mathbf{u}_i \mathbf{v}_i^T$$

Ricordiamo la proprietà per cui

$$\mathbf{A}_k = \sum_{i=1}^k \sigma_i \mathbf{u}_i \mathbf{v}_i^T$$

è quella matrice che minimizza $\|A - X\|_2$ tra tutte le matrici di X di rango k . Cioè è la migliore approssimazione di rango k di A valutata in norma 2 (ovvero in norma di Frobenius). Se vogliamo quindi approssimare una immagine A di dimensione $m \times n$ pixel attraverso una matrice di rango $k = \min(m, n)$, possiamo considerare la matrice A_k di migliore approssimazione in norma di Frobenius. Questo, ci consente di raggiungere lo scopo del nostro progetto, ovvero cercare di ridurre lo spazio di memoria occupato da una immagine $m \times n$ calcolando la DVS della matrice A che la rappresenta e memorizzando solo i primi k valori singolari. Quindi, l'obiettivo è cercare il k ottimale. Nella sezione "introduzione" si è già specificato che l'immagine presa in esame per questa sperimentazione è di tipo *RGB*.

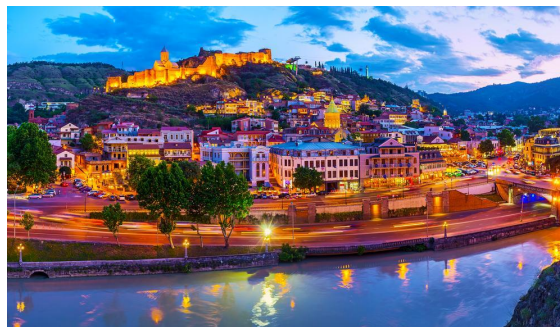


Figure 1: Immagine Originale

Ricordiamo che essendo *RGB* viene memorizzata con tre matrici di interi dove la prima registra le intensità di *rosso*, la seconda di *verde* e la terza di *blu* e quindi, si è dovuto procedere all'esecuzione della DVS per ogni matrice (R , G e B).

```
def svd_compress_rgb(img, k_r, k_g, k_b):
    """Given a matrix representing a RGB image, compress
    it by taking the largest k elements from its singular values"""
    # split into separate channels
    comp_r = svd_compress_gs(img[:, :, 0], k_r)
    comp_g = svd_compress_gs(img[:, :, 1], k_g)
    comp_b = svd_compress_gs(img[:, :, 2], k_b)
```

Figure 2: DVS applicata alla matrice R , G e B

Dopo aver eseguito la *Decomposizione ai Valori Singolari* per la matrice *Red*, *Green* e *Blue*, le si ricompongono in un'unica matrice che rappresenterà la nuova immagine compressa.

```

new_img = np.zeros(img.shape, img.dtype)
nrows = img.shape[0]
ncols = img.shape[1]
nchans = img.shape[2]
for i in range(nrows):
    for j in range(ncols):
        for c in range(nchans):
            val = 0
            if c == 0:
                val = comp_r[i][j]
            elif c == 1:
                val = comp_g[i][j]
            else:
                val = comp_b[i][j]
            # float64 values must be between -1.0 and 1.0
            if val < -1.0:
                val = -1.0
            elif val > 1.0:
                val = 1.0
            new_img[i][j][c] = val
return new_img

```

Figure 3: Ricomposizione delle tre matrici (R,G,B)

La matrice risultante (e di conseguenza, l'immagine) sarà determinata in base al valore di k ovvero in base ai k valori singolari che si è scelto di memorizzare. Ne consegue che più il valore di k è piccolo, minori saranno i valori singolari memorizzati e di conseguenza ne risulterà un'immagine di bassa qualità. Ad esempio, applicando eseguendo la DVS con $k=2$ l'immagine ricostruita dall'immagine risultante è la seguente:

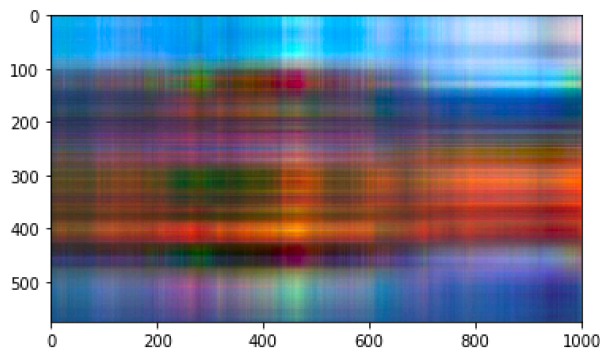


Figure 4: Immagine con $k=2$

Quindi, ne consegue che la scelta del k si rende fondamentale anche perché è un valore che non può essere determinato a priori. Per tale motivo, nel programma sviluppato il valore di k parte da un valore molto basso (nel caso specifico è stato impostato il valore di $k=2$) per poi aumentare in maniera costante. Emerge quindi il problema di scegliere una *soglia* che ci consenta di fermare l'esecuzione della DVS ma allo stesso tempo di ottenere una immagine compressa senza rinunciare alla qualità. Ne consegue che non possiamo scegliere come immagine ottimale quella la cui matrice minimizza la distanza in norma di Frobenius con la matrice dell'immagine originale in quanto sono metriche puramente quantitative. Si è scelto quindi di utilizzare metriche qualitative⁵ che ci consentono di misurare la qualità dell'immagine in relazione a come questa viene percepita⁶ dall'occhio umano. Tali metriche saranno discusse nel paragrafo successivo.

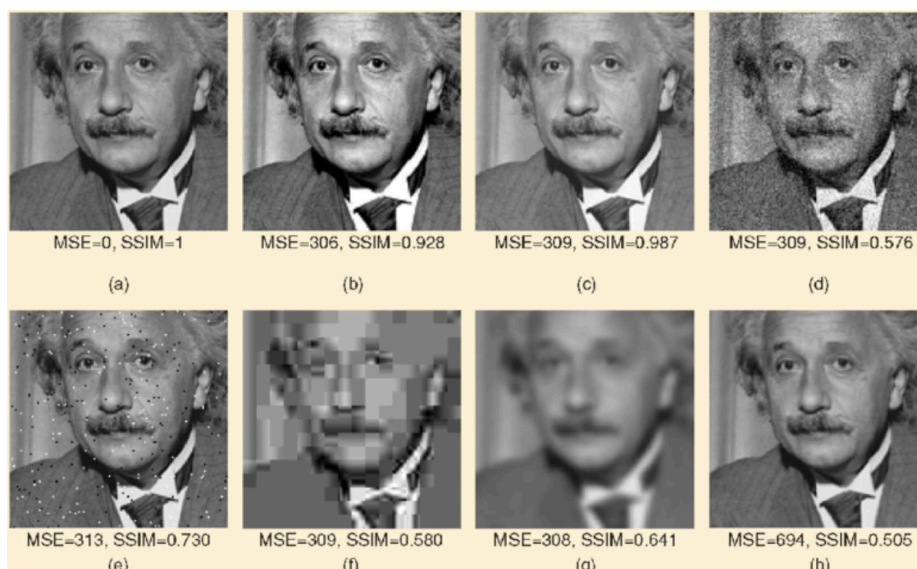
2. MSE e SSIM

SSIM (in italiano, *Indice strutturale di similarità*) è un metodo per predire la qualità percepita di una immagine o video digitale [2]. È usualmente utilizzato per misurare la similarità tra due immagini.

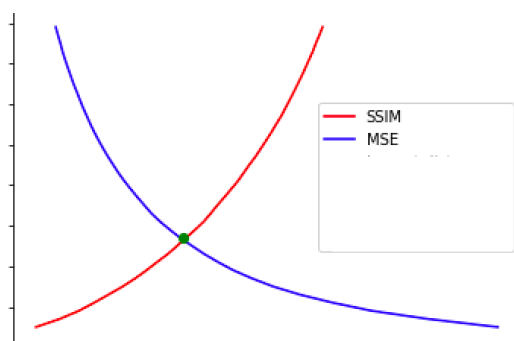
⁵ Una metrica di qualità è un metodo che permette di dare una valutazione qualitativa di un'immagine. È quindi fortemente legata al modello di percezione del sistema visivo umano.

⁶ La qualità percepita è un parametro soggettivo, dipendente oltre che dalle caratteristiche dell'immagine, anche da altri fattori quali: tecniche di compressione/decompressione che degradano l'immagine, le conversioni di formati...

La misurazione o la previsione della qualità dell'immagine si basa su un'immagine iniziale non compressa o priva di distorsioni. Questo è un modello basato sulla percezione che considera il degrado dell'immagine come un cambiamento percepito nelle informazioni strutturali ovvero, l'idea è che i pixel hanno forti interdipendenze specialmente quando sono spazialmente vicini. Queste dipendenze portano importanti informazioni sulla struttura degli oggetti nella scena visiva⁷. L'*Indice Strutturale di Similarità* può assumere un valore nell'intervallo tra 0 e 1 dove 1 indica la massima similarità. Poiché l'occhio umano è in grado di estrarre informazioni strutturali, la perdita di informazioni strutturali può essere usata per approssimare la distorsione di una immagine. L'*MSE* (in Italiano, *Errore quadratico medio*) serve per stimare l'errore medio quadratico tra due immagini; più tale indice è basso minore è la differenza tra le immagini ovvero minore è l'errore. Entrambi i metodi (*MSE* e *SSIM*) si comportano in maniera differente durante la compressione dell'immagine in quanto, all'aumentare di K (all'aumentare dei k valori singolari memorizzati), l'*MSE* tenderà a 0 invece l'*SSIM* tenderà a 1. Così come descritto nel precedente paragrafo, si è cercato, in questo esperimento, di trovare un *criterio di stop*⁸ che consenta di ottenere una immagine compressa percepibile all'occhio umano come di buona qualità. L'*MSE* a differenza del *SSIM* (che ha sia un valore minimo (0) che uno massimo (1) stabiliti), ha solo il valore minimo che è 0 mentre il massimo valore varia a seconda dell'immagine che si confronta con quella originale. Ad esempio:



Se proiettassimo all'interno di un grafico i diversi valori assunti dall'*MSE* e dal *SSIM* con k crescente avremmo:



⁷ La scena visiva è caratterizzata da fattori come la profondità, colore, variazioni di intensità luminosa che definiscono i bordi ed i contorni degli oggetti.

⁸ La DVS esegue la *Decomposizione a valori singolari* in base al valore k

Il punto colorato in *verde* nell'immagine sovrastante, è il punto di *intersezione* dei due *indicatori*, punto in cui l'immagine risulta risulterebbe avere un buon livello di *similarità* (*SSIM*) ed un buon livello di *errore* (*MSE*). Per poter eseguire e quindi implementare questa valutazione si necessita di normalizzare il valore dell'*MSE* in un intervallo da 0 a 1 (simile a quello del *SSIM*). Il valore normalizzato sarà dato da:

$$z_i = \frac{x_i - \min(x)}{\max(x) - \min(x)}$$

Ci occorre quindi sapere il valore massimo dell'*MSE* per applicare la precedente formula. Tale valore possiamo facilmente calcolarlo alla prima iterata del *DVS* in quanto, come descritto nel precedente paragrafo, nel nostro caso specifico il valore di *k* parte da 2 per poi incrementare. Con *K=2* abbiamo visto nella Figura 4 che l'immagine ha una pessima qualità e calcolando il valore dell'*MSE* confrontando l'immagine compressa con l'immagine originale, otteniamo quanto segue:



MSE: 1734.01 SSIM: 0.35

Dopo aver calcolato l'*MSE* massimo si potranno normalizzare tutti i vari valori dell'*MSE* delle diverse immagini compresse con i differenti valori di *k*:

```
for i in range(0, mcount):
    mNormalized.append(m[i]/m[0])
```

Figure 6: Normalizzazione di tutti i valori dell'*MSE* per le diverse immagini ottenute dai crescenti valori di *K*

Possiamo adesso visualizzare che i valori dell'*MSE* e del *SSIM* al crescere di *k* variano come ci aspettavamo, intersecandosi in un punto nel grafico:

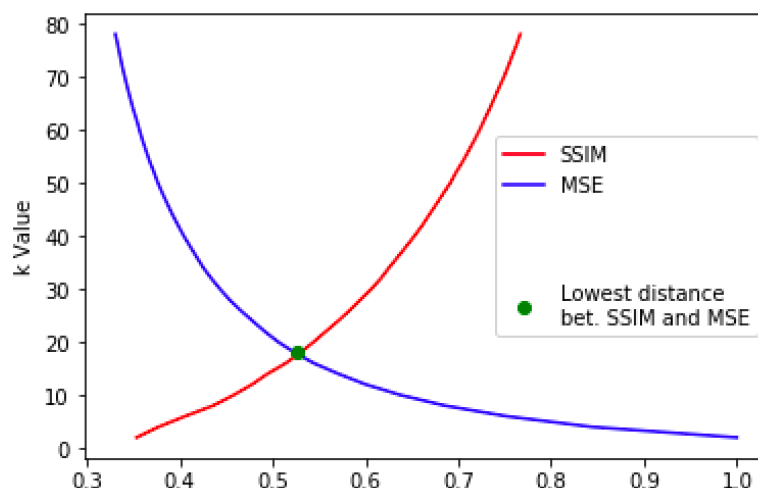


Figure 7: *SSIM* e *MSE*

Le due rette (Figura 7), costruite dai valori dei singoli indicatori, si intersecano quando entrambi gli indicatori raggiungono un valore molto simile ovvero quando si raggiunge la minima distanza in valore assoluto tra i due indicatori:

- $SSIM: 0.529$ $MSE: 0.523$

Dall'intersezione, ricaviamo quindi che il presunto valore ottimale di k è 18 la cui corrispondente immagine è:

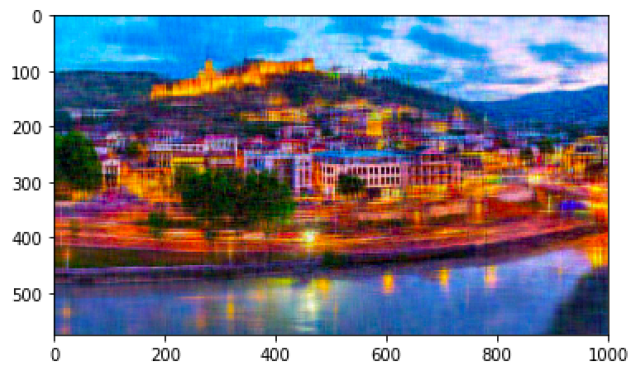


Figure 8: Immagine compressa con $k=18$

Con questo livello di compressione si ha una diminuzione di peso dell'immagine di 49.54 Kb passando dall'immagine originale che era di 137.29 Kb a 87.74 Kb .

3. Conclusione

Da un'attenta analisi dell'immagine in Figura 8, risulta che, nonostante $k=18$ sia un presunto punto ottimale (in quanto punto di intersezione tra i valori dei due indicatori), la qualità dell'immagine percepita dall'occhio umano risulta essere molto bassa. Avendo effettuato una ricerca approfondita in letteratura ([3] [4] [5] [6]), la maggior parte degli articoli scientifici considerano il $SSIM$ con un valore di almeno $0,7$ per garantire un'immagine qualitativamente buona alla percezione dell'occhio umano. A esempio in un articolo [7] è stata effettuata una sperimentazione che mirava a capire da quale livello di $SSIM$ in poi i partecipanti ritenevano l'immagine compressa identica a quella originale. L'esperimento è stato condotto su 30 partecipanti di diversa età al quale sono state sottoposte 6 immagini differenti dove, per ogni immagine vi erano 10 versioni della immagine compressa con differenti livelli di $SSIM$ (82, 84, 86, 88, 90, 92, 94, 96, 98, 99.9). È risultato che i partecipanti ritenevano l'immagine compressa con $SSIM=0.90$ e $MSE=0.23$ uguale all'originale. L'immagine considerata nel nostro caso, con $SSIM=0.90$ risulta essere:

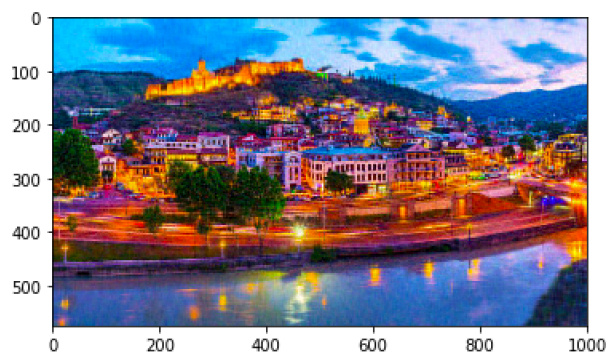


Figure 9: Immagine compressa con $k=200$, $SSIM=0.90$, $MSE=0.23$

Con un valore così elevato di $SSIM$ l'immagine, pur essendo compressa risulta pesare circa 134.75 Kb in confronto all'originale di 137.29 Kb , perdendo quindi, solo 2.54 Kb . Nel nostro caso, si è deciso di ricavare l'immagine cercando quel k dove la distanza in valore assoluto tra MSE e $SSIM$ fosse la più piccola imponendo però come vincolo $SSIM \geq 0.73$ (si è deciso per 0.73 in quanto è un giusto

compromesso tra la quantità d'errore che è del 30% circa e la similarità dell'immagine rispetto all'originale che è pari al 70% circa) in quanto in confronto con l'immagine con $SSIM=0.90$ si ha un errore maggiore di solo 0.10 che risulta essere tollerante visto che non incide significativamente sulla qualità percepita.

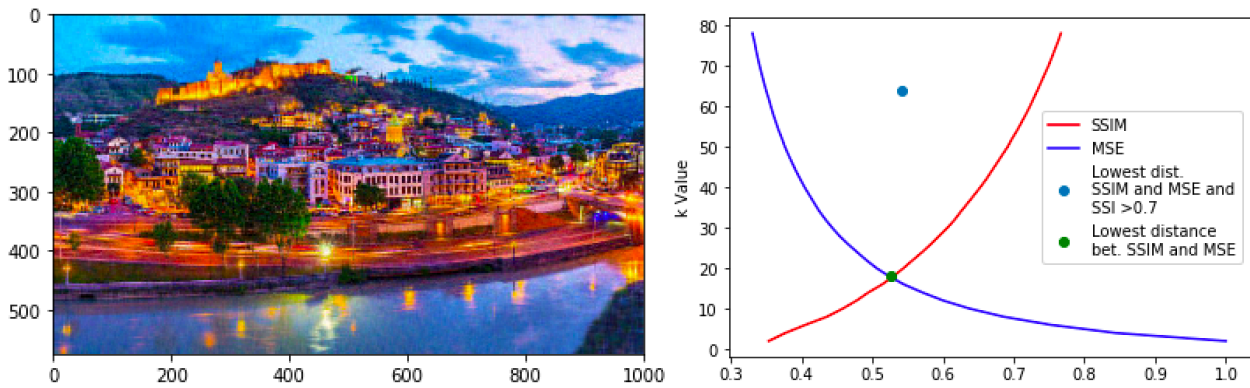


Figure 10: Immagine compressa con $k=64$, $SSIM=0.73$, $MSE=0.34$

Tale immagine, ottenuta con $k=62$, risulta essere più leggera dell'immagine originale di 23.48 Kb (pesando quindi 113.8 Kb). Impostare un $SSIM$ più alto garantirebbe certamente una qualità più alta dell'immagine ma non si otterrebbe nessuna diminuzione in termini di Kb significativa. Usualmente, i programmi di compressione vengono utilizzati dagli utenti poiché molte volte per il trasferimento di immagini, vengono imposti dei limiti massimi legati alle dimensioni. Per questo motivo, il programma sviluppato consente all'utente oltre che ottenere l'immagine compressa in maniera ottimale (Figura 10), anche di ottenere una immagine appositamente compressa specificando il limite di peso in Kb .

```

2)Eseguire SVD inserendo un limite specifico di peso di Kb
inserire scelta -> 2

Inserire il nome della foto es: foto.jpg
(questa deve essere posizionata nella stessa cartella)
inserire nome -> foto.jpg
-----

Inserire il limite dei Kb
Es: 80
inserire kb -> 120

Attendere...L'esecuzione può richiedere qualche minuto

Immagine Compressa

Kb image: 120.017
Value of K: 88

```

Figure 11: Esempio di esecuzione con specifica del limite in Kb

Bibliografia

- [1] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*, SIAM.
- [2] H. K. M. Y. H. W. S. M. PETER NDAJAH, «SSIM Image Quality Metric for Denoised Images,» *ADVANCES in VISUALIZATION, IMAGING and SIMULATION*, pp. 53-57, 2010.
- [3] D. Z. Alain Horé, «Image quality metrics: PSNR vs. SSIM,» *2010 International Conference on Pattern Recognitio*, pp. 2366-2369, 2010.
- [4] A. C. B. Sumohana Channappayya, «Structural Similarity Index Based Optimization,» *Encyclopedia of Multimedia*, 2008.
- [5] G. P. R. e. al., «Structural similarity index family for image quality assessment in radiological images,» *Journal of Medical Imaging*, vol. 4, n. 3, 2017.
- [6] S. W. e. al., «RATE-SSIM OPTIMIZATION FOR VIDEO CODING,» *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011.
- [7] J. R. F. W. A. Poole, «Image Quality Assessment Using the SSIM and the Just Noticeable Difference Paradigm,» *Lecture Notes in Computer Science*, vol. 8019, n. 978-3-642-39359-4, pp. 22-30, 2013.