

Image OSN Identification

Abstract

Questo paper mostra come eseguire la classificazione d'immagini digitali in base all'Online Social Network (OSN) da cui sono state scaricate, utilizzando principalmente come caratteristica discriminante le differenti tabelle di quantizzazione JPEG applicate dagli OSN, come Facebook e Badoo. Inoltre, si mostra il processo di elaborazione delle immagini applicato da questi OSN.

Il paper termina mostrando che non è possibile nascondere le informazioni nelle immagini pubblicate su Facebook adoperando tecniche standard di steganografia, e mostra come applicare tali strumenti su Google Picasa.

Questo paper è un riepilogo del lavoro svolto, maggiori informazioni sul contenuto di questo documento sono reperibili nella tesina allegata.

1. Introduzione

In generale, si definiscono Online Social Network (OSN) quei servizi Web che permettono la creazione di un profilo pubblico o semi-pubblico, l'articolazione di una lista di contatti, la possibilità di scorrere la lista degli amici dei propri contatti.

Oggigiorno le persone pubblicano le immagini sugli OSN per “rivivere” i ricordi in un secondo momento, trasmettere le sensazioni di un evento speciale a chiunque. Infatti, questi servizi stanno diventando sempre più strumento di “condivisione di ricordi”. Per questo le immagini ricoprono un ruolo sempre più importante, ad esempio, Facebook ha intensificato le funzionalità attinenti la pubblicazione di immagini digitali. In generale, le immagini possono conservare informazioni utili, come la descrizione di una traccia nella scena del crimine, e descrivono un evento in precedenza accaduto.

In questo scritto si parlerà di come classificare le immagini digitali scaricate dagli OSN, come Facebook e Badoo, i due principali OSN più utilizzati in Italia [1]. Inoltre, si farà cenno al neonato OSN Google Plus (*Google +*), ancora in fase di testing. Basti pensare che Facebook (<http://www.facebook.com/>) possiede in tutto il mondo oltre 500 milioni di utenti attivi (anno 2010) e il numero di foto caricate all'anno è 36 miliardi [2]. Mentre, Badoo (<http://www.badoo.it/>) ha oltre 120 milioni d'iscritti (Aprile 2011) e 4.3 milioni di foto e video caricati quotidianamente [3]. In Google Plus (<http://plus.google.com/>) sono stati superati i venti milioni di utenti in soli ventiquattro giorni dalla nascita, contro i più di 1000 giorni che sono stati necessari a Twitter e Facebook [11].

Esistono parecchi scenari in cui è possibile applicare un metodo che consente di classificare le immagini provenienti dagli OSN. Infatti, gli organi giudiziari hanno necessità a verificare se una certa immagine è stata scaricata da un determinato OSN e un consulente forense deve dare risposta a vari quesiti attinenti le immagini digitali. In questo scritto cercheremo di rispondere al quesito: “Data un'immagine digitale, da quale OSN è stata scaricata?”.

L'attività svolta rientra nel ramo della "Digital Image Forensic": "*Forensic Image analysis is the application of image science and domain expertise to interpret the content of an image or the image itself in legal matters*" (SWGIT) [4].

Nel seguente scritto tratteremo principalmente dei metadati presenti nelle immagini scaricate dagli OSN, e in particolare delle tabelle di quantizzazione JPEG presenti nelle foto scaricate. Infatti, adopereremo tale caratteristica come "un'impronta" dell'OSN, siccome ogni OSN utilizza tabelle personalizzate. Si adopera un metodo che dato un fattore di qualità restituisce le appropriate tabelle di quantizzazione JPEG.

Il paper è organizzato nel seguente modo: inizialmente parleremo del nome delle immagini digitali scaricate dagli OSN, in seguito discuteremo dei formati delle immagini utilizzate dai OSN di riferimento. Nella sezione quattro si parlerà della risoluzione delle immagini e delle anteprime utilizzate da facebook, badoo e infine google+. Nella sezione cinque tratteremo i metadati che vengono utilizzati dai OSN di riferimento il particolare nella sezione successiva introduciamo il metadato comment utilizzato da badoo. Infine, tratteremo delle tabelle di quantizzazione e l'applicazione realizzata. Ultimo argomento è la steganografia.

2. Nome immagini digitali scaricate dagli OSN

Facebook

Le immagini in Facebook sono raggruppate in album. Ogni album e di conseguenza ogni immagine sono etichettati da un identificativo generato da un algoritmo proprietario. L'URL di un'immagine è composta dall'identificativo dell'utente e dall'identificativo dell'album dove risiede l'immagine in esame. Supponiamo che l'id del nostro profilo sia 2221314, e ipotizziamo di aver caricato un'immagine sul nostro album, dove l'id dell'album dato da Facebook è 1111444111.0009. L'url dell'immagine conterrà *a.1111444111.0009.2221314*. Se l'id dell'immagine in esame generato da Facebook è 6666666, l'url che permette di raggiungere l'immagine conterrà la stringa *fbid=6666666&set=a.1111444111.0009.2221314*.

Invece, all'atto del download dell'immagine sul computer, il nome dell'immagine conterrà informazioni che racchiudono l'id associato all'immagine, generato da Facebook nella fase di upload, e l'id dell'utente che ha caricato l'immagine. Un esempio può essere: *6666666&set=a.1111444111.0009.2221314*, dove come indicato precedentemente il campo iniziale è l'id dell'immagine e quello finale è l'id dell'utente. Da notare che il processo di generazione di tali informazioni è sconosciuto all'utente.

Badoo

Badoo a differenza di Facebook non permette agli utenti di scaricare le immagini presenti nel profilo. Un modo per farlo è quello di aprire il sorgente della pagina html che contiene l'immagine e prendere il campo *URL SCR* associato all'immagine. Una volta scaricata l'immagine secondo la procedura presentata, essa avrà come nome l'identificativo generato da Badoo nella fase di upload.

Google+

Le immagini che l'utente inserisce all'interno dell'OSN saranno presenti all'interno del servizio d'*image hosting* Google Picasa. Adesso, per scaricare le immagini pubblicate (e non un'anteprima dell'immagine) bisogna accedere all'album contenete l'immagine presente in Google Picasa. Il nome dell'immagine scaricata è uguale al nome dell'immagine originaria, ossia prima dell'upload sull'OSN.

3. Formato delle immagini pubblicate dagli OSN

Il seguente test si pone come obiettivo comprendere i diversi formati delle immagini che sono supportati dagli OSN di riferimento: Facebook, Badoo e Google Plus.

Facebook

Sono state individuate otto immagini di riferimento, due per ogni formato (immagine con risoluzione minore rispetto a quella accettata dall'OSN e un'altra con risoluzione maggiore; maggiori informazioni sulla risoluzione delle immagini sono reperibili nel successivo paragrafo):

- PNG
- BMP
- GIF
- JPG

Una volta individuate le immagini, si è cercato di caricare le immagini all'interno dell'OSN. Dalle operazioni effettuate si evince che l'OSN Facebook accetta immagini con formato PNG, GIF e JPG e le memorizza al suo interno nel formato JPG.

Per verificare l'operazione di conversione da PNG/GIF nel formato JPG è stato compiuto un test. Sono state caricate dieci immagini digitali per ogni formato con una risoluzione di 640x480 pixel all'interno dell'OSN. In seguito, le immagini sono state scaricate ed è stata eseguita la stessa operazione di conversione sulle immagini originarie, ossia quelle di upload, con due software di riferimento: GIMP2 e IrfanView (utilizzando lo stesso fattore di qualità JPEG adoperato da Facebook, ossia 85; maggiori informazioni sono reperibili nelle successive sezioni). Le seguenti tabelle mostrano in media i pixel che risultano differenti confrontando le 10 immagini digitali per ogni formato scaricate da Facebook con quelle realizzate con GIMP2 e IrfanView. All'interno della tabella saranno presente le colonne che indicano il numero medio dei pixel che variano per ogni canale, il numero medio dei pixel che variano per ogni canale lungo il contorno e la differenza massima tra i pixel che esiste tra l'immagine di facebook e quella in analisi.

CONVERSIONE PNG → JPG

	Rossi tot. Modificati	Rossi cont. Modificati	Verdi tot Modificati	Verdi cont Modificati	Blu tot Modificati	Blu cont Modificati	Val max
GIMP2	0	0	0	0	0	0	0
IrfanView	44884,3	3,9	53659,2	11,2	53311,3	13,4	1

Tabella 1. Conversione PNG in JPG.

CONVERSIONE GIF → JPG

Prima della conversione di formato con entrambi i software, è stata compiuta anche la conversione nello spazio RGB.

	Rossi tot Modificati	Rossi cont Modificati	Verdi tot Modificati	Verdi cont Modificati	Blu tot Modificati	Blu cont Modificati	Val max
GIMP2	0	0	0	0	0	0	0
IrfanView	46818,6	9,9	52321,4	9,1	50935,9	6,6	1

Tabella 2. Conversione GIF in JPG.

Da entrambe le tabelle si evince che GIMP2 utilizza la stessa tecnica di conversione di formato di Facebook, mentre IrfanView no.

Badoo

Sono state individuate otto immagini di riferimento, due per ogni formato (immagine con risoluzione minore rispetto a quella accettata dall'OSN e un'altra con risoluzione maggiore):

- PNG
- BMP
- GIF
- JPG

Una volta individuate le immagini, si è cercato di caricare le immagini all'interno di Badoo. Dalle operazioni effettuate si evince che l'OSN Badoo accetta immagini con formato PNG, GIF, BMP e JPG e le memorizza al suo interno nel formato JPG.

Per verificare l'operazione di conversione da PNG/GIF/BMP nel formato JPG è stato eseguito un test. Sono state caricate dieci immagini digitali per ogni formato con una risoluzione di 640x480 all'interno dell'OSN. In seguito, le immagini sono state scaricate ed è stata eseguita la stessa operazione di conversione sulle immagini originarie con i due software di riferimento: GIMP2 e IrfanView (utilizzando lo stesso fattore di qualità di Badoo, ossia 91). Le seguenti tabelle mostrano in media i pixel che risultano differenti confrontando le 10 immagini digitali per ogni formato scaricate da Badoo con quelle realizzate con GIMP2 e IrfanView. All'interno della tabella saranno

presente le colonne che indicano il numero medio dei pixel che variano per ogni canale, il numero medio dei pixel che variano per ogni canale lungo il contorno e la differenza massima tra i pixel che esiste tra l'immagine di badoo e quella in analisi.

CONVERSIONE PNG → JPG

	Rossi tot Modificati	Rossi cont Modificati	Verdi tot Modificati	Verdi cont Modificati	Blu tot Modificati	Blu cont Modificati	Val max
GIMP2	71193,8	20,4	75561,2	21,4	82747,2	24,8	1
IrfanView	31341,9	3,1	33211,3	1,6	40767,6	1,2	1

Tabella 3. Conversione PNG in JPG.

CONVERSIONE GIF → JPG

Prima di effettuare la conversione di formato con entrambi i software, è stata effettuata anche la conversione dei colori nello spazio RGB.

	Rossi tot Modificati	Rossi cont Modificati	Verdi tot Modificati	Verdi cont Modificati	Blu tot Modificati	Blu cont Modificati	Val max
GIMP2	79690,1	20,6	83251,8	17,4	89584,9	18,6	1
IrfanView	31211,7	2,1	34938	3,3	41374,4	6,4	1

Tabella 4. Conversione GIF in JPG.

CONVERSIONE BMP → JPG

	Rossi tot Modificati	Rossi cont Modificati	Verdi tot Modificati	Verdi cont Modificati	Blu tot Modificati	Blu cont Modificati	Val max
GIMP2	39820,7	3,1	41089,7	6,4	44117,9	7,5	1
IrfanView	15965,7	0	16978,9	2,5	19887,7	4,1	1

Tabella 5. Conversione BMP in JPG.

Dall'analisi effettuata è possibile affermare che dalle tre operazioni di conversione, GIMP2 e IrfanView non realizzano la stessa immagine pubblicata su Badoo, a differenza di Facebook dove solo GIMP2 esegue la stessa conversione di formato.

Google+

Sono state individuate otto immagini di riferimento, due per ogni formato (immagine con risoluzione minore di quella accettata dall'OSN ed un'altra con risoluzione maggiore):

- PNG
- BMP
- GIF

- JPG

Una volta individuate le immagini, si è cercato di caricare le immagini all'interno dell'OSN. Dall'operazione effettuata si evince che il formato di output dell'OSN dipende dal quello in input, come mostrato dalla seguente tabella.

Formato Input	Dimensione del lato maggiore	Formato Output	Resize
PNG	>2048	PNG	Si
PNG	<2048	PNG	No
JPEG	>2048	JPEG	Si
JPEG	<2048	JPEG	No
BMP	>2048	JPEG	Si
BMP	<2048	BMP	No
GIF	>2048	PNG	Si
GIF	<2048	GIF	No

Tabella 6. Formati immagine accettati da Google Plus.

4. La risoluzione delle immagini pubblicate

Con il termine pixel, contrazione della locuzione inglese *picture element*, s'indica ciascuno degli elementi che compongono la rappresentazione di un'immagine digitale. Il numero di pixel in un'immagine, talvolta chiamato "risoluzione dell'immagine" determina la quantità di dettagli fini che possono essere rappresentati.

Facebook

Facebook permette di caricare le immagini in due modalità differenti: "*high resolution*" con una risoluzione di 2048 pixel sul lato maggiore e "*standard resolution*" con una risoluzione di 720 pixel sul lato più grande dell'immagine. Con l'opzione standard, se si carica un'immagine con il valore maggiore della risoluzione più grande di 720 pixel, si effettua il *resize* (ridimensionamento) portando l'immagine su tale risoluzione, altrimenti non si applica nessun algoritmo di ridimensionamento. Con l'altra opzione, se si carica un'immagine con il valore maggiore della risoluzione più grande di 2048 pixel, si effettua il *resize* portando l'immagine su tale risoluzione; altrimenti non si applica nessun algoritmo di ridimensionamento. Inoltre, in Facebook sono presenti molte anteprime d'immagini, come le *thumbnail* (miniature d'immagini) associate all'immagine di profilo di un utente, le quali hanno una risoluzione di 50x50 pixel. Inoltre, precisiamo che le *thumbnail images* (anteprime d'immagini) di risoluzione 50x50 pixel in Facebook hanno il campo Comment con valore "*CREATOR: gd-jpeg v1.0 (using IJG JPEG v62), quality = 95.*", ossia, si è adoperato GD Graphics Library, un tool di generazione di grafica dinamica per siti Web, o meglio

una libreria di codici open source per la creazione dinamica d'immagini da parte dei programmatori. Facebook permette di caricare l'immagine che vogliamo associare al nostro profilo a patto che sia di dimensioni inferiori a 4MB e di risoluzione con larghezza(x) maggiore di 180 pixel. al quanto detto, i nostri test si suddividono in tre categorie. I quali sono:

- Immagini con larghezza maggiore di 180 pixel e altezza minore di 180 pixel. Infatti, le immagini che abbiamo utilizzato nella fase di test avevano risoluzione: 1156x25pixel, 1156x45pixel, 1156x50pixel e 729x90.
- Immagini con larghezza maggiore di 180 pixel e altezza maggiore di 180pixel. le immagini che sono state utilizzate per eseguire il test avevano risoluzione: 300x300pixel, 200x200pixel, 180x180pixel e 1500x1061.
- Immagini con larghezza minore di 180pixel e altezza maggiore di 180 pixel. Tali immagini sono state respinte dal sociale network.

Risoluzione	Risultato test
Immagini larghezza \geq 180pixel e altezza $<$ 180pixel	Le immagini in prima fase hanno attraversato un algoritmo di resize non noto agli utenti. invece nella seconda fase, le immagini hanno attraversato un processo di crop. In base ai test che sono stati effettuati tale processo viene effettuato al centro delle immagini. E infine, siccome l'altezza delle immagini utilizzate per il test erano minori di 180pixel, le immagini attraverso ancora una volta l'algoritmo di resize per riportarle a 50 x50pixel .
Immagini larghezza $<$ 180pixel e altezza \geq 180pixel	Il social network risponde con "Prova a usarne una più quadrata".
Immagini con larghezza \geq 180pixel e altezza \geq 180pixel	In questa categoria si trovano le immagini che hanno risoluzione quadrate o rettangolare con larghezza e altezza superiore a 180pixel. Data la discrezione dei dettagli contenute nelle immagini ottenuto dopo il caricamento, possiamo dedurre che il processo utilizzato da facebook è identico al primo facendo solo le prime due fasi. Cioè, nella prima fase le immagini attraversano l'algoritmo di resize, tale algoritmo non è noto all'utente. E successivamente, il processo di crop ancora una volta al centro delle immagini. Poiché le immagini che si trovano in questa categoria hanno risoluzione maggiori di 180pixel sia d'altezza che di larghezza, dopo il crop non hanno bisogno di essere riportate alla risoluzione 50x50pixel

Tabella 7 indica le operazioni sulle diverse modalità d'inserimento

Per verificare quale algoritmo di resize adoperava Facebook è stato realizzato il seguente test prevede di caricare un'immagine JPEG sull'OSN e in seguito scaricarla. Successivamente, è stata presa l'immagine di partenza ed è stato effettuato il resize utilizzando tutti gli algoritmi a disposizione dei software Gimp2 e IrfanView, per verificare la corrispondenza in pixel della stessa immagine caricata sull'OSN (nel salvataggio è stato utilizzato lo stesso fattore di qualità adoperato facebook ossia 85 e impostazioni di default dei software; maggiori informazioni sono reperibili nella sezione 5). Dal seguente test utilizzando il software Matlab si evince che le immagini generate da Facebook sono differenti da quelle elaborate da tali software di resizing in termini di pixel. Questo significa che Facebook non utilizzano gli algoritmi di resize presenti in tali software.

Badoo

In Badoo, se l'immagine da pubblicare ha il valore maggiore della risoluzione più grande di 920 pixel, si compie l'operazione di ridimensionamento, adattando l'immagine su tale valore, altrimenti l'immagine è inalterata dal punto di vista del resize. Anche Badoo associa all'immagine inserite le thumbnail, ma a differenza di Facebook sono presenti diversi step di risoluzione:

- Thumbnail Xx48 pixel
- Thumbnail Xx96 pixel
- Thumbnail Xx192 pixel
- Thumbnail Xx300 pixel

Dove la notazione Xx48 indica che la dimensione dell'anteprima è un valore variabile, X, per 48. Precisiamo che Badoo accetta immagini con risoluzione minimo 200x200 pixel.

Per verificare quale algoritmo di resize adoperava Badoo è stato realizzato il seguente test prevede di caricare un'immagine JPEG sull'OSN e in seguito scaricarla. Successivamente, è stata presa l'immagine di partenza ed è stato effettuato il resize utilizzando tutti gli algoritmi a disposizione dei software Gimp2 e IrfanView, per verificare la corrispondenza in pixel della stessa immagine caricata sull'OSN (nel salvataggio è stato utilizzato lo stesso fattore di qualità adoperato badoo ossia 91 e impostazioni di default dei software; maggiori informazioni sono reperibili nella sezione 5). Dal seguente test utilizzando il software Matlab si evince che le immagini generate da Badoo sono differenti da quelle elaborate da tali software di resizing in termini di pixel. Questo significa che Badoo non utilizzano gli algoritmi di resize presenti in tali software.

Siccome dai dati Exif di tutte le anteprime di 50x50 pixel abbiamo reperito che queste vengono generate dal software GD Library, abbiamo effettuato successivi test di resizing per Facebook adoperando tale libreria. Quindi, come prima operazione è stato realizzato lo script per caricare l'immagine, ridimensionarla, e quindi salvarla, adoperando le funzioni di tale libreria. Nella funzione di salvataggio è stato impostato il fattore di qualità associato all'OSN di riferimento (85).

Si mostra il codice PHP utilizzato per questo esperimento.

```
<?php
// File and new size
$filename = 'Immagine.jpg';

// Content type
header('Content-Type: image/jpeg');

// Get new sizes
list($width, $height) = getimagesize($filename);
$newwidth = 2048;
$newheight = 1365;

// Load
$img = imagecreatetruecolor($newwidth, $newheight);
$source = imagecreatefromjpeg($filename);

// Resize
imagecopyresized($img, $source, 0, 0, 0, 0, $newwidth, $newheight, $width, $height);

// Output
//imagejpeg($img);
imagejpeg($img, 'IMG2.JPG', 85);
?>
```

Una volta salvata l'immagine sul disco locale, ed è stata confrontata con quella scaricata da Facebook. Anche in questo caso, l'immagine pubblicata da Facebook e quella ridimensionata da GD sono differenti in termini di pixel.

Lo stesso esperimento è stato eseguito anche sulle anteprime 50x50 pixel, ma abbiamo registrato lo stesso esito.

Notiamo che ogniquale volta è utilizzata la libreria GD, all'interno del metadato comment dell'immagine compare il nome della libreria utilizzata. Infatti, è stato effettuato un semplice test che caricava l'immagine e la salvava sul disco.

Google+

In Google+ la risoluzione consentita per caricare le immagini di 2048 sul lato maggiore. Quindi, se l'immagine ha un lato maggiore più piccolo di 2048 google+ carica direttamente l'immagine originale senza compiere nessuna operazione, altrimenti applica l'algoritmo di resize per far sì che l'immagine digitale possa essere presente all'interno del online social network.

Infine, Picasa, da cui deriva Google Plus, permette una gestione migliore sulla risoluzione legata alle immagini, in quanto esso non è un online social network ma un servizio di foto sharing. Esso permette di caricare le immagini con diverse opzioni legate alla dimensione:

- Dimensione originale
- Consigliate: 1600 pixel (per stampa, salvaschermo e condivisione)
- Medie: 1024 pixel (per condivisione)
- Piccole: 640 pixel (per blog e pagine web)

La seguente tabella mostra il caricamento di un'immagine digitale con una risoluzione di 3888x2592 inserita nei diversi online social network e sul sito di foto sharing picasa.

<i>Servizio</i>	<i>Immagine modificata</i>	<i>Risoluzione immagine pubblicata</i>
Facebook high resolution	Si	2048x1365
Facebook standard resolution	Si	720x480
Facebook Thumbnail di profilo	Si	50x50
Badoo	Si	920x613
Badoo Thumbnail 1	Si	72x48
Badoo Thumbnail 2	Si	144x96
Badoo Thumbnail 3	Si	288x192
Badoo Thumbnail 4	Si	450x300
Picasa Web	No	3888x2592
Picasa dimensione originale	No	3888x2592
Picasa dimensione consigliate	Si	1600x1067
Picasa dimensione medie	Si	1024x683
Picasa dimensione piccole	Si	640x427
Google+	Si	2048x1365
Twitpic	No	3888x2592

5. Metadati presenti nelle immagini scaricate dagli OSN

Per estrarre i metadati dalle immagini scaricate dagli OSN abbiamo utilizzato vari tool, tra questi ExifTool [6], una libreria che consente di leggere e scrivere i metadati di un'ampia varietà di file (es. JPEG).

Facebook

Se carichiamo un'immagine su Facebook e confrontiamo l'immagine di partenza (presente sul nostro computer) con quella pubblicata, rileviamo le seguenti differenze: il nome dell'immagine cambia, la dimensione totale del file è sempre inferiore a quello di partenza (principalmente per

mezzo della compressione JPEG e dell'eliminazione dei dati EXIF sull'immagine), il tipo di file è sempre JPEG, e cambia eventualmente la risoluzione in pixel (vedere sezione 2). Facebook applica durante il processo di compressione JPEG alcuni parametri fissi al variare delle immagini, come "Baseline DCT", "Huffman coding", tabelle di quantizzazione fissate (maggiori informazioni nella sezione 5) e 24 bit per ogni pixel RGB. Inoltre, adopera un particolare metodo di sottocampionamento sull'immagine originaria (YCbCr4:2:0 (2 2)), elimina i dati EXIF presenti sull'immagine originaria e aggiunge informazioni di *rendering* nei metadati (ICC Profile). Infatti, Facebook inserisce i metadati riguardanti l'International Color Consortium (ICC) Profile con lo scopo di visualizzare l'immagine in modo ottimale su diversi dispositivi.

Badoo

Se si carica un'immagine su Badoo e si confrontano i metadati dell'originale con quelli dell'immagine pubblicata, si evince: il nome dell'immagine cambia, la dimensione dell'immagine è inferiore nell'immagine scaricata (causa compressione JPEG ed eliminazione EXIF), l'immagine scaricata è nel formato JPEG. Inoltre, l'immagine di Badoo ha una risoluzione concorde al resizing esplorato nella precedente sezione. Badoo applica durante il processo di compressione JPEG alcuni parametri fissi al variare delle immagini, come "Baseline DCT", "Huffman coding", tabelle di quantizzazione fissate (maggiori informazioni nella sezione 5) e 24 bit per ogni pixel RGB. Inoltre, adopera un particolare metodo di sottocampionamento sull'immagine originaria (YCbCr4:2:0 (2 2)), come Facebook. Oltre a ciò, dai metadati si evince che Badoo utilizza il campo COMMENT (maggiori informazioni nella sezione 6).

Google+

La gestione dei metadati all'interno del OSN Google+ plus si differenzia in base alla risoluzione con cui viene caricata l'immagine al suo interno. Se le immagini hanno una risoluzione maggiore di 2048, quindi viene applicato l'algoritmo di resize i metadati possono esser visualizzati solo all'interno del social network e non sono presenti nell'immagine che viene scaricata. Mentre, se il lato maggiore ha una risoluzione minore di 2048 allora l'immagine digitale viene lasciata inalterata e quindi i metadati sono uguali a quelli dell'immagine originale.

5.1 Metadato: Campo Comment Badoo

L'OSN Badoo quando un utente scarica un'immagine (dal proprio profilo o da un profilo di un altro utente) esso inserisce un metadato chiamato *Comment* il quale permette di ottenere informazioni utili del proprietario dell'immagine. Infatti, il campo *Comment* è una stringa esadecimale, la quale contiene informazioni sul contenuto dell'immagine, sul profilo dell'utente da cui è stata scaricata l'immagine e infine sul nome originario dato da Badoo all'immagine. Questo vuol dire che una volta scaricata l'immagine e modificato il nome è possibile ottenere sempre quello di badoo. Le informazioni possono essere utili per individuare l'immagine all'interno del OSN Badoo. Per esempio, Scaricando l'immagine dal seguente indirizzo <http://badoo.com/0238508272/p50176> possiamo scaricare tale immagine. Consideriamo ora di voler ottenere nuovamente l'URL

dell'immagine all'interno di Badoo questo è possibile analizzando il campo Comment dell'immagine (basta analizzare solo la parte evidenziata):

zU09393951E28660D0D0FD5D5D59292941E286D783F2CD7D7D78D8C911D2765131416CEC
DC9828284946750121315C1C2C40E3758F00000C400

Effettuando una conversione da esadecimale a decimale, otteniamo:

0E3758F0 → 0238508272 (utente)

0000C400 → 50176 (immagine)


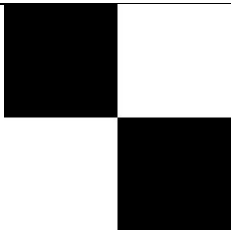
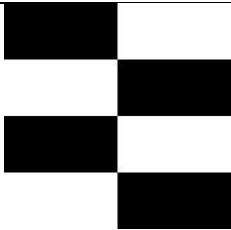
NOME IMG → 50176_920 (920 = risoluzione immagine; non anteprima)

[p = picture]

Con le seguenti informazioni siamo in grado di ottenere nuovamente URL:
<http://badoo.com/0238508272/p50176>

Precisiamo che la prima parte del campo comment non varia, se si pubblica un'immagine in due differenti momenti.

Il campo *Comment* oltre ad esser presente all'interno dell'immagine è presente anche nell'anteprima dell'immagine di dimensione 400 x 300.

Tipo Immagine	Immagine	Nome Immagine Badoo	Risoluzione in pixel
2		86001_920.jpg	200x200
4		86004_920.jpg	200x200
8		86012_920.jpg	200x200

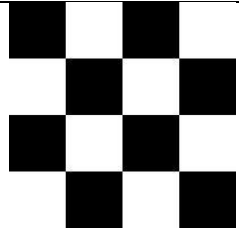
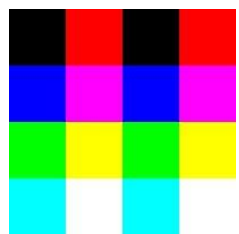
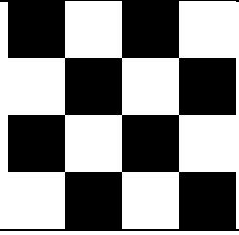
16		86036_920.jpg	200x200
16 Color		86059_920.jpg	200x200
16 bis		87800_920.jpg	400x400

Tabella 8 indica le diverse immagini utilizzate per il test

In seguito mostreremo come cambia la prima parte del campo Comment, al variare dell'immagine pubblicata su Badoo. In particolare abbiamo creato artificialmente con Matlab le immagini mostrate nella successiva tabella, e le abbiamo caricate su Badoo sul profilo di Alfredo Pierro, che possiede id=0238508272. Abbiamo utilizzato immagini originali quadrate di 200x200 pixel e una 400x400 pixel.

Abbiamo adoperato tali immagini poiché, secondo noi, erano le più adatte a mostrare la mutabilità del campo comment.

Per ogni immagine scaricata da Badoo, la prossima tabella, mostra l'intero campo Comment, il quale contiene le informazioni cromatiche, l'ID dell'utente e il nome dell'immagine.

Tipo Immagine	Campo Comment
2	zU0000000000000000FFFFFFFFFFFFFFFF000000000000FFFFFFFFFFFFFFFF000000000000FFFFFFFFFFFFFFFF000000000000E3758F000014FF1
4	zU0000000000000000FFFFFFFFFFFFFFFF000000000000FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF000000000000FFFFFFFFFFFFFFFF000000000000E3758F000014FF4
8	zU0000000000000000FF000000000000000000000000FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF000000000000E3758F000014FFC
16	zU00000000FFFFFFFF000000FFFFFFFFFFFFFFFF000000FFFFFFFF0000000000000000FFFFFFFF000000FFFFFFFFFFFFFFFF000000FFFFFFFF00000000E3758F000015014
16 Color	zU00000000FE0000000000FE00000000FEFF00FE0000FEFF00FE00FF01FFFF0000

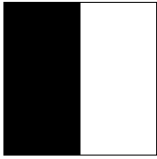
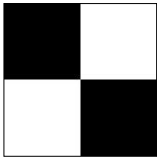
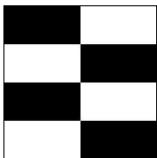
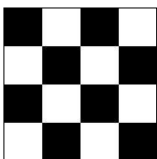
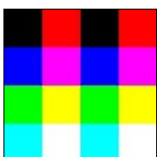
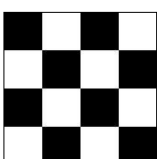
Tipo Immagine	Immagine	Nome Immagine	Prima Parte Campo Comment
2		86001_920.jpg	000000 000000 FFFFFFFF FFFFFFFF 000000 000000 FFFFFFFF FFFFFFFF 000000 000000 FFFFFFFF FFFFFFFF 000000 000000 FFFFFFFF FFFFFFFF
4		86004_920.jpg	000000 000000 FFFFFFFF FFFFFFFF 000000 000000 FFFFFFFF FFFFFFFF FFFFFFF FFFFFFFF 000000 000000 FFFFFFF FFFFFFFF 000000 000000
8		86012_920.jpg	000000 000000 FFFFFFFF FFFFFFFF FFFFFFF FFFFFFFF 000000 000000 000000 000000 FFFFFFFF FFFFFFFF FFFFFFF FFFFFFFF 000000 000000
16		86036_920.jpg	000000 FFFFFFFF 000000 FFFFFFFF FFFFFFF 000000 FFFFFFFF 000000 000000 FFFFFFFF 000000 FFFFFFFF FFFFFFF 000000 FFFFFFFF 000000
16 Color		86059_920.jpg	000000 FE0000 000000 FE0000 0000FE FF00FE 0000FE FF00FE 00FF01 FFFF00 00FF01 FFFF00 00FFFF FFFFFFFF 00FFFF FFFFFFFF
16 bis		87800_920.jpg	000000 FFFFFFFF 000000 FFFFFFFF FFFFFFF 000000 FFFFFFFF 000000 000000 FFFFFFFF 000000 FFFFFFFF FFFFFFF 000000 FFFFFFFF 000000

Tabella 11 indica il comment associato alle immagini caricate e poi scaricate

Analizzando la prima parte del campo comment, rileviamo che la struttura di ogni immagine si rispecchia nel corrispondente campo comment. Infatti, come si nota dalla tabella, abbiamo espresso la prima parte del campo Comment in una particolare notazione, volta ad associare ogni “pixel comment” alla corrispondente zona nell’immagine.

Nell’immagine 86036_920.jpg, i 16 “pixel del comment” sintetizzano i colori dei sedici quadrati nell’immagine. La stessa cosa vale per l’immagine a colori, 86059_920.jpg, dove ogni “pixel comment” rappresenta il colore più diffuso nella corrispondente zona, ad esempio, al secondo quadrato nella prima riga dell’immagine partendo da sinistra, è associato il colore rosso, infatti in tale zona dell’immagine ci sono molti pixel che hanno il colore rosso. Precisiamo che ogni singolo

quadrato nell'immagine a colori non contiene tutti i pixel settati allo stesso valore, perché è stata applicata la compressione JPEG da Badoo, trasformazione che modifica i colori dei pixel nella cornice di ogni quadrato.

Siccome abbiamo visto che dividendo l'immagine in sedici blocchi è possibile rilevare l'informazione cromatica di ogni blocco, sono stati effettuati parecchi esperimenti volti a rilevare come calcolare ogni "pixel comment" associato ad un blocco. Tutti gli esperimenti che presentiamo sono stati svolti con l'ausilio di Matlab. I dettagli di tutti gli esperimenti, insieme con gli esiti sono mostrati nella tesina allegata.

Un primo test è stato quello di prendere un'immagine, pubblicarla e poi scaricarla da Badoo (immagine 87808_920.jpg). Una volta letto il campo comment, abbiamo estratto la parte del comment che racchiude i "pixel comment", ossia i colori. Analizzando tali informazioni sui colori, abbiamo compiuto vari esperimenti su ognuno dei sedici blocchi dell'immagine scaricata da Badoo, alla ricerca di come venivano generate le informazioni sui "pixel comment". In particolare, sapendo che ogni pixel comment è associato a un blocco, abbiamo cercato i pixel RGB nel blocco in esame che avevano come valore quello del "pixel comment".

Dall'analisi sull'immagine 87808_920.jpg si evince che il primo blocco ha sedici pixel che hanno come valore quello del corrispondente "pixel comment", mentre l'ultimo blocco analizzato non ha nessun pixel settato al valore del corrispondente "pixel comment". Il precedente esperimento è stato compiuto anche sull'immagine originale, ossia l'immagine che è stata pubblicata su Badoo (quella da cui deriva 87808_920.jpg). Precisiamo, che abbiamo utilizzato sempre le informazioni dei colori estratte dal campo comment dell'immagine pubblicata su Badoo (87808_920.jpg). In questo caso, nel primo blocco abbiamo trovato sedici pixel compatibili con il corrispondente "pixel comment", mentre nell'altro blocco abbiamo rilevato un solo pixel compatibile.

Lo stesso esperimento è stato compiuto anche su un'altra immagine scaricata da Badoo, e abbiamo verificato che non esiste nessuna correlazione sul fatto che il "pixel comment" di un blocco X è un determinato pixel nel blocco X.

Inoltre, sono stati compiuti parecchi esperimenti per verificare se il "pixel comment" del blocco X corrisponde alla media (oppure al valore massimo/minimo, o alla varianza) dei pixel dell'immagine nel blocco X, operando per ogni canale di colore. L'esecuzione di tale test ha dato esito negativo sia analizzando l'immagine originaria, sia quella pubblicata su Badoo.

Lo stesso esperimento è stato compiuto sull'immagine originale, in quanto non è da escludere che il campo Commenti sia stato calcolato sull'immagine prima dell'elaborazione da parte di Badoo. Quindi, da questi esperimenti si evince che il "pixel comment" associato a un blocco non può essere la media (oppure il valore massimo/minimo) dei colori nel corrispondente blocco.

Un altro esperimento è stato verificare se il "pixel comment" associato a un blocco X corrisponde al colore più diffuso nel blocco X. Anche questo esperimento ha dato esito negativo.

Riportiamo per completezza l'esito del test che verifica se ogni "pixel comment" è la media dei pixel nel corrispondente blocco. Sono state utilizzate 5 immagini JPEG con risoluzione 400x400 pixel. Una volta caricate le immagini su Badoo, abbiamo rilevato i seguenti campi comment dalle corrispondenti immagini scaricate.

ID Immagine	Immagine	Comment
1		zU0 A5C7F5A7C9F7B0CFFEB9C5D5 CAD3E4C8E0FC041A32C2C28E C8D0E3162E460E3056A6B7E3 0208184B6574070E2A001936 0E3758F0 0001570C
2		zU0 C1E0F2C9E6F8CCE8F3BEE1F5 975032C96936DB6E1BA13906 592913170501C0480BC24809 440D06561B0DB34C2D953117 0E3758F0 000157B5
3		zU0 FB4805CC0200FF721C900201BA01 069B0000DE1207FF811C5C04026C 0006D81A02F47413A804026F0102 E02409F44413 0E3758F0 00015B1C
4		zU0 52483E050100D9C7BBB7AEA7 8A837BDED4CBC1B4AC22241F 766E636C655F6F645E747067 645B4C8F88805450440E0D09 0E3758F0 00015B1E
5		zU0 0D500059800D001300061F02 020401788162E2E4E3082506 020E008A9060D8D36B040005 021D001D45067577C1000300 0E3758F0 00015B20

Tabella 12 analisi del campo comment su immagini generiche

Inoltre, abbiamo analizzato la prima parte del campo comment di ogni immagine. Infatti, la seguente tabella mostra per ogni immagine i "pixel comment". In particolare ogni tripla corrisponde al "pixel comment" in RGB del corrispondente blocco, espresso in esadecimale.

Immagine	Colori nel Comment (valori in esadecimale)											
1	A5	C7	F5	A7	C9	F7	B0	CF	FE	B9	C5	D5
	CA	D3	E4	C8	E0	FC	04	1A	32	C2	C2	8E
	C8	D0	E3	16	2E	46	0E	30	56	A6	B7	E3
	02	08	18	4B	65	74	07	0E	2A	00	19	36
2	C1	E0	F2	C9	E6	F8	CC	E8	F3	BE	E1	F5
	97	50	32	C9	69	36	DB	6E	1B	A1	39	06
	59	29	13	17	05	01	C0	48	0B	C2	48	09
	44	0D	06	56	1B	0D	B3	4C	2D	95	31	17
3	FB	48	05	CC	02	00	FF	72	1C	90	02	01
	BA	01	06	9B	00	00	DE	12	07	FF	81	1C
	5C	04	02	6C	00	06	D8	1A	02	F4	74	13
	A8	04	02	6F	01	02	E0	24	09	F4	44	13
4	52	48	3E	05	01	00	D9	C7	BB	B7	AE	A7
	8A	83	7B	DE	D4	CB	C1	B4	AC	22	24	1F
	76	6E	63	6C	65	5F	6F	64	5E	74	70	67
	64	5B	4C	8F	88	80	54	50	44	0E	0D	09
5	0D	50	00	59	80	0D	00	13	00	06	1F	02
	02	04	01	78	81	62	E2	E4	E3	08	25	06
	02	0E	00	8A	90	60	D8	D3	6B	04	00	05
	02	1D	00	1D	45	06	75	77	C1	00	03	00

Tabella 13 comment originale in esadecimale

Effettuando la conversione in decimale, otteniamo i seguenti “pixel comment”.

Immagine	Colori nel Comment (valori in decimale)											
1	165	199	245	167	201	247	176	207	254	185	197	213
	202	211	228	200	224	252	4	26	50	194	194	142
	200	208	227	22	46	70	14	48	86	166	183	227
	2	8	24	75	101	116	7	14	42	0	25	54
2	193	224	242	201	230	248	204	232	243	190	225	245
	151	80	50	201	105	54	219	110	27	161	57	6
	89	41	19	23	5	1	192	72	11	194	72	9
	68	13	6	86	27	13	179	76	45	149	49	23
3	251	72	5	204	2	0	255	114	28	144	2	1

	186	1	6	155	0	0	222	18	7	255	129	28
	92	4	2	108	0	6	216	26	2	244	116	19
	168	4	2	111	1	2	224	36	9	244	68	19
4	82	72	62	5	1	0	217	199	187	183	174	167
	138	131	123	222	212	203	193	180	172	34	36	31
	118	110	99	108	101	95	111	100	94	116	112	103
	100	91	76	143	136	128	84	80	68	14	13	9
5	13	80	0	89	128	13	0	19	0	6	31	2
	2	4	1	120	129	98	226	228	227	8	37	6
	2	14	0	138	144	96	216	211	107	4	0	5
	2	29	0	29	69	6	117	119	193	0	3	0

Le seguenti tabelle mostrano per ogni immagine la media aritmetica dei pixel in un blocco, sia analizzando l'immagine originaria sia quella scaricata da Badoo. Tale media è calcolata su ogni canale di colore del blocco in esame, e i valori sono espressi in decimali con la virgola. Abbiamo effettuato tale esperimento per mostrare che i “pixel comment” non sono calcolati in tale maniera.

ID Image	Media immagine originale (valori in decimale)											
1	125.7949	148.5877	165.6584	139.6424	176.338	215.4344	141.2305	172.0024	202.1375	133.8943	155.5457	169.0574
	195.7989	200.9888	194.9602	195.4908	213.6772	247.5894	124.3905	138.1716	150.6816	112.7959	130.4743	146.8167
	137.7408	151.4247	174.9295	81.9026	103.3656	123.103	31.9095	53.201	80.4256	88.5629	101.1137	129.129
	134.2555	145.874	164.0236	139.2583	153.1526	166.9341	66.3401	76.8579	94.1169	24.6392	35.3624	55.4733
2	172.9889	206.8707	236.8179	181.0285	213.4398	241.7748	179.6947	204.0089	224.0449	172.74	206.3323	235.6062
	166.2313	148.8933	141.1943	192.8174	157.2879	137.5702	188.7324	104.8589	54.0709	163.0579	120.5303	102.3017
	85.4904	37.3309	18.7087	117.1965	49.4231	22.1894	179.213	67.0135	14.7503	150.4412	52.9379	13.8072
	56.2884	18.1628	9.4575	79.584	26.1214	11.7943	104.3693	34.3423	11.959	85.2171	26.5165	9.7177
3	240.69	72.1468	15.6353	232.1785	50.3653	11.7306	230.9408	62.5156	16.0849	205.5548	43.3061	16.2738
	204.4755	30.2203	10.3792	163.5798	7.7401	5.8664	212.2096	35.9434	10.6928	205.8766	38.8576	12.483
	184.9864	20.256	8.0632	144.8817	7.655	9.4138	195.1983	24.8919	6.9109	214.9716	51.9535	14.0281
	216.2334	45.262	12.5892	183.8873	24.0039	9.2839	212.3904	33.6332	9.7294	190.829	33.3723	10.8874
4	127.1776	111.8	92.878	142.1168	133.3865	121.638	176.1412	161.5453	137.7284	139.1777	133.3463	113.3722
	132.0629	116.4569	100.3798	154.1338	145.8185	138.9526	161.8446	151.3396	144.0556	92.2598	89.1052	84.8471
	100.8596	82.823	68.8309	140.6474	129.4162	119.8563	109.7631	107.0199	101.4968	132.2999	121.019	107.3445
	122.2393	100.4494	83.0968	140.6897	127.5336	110.7791	54.8785	52.688	44.7341	79.3512	71.3237	59.9845
5												

ID Image	Media immagine originale (valori in decimale)											
	12.9379	54.1502	5.1542	47.1113	86.7761	25.8452	9.6037	22.8492	6.2373	16.2082	45.6534	7.6184
	8.2327	24.9369	5.4977	131.4707	141.948	109.0079	164.2707	167.5203	139.6915	20.1478	50.4673	11.7665
	26.1346	60.0403	17.2582	141.0708	146.5367	124.7925	153.5167	154.7765	98.4093	39.4338	64.0767	21.1784
	20.2716	62.9048	9.8504	26.1437	64.7091	18.5531	55.6419	84.5029	58.8981	46.9173	69.4619	44.4604

ID Image	Media immagine scaricata da Badoo (valori in decimale)											
1	125.9951	148.6232	166.1261	139.8302	176.3656	215.4565	141.2908	172.0157	202.2208	134.0381	155.5035	169.3992
	195.7304	200.9506	195.0284	195.4804	213.7096	247.4702	124.7039	138.1441	150.8626	113.17	130.4718	146.9416
	138.0498	151.3817	175.0261	82.4035	103.3279	123.1374	32.7278	53.1707	80.5815	89.0109	101.0815	129.209
	134.3589	145.8745	164.0308	153.1593	139.6507	166.9163	66.9417	76.829	94.1515	25.3809	35.3089	55.5669
2	172.9932	206.8755	236.8136	181.0377	213.449	241.784	179.7018	204.0113	224.0938	172.7634	206.3431	235.5948
	166.2194	148.898	141.2213	192.8701	157.2371	137.5805	188.7185	104.876	54.8071	163.0373	120.5199	102.5705
	85.4888	37.37	19.4026	117.1761	49.6079	23.8836	179.193	67.0235	15.4718	150.4348	52.8792	14.6994
	56.2047	18.5105	11.5798	79.5323	26.4219	14.2034	104.3243	34.472	14.2838	85.2099	26.7112	11.6695
3	239.5769	72.1492	16.987	231.3964	50.4765	13.0172	230.0674	62.6367	17.443	204.6318	43.7829	18.5165
	203.8217	30.5967	12.401	163.3	8.6341	7.5478	211.6473	36.1971	11.9757	205.3708	39.1349	13.8643
	184.4742	20.7806	9.9164	144.7391	9.0822	11.0344	194.7412	25.2818	8.5974	214.3392	52.1728	15.266
	215.6631	45.3977	14.0998	183.5236	24.5908	10.9558	211.9309	33.8876	11.1572	190.5281	33.6461	12.1464
4	127.1189	111.7565	92.8271	141.9935	133.3543	121.5534	175.9224	161.5549	137.6717	139.1331	133.3311	113.3769
	132.0939	116.4454	100.4323	153.9749	145.7669	138.8791	161.7796	151.2791	143.9826	92.273	89.1587	84.8769
	100.8095	82.8551	68.9087	140.6135	129.3891	119.8234	109.7266	107.0254	101.4903	132.2986	121.0008	107.3343
	122.1996	100.4671	83.1318	140.6609	127.5383	110.7731	54.8966	52.7066	44.7681	79.3436	71.3236	59.9329
5	13.102	54.0507	6.2449	47.3102	86.7153	26.6017	9.7682	22.9354	6.7348	16.2684	45.6065	8.0572
	8.3906	24.9574	5.8914	131.4512	141.9426	109.1326	164.2892	167.5566	139.7156	20.1717	50.4493	11.9602
	26.263	60.0298	17.7835	141.1236	146.5709	124.9643	153.551	154.7832	98.8921	39.468	64.0673	21.6251
	20.4824	62.8361	10.8258	26.3537	64.6831	19.4057	55.8268	84.5026	59.3838	46.9941	69.4585	44.7161

6. Metadato: QT-JPEG negli OSN

Lo standard JPEG ha come step di compressione principale l'adozione della “quantizzazione scalare dei coefficienti della DCT”[7], fase che adopera due tabelle, una per la componente cromatica (crominanza) dell'immagine e una per la componente luminosa (luminanza). Le tabelle di quantizzazione variano secondo l'implementazione dello standard JPEG, ciò è dovuto al fattore di qualità (QF) ossia il livello di compressione desiderato. Innanzitutto è opportuno evidenziare che

molti applicativi di elaborazione delle immagini utilizzano impostazioni che consentono di diminuire la dimensione in byte di un'immagine, a scapito della qualità. Solitamente, un valore denominato QF (*quality factor* o fattore di qualità), compreso tra 1 e 100, consente di modificare la qualità dell'immagine durante il processo di compressione JPEG. Valori bassi di QF producono una scarsa qualità dell'immagine compressa, ma la dimensione in byte della foto compressa è nettamente più piccola rispetto a quella di partenza. Lo standard JPEG adoperava una tabella di quantizzazione standard per la luminanza e una per la cromaticanza (JPEG spec section K.1), ma l'utente può utilizzarne altre secondo il suo piacimento, l'importante è memorizzarle sempre nell'immagine compressa. Tuttavia, esistono dei metodi di *scaling*, che dato in input un fattore di qualità (QF), solitamente compreso tra 1 e 100, restituiscono un'apposita tabella di quantizzazione per la luminanza/cromaticanza creata a partire dalla tabella standard della luminanza/cromaticanza. Il metodo di scaling che utilizzeremo in questo scritto è quello implementato da IJG (Independent JPEG Group) code library. In realtà, è possibile utilizzare un QF per generare la tabella di quantizzazione della luminanza e uno differente per generare la tabella di quantizzazione della cromaticanza.

Tale metodo restituisce una tabella di quantizzazione fornita nello standard JPEG (*JPEG spec, section K.1*) e scalata usando un *quality factor* (o *quality*). Il metodo di scaling mostrato nella figura sottostante è lo stesso di quello usato da IJG (Independent JPEG Group) code library [<http://www.ijg.org/>] un gruppo che sviluppa e distribuisce una libreria gratuita, ampiamente utilizzata, per la compressione delle immagini JPEG.

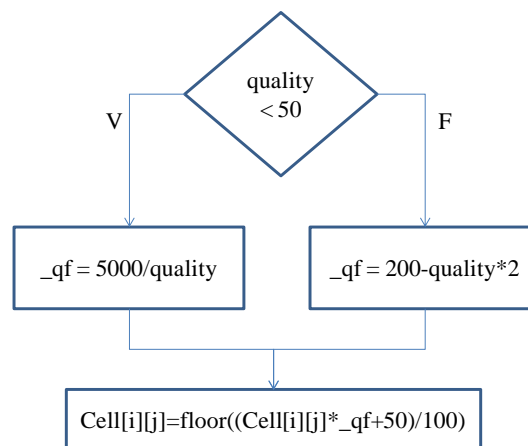


Figura 1 In figura, il metodo di scaling utilizzato da IJG (Independent JPEG Group) code library.

Quantization Table: LUMINACE							
16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

Quantization Table: CHROMINANCE							
17	18	24	47	99	99	99	99
18	21	26	66	99	99	99	99
24	26	56	99	99	99	99	99
47	66	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99

Figura 2. Queste sono le quantization table / matrix coefficients che sono raccomandate/suggerite nell'allegato dello standard JPEG.

Facebook

Dagli esperimenti realizzati, abbiamo notato che qualunque immagine caricavamo in Facebook, le tabelle di quantizzazione dell'immagine scaricata risultavano sempre le stesse. In particolare, abbiamo rilevato, sia con l'opzione "risoluzione standard", che con l'opzione "alta risoluzione", sempre le stesse tabelle di quantizzazione: una per la luminanza e una per la cromaticanza. Empiricamente, abbiamo rilevato che utilizzando il metodo di *scaling* elencato in precedenza, con QF=85, si generano proprio le tabelle evidenziate nelle immagini scaricate da Facebook. Quindi, siamo arrivati alla conclusione che Facebook effettua la compressione JPEG con QF=85, in accordo al metodo di scaling visto in precedenza.

Precisiamo che indipendentemente dalla compressione dell'immagine di partenza, Facebook effettua sempre la compressione. Infatti, se carichiamo un'immagine su Facebook che adopera le tabelle di quantizzazione usate da quest'OSN, si nota che Facebook effettua comunque la compressione. In tutto questo documento indicheremo che Facebook adopera QF=85, per indicare che con il metodo di scaling visto su input 85, otteniamo le tabelle di quantizzazione presenti in tutte le immagini scaricate da Facebook.

FACEBOOK: QT LUMINANCE							
5	3	3	5	7	12	15	18
4	4	4	6	8	17	18	17
4	4	5	7	12	17	21	17
4	5	7	9	15	26	24	19
5	7	11	17	20	33	31	23
7	11	17	19	24	31	34	28
15	19	23	26	31	36	36	30
22	28	29	29	34	30	31	30

FACEBOOK: QT CHROMINANCE							
5	5	7	14	30	30	30	30
5	6	8	20	30	30	30	30
7	8	17	30	30	30	30	30
14	20	30	30	30	30	30	30
30	30	30	30	30	30	30	30
30	30	30	30	30	30	30	30
30	30	30	30	30	30	30	30
30	30	30	30	30	30	30	30

Figura 3. In figura, la tabella di quantizzazione della luminanza adoperata da Facebook.

Figura 4. In figura, la tabella di quantizzazione della cromaticanza adoperata da Facebook.

Badoo

Discorso analogo vale pure per Badoo. Infatti, abbiamo notato che le tabelle di quantizzazione risultavano sempre le stesse sull'immagine scaricata, al variare dell'immagine di partenza.

Empiricamente, abbiamo rilevato che utilizzando il metodo di scaling elencato in precedenza, con QF=91, si generano proprio le tabelle evidenziate nelle immagini scaricate da Badoo. Quindi, Badoo compie la compressione JPEG con QF=91, in accordo al metodo di scaling visto in precedenza. Inoltre, indipendentemente dalla compressione dell'immagine di partenza, Badoo compie sempre la compressione. In tutto questo documento indicheremo che Badoo adopera QF=91, per indicare che con il metodo di scaling visto su input 91, otteniamo le tabelle di quantizzazione presenti in tutte le immagini scaricate da Badoo.

BADIOO: QT LUMINANCE							
3	2	2	3	4	7	9	11
2	2	3	3	5	10	11	10
3	2	3	4	7	10	12	10
3	3	4	5	9	16	14	11
3	4	7	10	12	20	19	14
4	6	10	12	15	19	20	17
9	12	14	16	19	22	22	18
13	17	17	18	20	18	19	18

Figura 5. In figura, la tabella di quantizzazione della luminanza adoperata da Badoo.

BADIOO: QT CHROMINANCE							
3	3	4	8	18	18	18	18
3	4	5	12	18	18	18	18
4	5	10	18	18	18	18	18
8	12	18	18	18	18	18	18
18	18	18	18	18	18	18	18
18	18	18	18	18	18	18	18
18	18	18	18	18	18	18	18
18	18	18	18	18	18	18	18

Figura 6. In figura, la tabella di quantizzazione della crominanza adoperata da Badoo.

Google+

Per quanto riguarda il discorso legato alle tabelle di quantizzazione in Google+ è differente con quello di facebook e badoo. Questo perché esso permette di memorizzare le immagini nel formato PNG,GIF e BMP. Inoltre, per le immagini con una risoluzione minore di 2048 nel formato JPG essendo che le immagini rimangono intatte le tabelle sono uguali a quelle dell'immagine originale. Infine, per le immagini nel formato JPG con risoluzione maggiore di 2048 dai test effettuati si evince che Google+ le tabelle per tutte le immagini in cui è stato applicato l'algoritmo di resize sono:

Google+ : QT LUMINANCE							
5	3	15	15	13	13	13	13
4	16	15	13	13	13	13	13
16	15	13	13	13	13	13	13
13	13	13	13	13	15	13	13
13	13	13	13	13	21	19	13
13	13	13	13	13	19	22	16
13	13	13	14	19	24	24	18
13	16	17	17	22	18	20	18

Figura 7. In figura, la tabella di quantizzazione della luminanza adoperata da Google+.

Google+ : QT CHROMINANCE							
5	5	8	15	20	20	20	20
5	7	9	20	20	20	20	20
8	9	15	20	20	20	20	20
15	20	20	20	20	20	20	20
20	20	20	20	20	20	20	20
20	20	20	20	20	20	20	20
20	20	20	20	20	20	20	20
20	20	20	20	20	20	20	20

Figura 8. In figura, la tabella di quantizzazione della crominanza adoperata da Google+.

Dall'analisi delle tabelle di quantizzazione, abbiamo rilevato una "traccia" comune a tutte le immagini scaricate da Facebook. Naturalmente, lo stesso discorso vale per Badoo e Google Plus. Precisiamo che le tabelle di quantizzazione riportate non sono state controllate sulle miniature (anteprime) delle immagini, in quanto il discorso varia leggermente.

E' da precisare che sono stati compiuti altri esperimenti per vedere se Facebook, Badoo e Google+ applicavano qualche informazione nelle immagini o alteravano i pixel. Tutti gli esperimenti hanno dato esito negativo, infatti, abbiamo replicato il lavoro di Facebook su una particolare immagine. E' stata presa un'immagine originale, aperta e salvata con GIMP versione 2.6.11 su Windows 7. Quest'ultima è stata salvata con il QF=85 e le altre impostazioni di default. Inoltre, è stata caricata l'immagine originale su Facebook, ed è stata in seguito scaricata. Come risultato l'immagine di Facebook e quella compressa con gimp sono identiche pixel per pixel, in ogni canale di colore. Le dimensioni delle immagini sono 1280 x 1024 pixel, infatti si è cercato di non applicare il resizing dell'immagine.

6.1 Metodo basato sulle QT-JPEG

Il metodo che abbiamo sviluppato sfrutta proprio le tabelle di quantizzazione (QT). Una volta dimostrato sperimentalmente che Facebook utilizza QF=85 e Badoo QF=91, abbiamo realizzato un prototipo di applicazione adoperando *Matlab*. MATLAB (abbreviazione di *Matrix Laboratory*) è un ambiente per il calcolo numerico e l'analisi statistica che comprende anche l'omonimo linguaggio di programmazione creato dalla MathWorks [8]. Abbiamo utilizzato tale linguaggio perché ci consente di utilizzare un'ampia varietà di strumenti per l'analisi delle immagini JPEG, come l'*Image Processing Toolbox* e il *JPEG Toolbox* [9].

L'applicativo funziona nel seguente modo:

- Per ogni immagine da classificare si estraggono le tabelle di quantizzazione;

- Queste ultime sono confrontate con le tabelle di quantizzazione generate dai fattori di qualità (QFs) degli OSN (memorizzati in un file), secondo il metodo di scaling presentato.
- Se le tabelle di quantizzazione risultano uguali, sia per la luminanza sia per la crominanza, allora l'applicazione restituisce il nome dell'immagine seguito dal nome dell'OSN che adopera quelle determinate tabelle.

Il file dei fattori di qualità degli OSN è così strutturato:







OSN NAME		Quality Factor
	<i>FACEBOOK</i>	85
	<i>ANTEPRIMA_FACEBOOK(50x50_pixel)</i>	95
	<i>PUBBLICITA'_FACEBOOK</i>	100
	<i>BADOO</i>	91
	<i>ANTEPRIMA_BADOO(Xx48)</i>	97
	<i>ANTEPRIMA_BADOO(Xx96, Xx192, Xx300)</i>	94

Tabella 15 indicate i QF per ogni tipo d'immagine

La notazione "Xxyy" (dove yy può essere 48, 96, 192, 300) indica che il corrispondente lato della risoluzione dell'immagine è yy, mentre l'altro lato è variabile in proporzione a questo valore.

Quindi, data un'immagine si ottiene il nome dell'OSN che applica quelle tabelle di quantizzazione. Ad esempio, l'immagine *158485_920.jpg* proviene da *BADOO* (QF=91), mentre l'immagine *218919_103315643091724_100002398663355_30764_5605087_o.jpg* proviene da *FACEBOOK* (QF=85).

7. Gli OSN e il rumore PNU

Sia su Facebook sia su Badoo è stato applicato il filtro PNU, ossia il metodo che consente di estrarre da ogni immagine digitale il rumore PNU (disturbo caratteristico del sensore digitale), adoperando il cosiddetto “filtro PNU” [5]. Tale metodo consente di riconoscere quale fotocamera ha scattato un’immagine digitale distinguendo tra due fotocamere della stessa marca e modello, utilizzando le imperfezioni di fabbricazione del sensore digitale.

Facebook

E’ stata presa un’immagine ed è stata pubblicata su Facebook, in seguito su entrambe le immagini (immagini di partenza e quella pubblicata), è stato applicato il filtro PNU, per verificare se Facebook applica un filtro di denoising per estrarre il rumore caratteristico del sensore digitale (PNU). E’ da precisare che le due immagini, quella di partenza e quella scaricata, sono pressoché identiche, infatti, tali foto non sono uguali perché su quella di Facebook è stata applicata la compressione JPEG. A tal proposito precisiamo che Facebook esegue sempre la compressione JPEG, indipendentemente dall’immagine di partenza, e che il rumore PNU tende a diminuire se si esegue una forte compressione JPEG. Dagli esperimenti svolti, si è analizzato che Facebook non applica nessuna trasformazione esplicita volta a eliminare o attenuare il rumore PNU, tranne il fatto che applica la compressione JPEG su qualunque immagine di partenza, anche se è stata già compressa. L’applicazione ricorsiva di compressione JPEG e/o una forte compressione determinano l’eliminazione del PNU. Un altro effetto implicito che potrebbe attenuare il PNU è il resize dell’immagine.

Badoo

Gli stessi esperimenti sono stati effettuati anche su Badoo, ed hanno dato gli stessi esiti: Badoo non applica nessuna trasformazione esplicita volta a eliminare o attenuare il rumore PNU, tranne il fatto che applica la compressione JPEG su qualunque immagine di partenza, anche se è stata già compressa.

Google+

Dai test effettuati in Google+ si evince che con immagini avente risoluzione del lato più grande maggiore di 2048 non applica nessuna trasformazione esplicita volta a eliminare o attenuare il rumore PNU, tranne che applicare l’operazione di resize. Mentre, per le immagini digitali con lato più grande minore di 2048 lasciando le immagini intatte non elimina quindi il PNU.

Test Massivi

Un test massivo sul metodo Camera Identification applicato a immagini scaricate dagli OSN è reperibile nel paper Source Camera Identification in Online Social Networks di A. Castiglione, G. Cattaneo, M. Cembalo, U. Ferraro Petrillo. Dal seguente articolo sono state estratte informazioni riguardanti i test che sono stati realizzati su PNU con immagini scaricate dall’OSN facebook (consideriamo solo facebook questo perché esso ha lo stesso comportamento di badoo come si nota

dai diagrammi in allegato). Come risoluzione di facebook consideriamo la risoluzione standard ossia lato maggiore con 720 pixel:

Modello	Image Size	Threshold	Numero Immagini	Immagini Rifiutate
Canon EOS 400D	3888x2592	0,0062	178	165
Canon EOS 1000D	3888x2592	0,0076	176	175
Canon Powershot A 400 istanza A	2048x1536	0,0063	171	18
Canon Powershot A 400 istanza B	2048x1536	0,0062	159	2
Panasonic Lumix DMC-FZ20	2048x1536	0,0061	180	14
Panasonic Lumix DMC-FS5	3648x2736	0,0040	180	84
Kodak EasyShare CX 7530	2560x1920	0,0072	2	4
HP PhotoSmart E327	2560x1920	0,0069	178	68

Tabella 16 immagini testate su facebook

Come si evince dalla tabella le immagini che sono state maggiormente rifiutate sono le immagini digitali che nell'operazione di resize sono state fortemente compresse, infatti vediamo che per le immagini digitali in cui si è passato dal lato maggiore 3888 a 720 abbiamo un alto numero di immagini rifiutate. Mentre, per le immagini digitali di cui si è passato dal lato maggiore da 2048 a 720 pixel abbiamo un basso numero di immagini rifiutate. Questo significa che il rifiuto delle immagini è associato al resize, quindi consideriamo il servizio utilizzato da twitter photobucket il quale di default inserisce immagini con lato maggiore 2048 pixel.

Modello	Image Size	Threshold	Numero Immagini	Immagini Rifiutate
Canon EOS 400D	3888x2592	0,0057	178	36
Canon EOS 1000D	3888x2592	0,0090	176	105
Canon Powershot A 400 istanza A	2048x1536	0,0067	171	11
Canon Powershot A 400 istanza B	2048x1536	0,0073	159	2
Panasonic Lumix DMC-FZ20	2048x1536	0,0193	180	--
Panasonic Lumix DMC-FS5	3648x2736	0,0035	180	10
Kodak EasyShare CX 7530	2560x1920	0,0070	2	2
HP PhotoSmart E327	2560x1920	0,0080	178	2

Tabella 17 immagini caricate su Photobucket

Come vediamo aumentando la risoluzione dell'immagine che si può caricare il numero di immagini rifiutate è diminuito drasticamente.

Per validare come impatta la compressione JPEG sull'affidabilità dell'identificazione della fotocamera (Camera Identification) sono stati compiuti vari test. Il metodo d'identificazione è robusto alla compressione JPEG, sebbene quest'alterazione diminuisca il valore medio delle correlazioni tra il *residual noise* dell'immagine (rumore PNU estratto dall'immagine) e il modello di riferimento corretto della fotocamera. Tale riduzione è graduale, infatti, al diminuire del fattore della qualità la correlazione cala.

Il seguente grafico mostra tale andatura per le immagini della Canon EOS 400D seriale 7615-51577. In particolare, sono state scattate trenta immagini a risoluzione massima, memorizzate con il

fattore di qualità JPEG più alto. Tali immagini sono state ricomprese con il software IrfanView (reperibile all'indirizzo web: <http://www.irfanview.com/>), utilizzando vari fattori di qualità (90, 75 e 50). Tutti i rumori di queste immagini sono stati correlati con il camera reference pattern della Canon EOS 400D in esame.

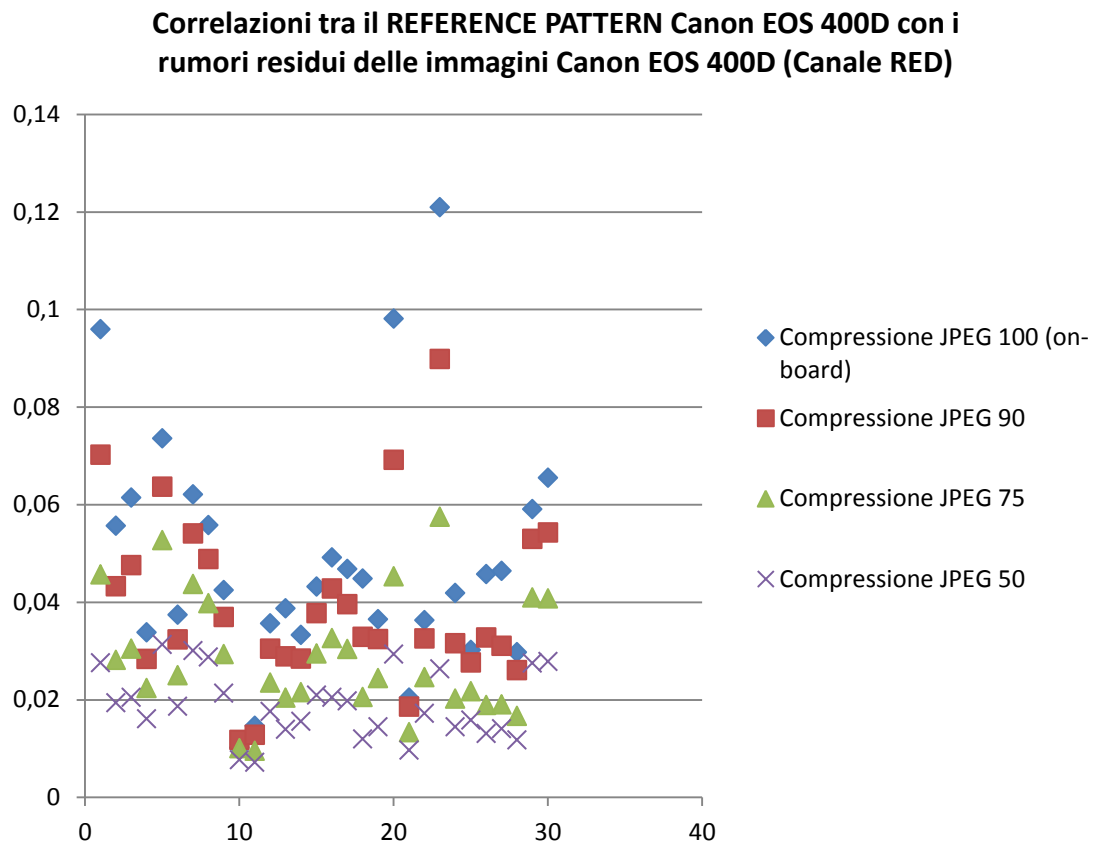


Figura 9. Il grafico mostra le correlazioni tra i rumori residui delle immagini scattate dalla Canon EOS 400D 7615-51577, ricomprese con il software IrfanView a varie qualità JPEG, con il reference pattern della stessa camera. Si considera solo il canale di colore rosso, poiché per gli altri canali vale lo stesso discorso.

Come si nota dal grafico, una riduzione del fattore di qualità determina il calo della correlazione tra il rumore residuo di un'immagine con il reference pattern della fotocamera che ha scattato tale immagine. Il motivo di tutto ciò è dato dalla compressione JPEG che diminuisce la qualità dell'immagine, visibile facilmente dall'introduzione di artefatti che prendono il nome di blocking (quadrettature). Di conseguenza, si ha anche una notevole perdita del rumore caratteristico della fotocamera, quindi, un sistema di machine learning per Camera Identification, non correttamente addestrato, potrebbe non considerare tali differenze nel rumore residuo e invalidare i risultati della classificazione.

8. Informazioni nascoste nelle immagini pubblicate su Facebook

Su Facebook e Badoo non si possono inserire informazioni nascoste nei pixel delle immagini, causa la compressione JPEG. Mentre adoperando Google Picasa è possibile nascondere informazioni nelle immagini.

Ad avvalorare la prima tesi, abbiamo compiuto un test inserendo alcune informazioni nascoste in un'immagine, adoperando il software jpegX. JpegX è un software di steganografia per inserire dati crittografati in file JPEG (utilizzato su sistemi Windows). In particolare i dati cifrati incorporati sono aggiunti al file in modo coerente e sono facilmente identificabili. JPEGX non inserisce le informazioni nei pixel, perché confrontando con la funzione `isequal()` di Matlab l'immagine originale (senza steganografia) e quella con le informazioni nascoste, queste foto risultano uguali pixel a pixel. Quindi, JPEGX inserisce queste informazioni nascoste in metadati che non possono essere estratti da Exiftool. Il software permette di inserire le informazioni all'interno di parti ridondanti del file utilizzando la tecnica del file camouflage (tecnica utilizzata per nascondere le informazioni in un file).

In seguito, abbiamo caricato l'immagine con l'informazione nascosta sull'OSN Facebook e poi l'abbiamo scaricata. Utilizzando poi jpegX, abbiamo cercato di estrarre l'informazione dall'immagine scaricata da Facebook, ma non è stato possibile. Quindi, le operazioni eseguite da Facebook sulle immagini non consentono di inserire informazioni all'interno della stessa, adoperando metodi classici di steganografia.

9. Informazioni nascoste nelle immagini pubblicate su Google Picasa

Il software Google Picasa [10] ci consente di organizzare, modificare e caricare le immagini sul Web in modo semplice e veloce sul repository dell'utente. Il nostro obiettivo è verificare se Picasa, nella fase di upload delle immagini, non altera il contenuto delle immagini in termini di pixel. Infatti, un primo test effettuato è stato quello di caricare un'immagine sul nostro Web album, utilizzando la configurazione "Dimensioni originali", ossia senza applicare il ridimensionamento dell'immagine di partenza. In seguito, abbiamo scaricato l'immagine dal nostro album Picasa e abbiamo esaminato l'immagine confrontando il contenuto di quella scaricata con il contenuto dell'immagine di partenza. Abbiamo notato che le due immagini sono uguali pixel a pixel, quindi è possibile nascondere informazioni all'interno delle immagini utilizzando un qualunque software di steganografia. Nel nostro caso abbiamo utilizzato JPEGX, il quale ci permette di inserire e cifrare un qualsiasi testo nel file dell'immagine. Abbiamo ripetuto il precedente test adoperando un'immagine di partenza che conteneva alcune informazioni nascoste con JPEGX. L'immagine scaricata da Picasa possiede l'intero messaggio inserito nella prima fase del test. Ciò dimostra che Picasa non altera il contenuto dell'immagine, è dà la possibilità di eseguire operazioni di steganografia sulle immagini. Inoltre, l'immagine di partenza e quella scaricata da Picasa hanno lo stesso valore hash, sia MD5 sia SHA-1, ciò dimostra che questo servizio Web, con l'opzione "Dimensioni originali", non altera il contenuto dell'immagine.

Siccome Picasa permette agli utenti differenti modalità di uploading, come caricare le immagini con tre diverse risoluzioni in pixel (1600, 1200 e 640), sono state esplorate tali caratteristiche. Ciò indica che Picasa utilizza un algoritmo di resize se l'utente sceglie una configurazione di upload

non concorde alla risoluzione dell'immagine originale. Abbiamo verificato se cambiando la configurazione di upload dell'immagine è possibile ottenere le informazioni nascoste. Dal test che abbiamo eseguito, seguendo lo stesso procedimento del secondo test, con la configurazione "dimensione 1600x1200", non siamo riusciti a recuperare il messaggio che abbiamo nascosto nell'immagine. Ciò indica che utilizzando una configurazione che modifica la risoluzione dell'immagine, i valori dei pixel vengono modificati, come ci aspettiamo.

Picasa, nel caso in cui l'immagine di partenza risulti di uguale o inferiore risoluzione rispetto a quella della configurazione desiderata, nella fase di upload, non applica l'algoritmo di resize, quindi è ancora possibile estrarre informazioni nascoste nel caso in cui vengano inserite.

10. Conclusioni

Bado, Facebooke Google+ non applicano particolari metodi di elaborazione delle immagini. Data un'immagine scaricata da Facebook o da Badoo o da Google+, e memorizzata in locale, siamo riusciti ad affermare se tale immagine è stata scaricata da uno di questi due OSN. Inoltre, data un'immagine scaricata da Badoo, è possibile risalire all'utente che ha pubblicato quell'immagine. Causa la compressione JPEG, gli OSN non permettono di fare steganografia classica sulle immagini ad eccezione di Google+ che in alcuni casi sì, mentre Google Picasa permette di fare steganografia sulle immagini.

La seguente tabella mostra la pubblicazione su vari OSN di un'immagine JPEG di risoluzione 3888x2592.

La prima colonna mostra dove l'immagine è stata pubblicata, e qual è l'immagine considerata; poi viene mostrata la risoluzione in pixel, il fattore di qualità JPEG dell'immagine pubblicata; se vengono inseriti i metadati dell'ICC-Profile; se è presente il campo Comment; infine, se i metadati sono alterati (creati, eliminati, modificati).

OSN\Caratteristiche	Risoluzione in pixel	Fattore di Qualità		ICC-Profile	Comment	Metadati originali
		LUM	CROM			
Facebook high resolution	2048x1365	85	85	Si	No	No
Facebook standard resolution	720x480	85	85	Si	No	No
Facebook Thumbnail 50x50 pixel	50x50	95	95	No	Si	No
Badoo	920x613	91	91	No	Si	No
Badoo Thumbnail 1	72x48	97	97	No	No	No
Badoo Thumbnail 2	144x96	94	94	No	No	No
Badoo Thumbnail 3	288x192	94	94	No	No	No
Badoo Thumbnail 4	450x300	94	94	No	Si	No
Picasa dimensione originale	3888x2592	uguale	uguale	No	No	Si
Picasa dimensione consigliate	1600x1067	81,45	88,78	No	No	Si
Picasa dimensione medie	1024x683	78,58	88,60	No	No	Si

Picasa dimensione piccole	640x427	81,04	88,78	No	No	Si
Google+	2048x1365	80,60	88,76	No	No	No
Twitter (Photobucket)	3888x2592	uguale	uguale	No	No	Si

Tabella 18 tabella riassuntiva

11. Riferimenti bibliografici

- [1] Ricerca OSN, Google Trends for Websites/ Alexa, fonte URL <http://www.vincos.it>.
- [2] Ricerca Facebook, fonte URL <http://lissimattia.com>.
- [3] Ricerca Badoo, fonte URL <http://www.vincos.it>.
- [4] Scientific Working Group on Imaging Technology (SWGIT) URL www.fbi.gov.
- [5] Aniello Castiglione, Giuseppe Cattaneo, Maurizio Cembalo, Umberto Ferraro Petrillo: Source Camera Identification in Real Practice: A Preliminary Experimentation. BWCCA 2010: 417-422
- [6] ExifTool <http://www.sno.phy.queensu.ca/~phil/exiftool/>.ExifTool
- [7] Standard JPEG
- [8] Matlab URL <http://www.mathworks.com/>.
- [9] Matlab JPEG Toolbox di Phil Sallee. Questo toolbox contiene routine Matlab per manipolare file JPEG, esempio coefficienti JPEG e le tabelle di quantizzazione, URL <http://www.philsallee.com/jpegtbx/index.html>.
- [10] Google Picasa <http://picasaweb.google.com/home?hl=it>.
- [11] <http://www.plusitalia.it/2011/07/24/20-milioni-di-utenti-su-google-plus-trend-in-calo/>