Università degli Studi di Salerno

Digital Watermarking

Corso di **Sicurezza su Reti 2**
a.a. 2008/09
Prof. A. De Santis

Daniele Giardino
0521000834

Sommario

- Introduzione
  - Di cosa parliamo
  - Motivazioni
  - Un po' di storia
  - Ambienti applicativi
  - Proprietà e tipologie
  - Schemi generali
- Image Watermarking
  - Tecniche del dominio spaziale
  - Tecniche del dominio delle frequenze
  - Possibili attacchi
- Audio Watermarking
  - Schemi di Watermarking
  - *Audio Watermarking Technique Could Locate Movie Pirates*
- Curiosità

Sommario

- **Introduzione**
  - Di cosa parliamo
  - Motivazioni
  - Un po' di storia
  - Ambienti applicativi
  - Proprietà e tipologie
  - Schemi generali
- Image Watermarking
  - Tecniche del dominio spaziale
  - Tecniche del dominio delle frequenze
  - Possibili attacchi
- Audio Watermarking
  - Schemi di Watermarking
  - *Audio Watermarking Technique Could Locate Movie Pirates*
- Curiosità

Che cos'è un "watermark"?

- Può avere significati diversi a seconda del contesto
  - Volendo tradurre in italiano possiamo usare *"filigrana"*
  - In ambito informatico si parla di *digital watermark*
- In generale questo termina indica *informazioni* aggiunte ad un documento che possono essere successivamente rilevate o estratte
- E il watermarking?
  - È il processo di inserimento di tali informazioni

A cosa serve?

- Nell'era dell'informatica è usato per:
  - *Copy-protection*
  - *Copyright-protection*
- Storicamente è usato per inviare informazioni sensibili nascoste
- Informazioni nascoste?
  - Crittografia?
  - Steganografia?

Crittografia ≠ Steganografia ≠ Watermarking

- Crittografia
  - Protezione del contenuto di un messaggio
  - Attacco: decifrare il significato del messaggio
- Steganografia
  - Nascondere la presenza di una comunicazione
  - Attacco: scoprire la presenza della comunicazione
- Watermarking
  - Aggiungere informazioni di protezione
  - Attacco: eliminare o modificare tali informazioni

Motivazioni

- Rendere visibile a tutti gli utenti il legittimo proprietario
- Dimostrare l'originalità di un documento non contraffatto
- Evitare la distribuzione di copie non autorizzate
- Marcare alcune caratteristiche specifiche del documento
- Segnare il percorso di vendita

Un po' di storia…

- Si parte dal 1282, Bologna…
  - *Dandy Roll:* si imprime uno stampo metallico sulla carta durante la produzione
    - usato dai produttori di carta per identificare i propri prodotti
- …passando per il 1848…
  - *Cylinder mould watermark:* tramite aree in rilievo su di un rullo viene creata un'immagine in scala di grigi
- …arrivando ai giorni nostri
  - *Digital Watermarking*

Scenari applicativi

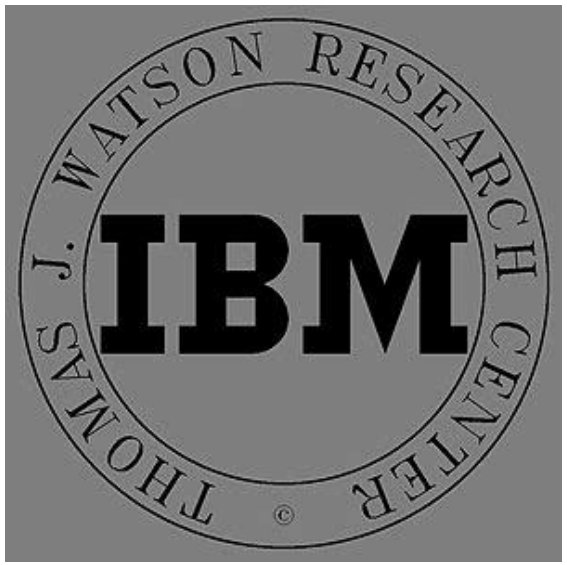
- Autenticazione ed integrità dei dati
- Copie non autorizzate
- Identificazione del distributore illegale
- Affermazione di paternità
- Protezione dei contenuti
- Aggiunta di informazioni

Tipologie di Watermarking

- Blind (Public) Watermarking
  - Non è necessario *l'originale* per individuare il watermark
  - Si tratta della tecnica più difficile da implementare, ma anche di quella che garantisce la migliore flessibilità del sistema
- Semi-blind Watermarking
  - È necessario l'accesso ad alcune informazioni
    - Di solito al decodificatore viene fornito il watermark e in output si ottiene un booleano che ne identifica la presenza o meno all'interno dell'immagine
- Non-blind (Private) Watermarking
  - È necessario *l'originale* per individuare il watermark
  - Si tratta della tecnica più semplice e più affidabile
    - Ma la necessità di possedere anche l'originale ne limita l'utilità

Tipologie di Watermarking

- Visibili



Watermark

Immagine con Watermark

- Invisibili

Tipologie di Watermarking

- Fragile
  - Altamente sensibile alle modifiche
  - Serve per identificare ogni cambiamento e dove questo è stato messo in atto *(manomissioni)*

- Robusto
  - Deve essere difficile da rimuovere
  - Deve resistere sia ad attacchi *voluti* che *non voluti*
  - Si definisce in termini di *livello di degradazione*
  - Serve per "trasportare" informazioni che devono accompagnare *sempre* il documento *(copyright?)*

Proprietà del Watermarking

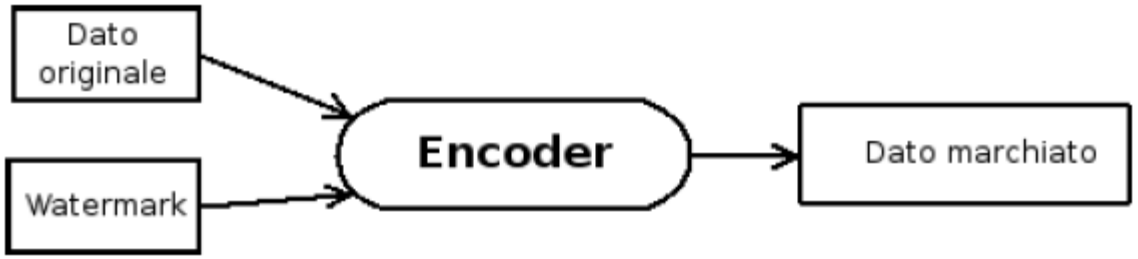
- *La sicurezza non deve essere basata sull'algoritmo (Kerkhoff)*

- Impercettibilità
  - Il documento *marked* deve avere la stessa utilità dell'originale

- Individuabilità
  - Se richiesto, il watermark deve essere individuabile *efficientemente*

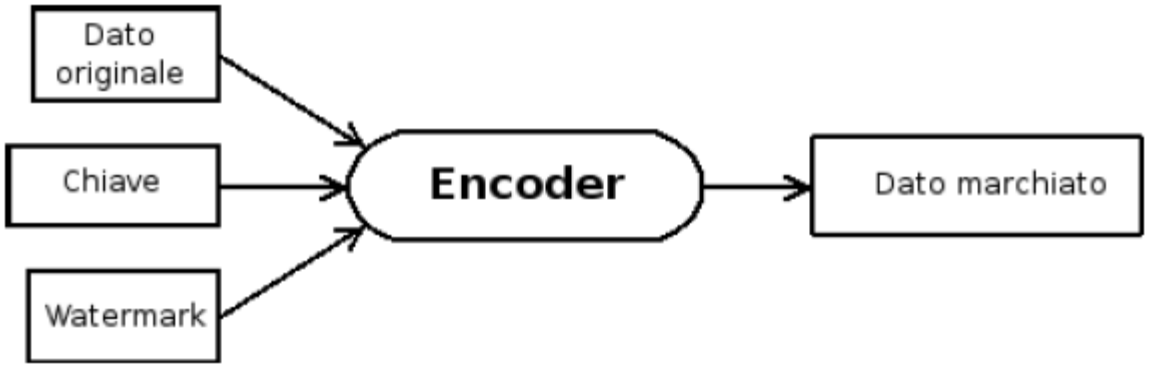
Schema generale di codifica

- $E(W , I) = I_m$
  - E: funzione di inserimento del watermark
  - W: il watermark
  - I: Oggetto su cui inserire il watermark
  - I_m: Oggetto *marchiato*



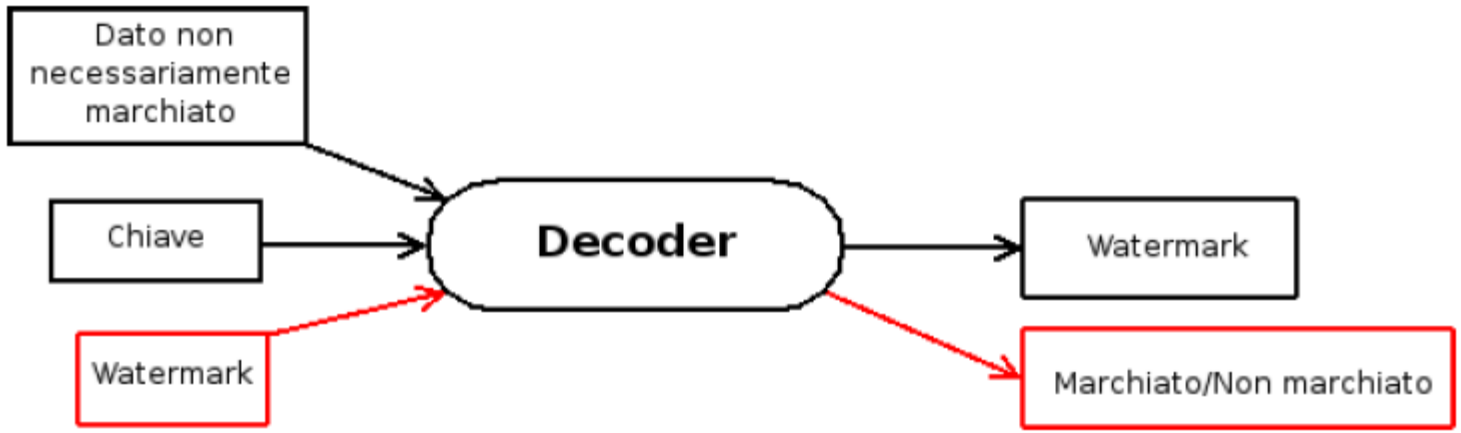
Schema generale di codifica con chiave

- $E(W, I, K) = I_m$
  - E: funzione di inserimento del watermark
  - W: il watermark
  - I: Oggetto su cui inserire il watermark
  - K: chiave
  - I_m: Oggetto *marchiato*



Schema generale di decodifica con chiave

- $D(I_m, K) = W$
  - D: funzione di verifica del watermark
  - I_m: Oggetto da controllare
  - K: chiave
  - W: il watermark
- $D(I_m, K, W) = \textit{true/false}$



Sommario

- Introduzione
  - Di cosa parliamo
  - Motivazioni
  - Un po' di storia
  - Ambienti applicativi
  - Proprietà e tipologie
  - Schemi generali
- **Image Watermarking**
  - Tecniche del dominio spaziale
  - Tecniche del dominio delle frequenze
  - Possibili attacchi
- Audio Watermarking
  - Schemi di Watermarking
  - *Audio Watermarking Technique Could Locate Movie Pirates*
- Curiosità

Image Watermarking

- La gran parte degli schemi di watermarking operano essenzialmente in due modalità:
  - Dominio spaziale
  - Dominio delle trasformate

Originale
(SHA1) 7cfd6a1838d56b47f8f398a16eeca42828e8f172

Marcata
(SHA1) 3cf51bdacd5b2e2f3d40a40da41958757018470d

Image Watermarking

- Modello generale
  - Una immagine può essere vista come una matrice
    - Gli elementi della matrice sono i *pixel*
    - Un pixel è associato ad una stringa binaria di 8, 16, 24, 32 bit
      - Il numero di bit utilizzati per descrivere il pixel è la *risoluzione in ampiezza*
      - Il numero di righe della matrice è la *risoluzione orizzontale*
      - Il numero di colonne della matrice è la *risoluzione verticale*

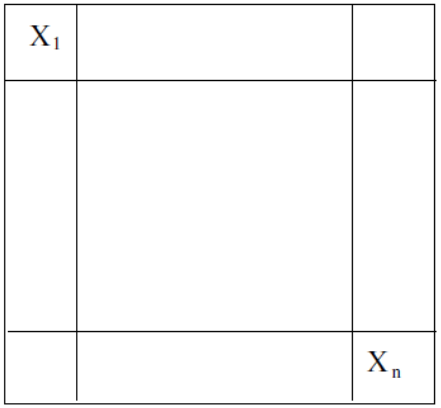


Image Watermarking

- Modello generale
  - Ogni stringa punta ad una *matrice di colori*
  - Per le immagini a toni di grigio
    - Con una stringa di 8 bit possiamo rappresentare 256 colori diversi

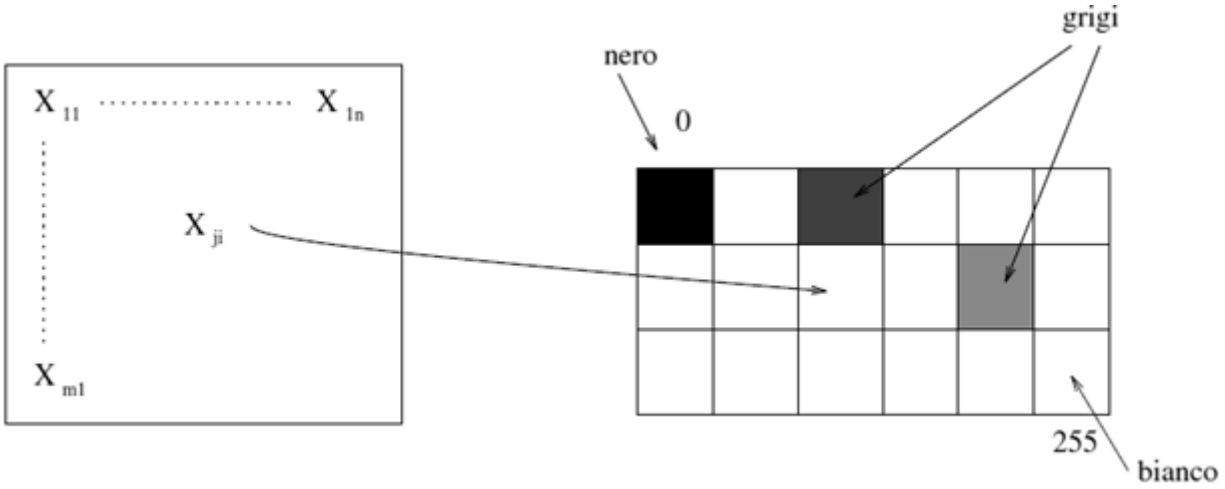


Image Watermarking

- Modello generale
  - Il colore è espresso mediante la combinazione di tre componenti:
    - Rosso, Verde e Blu
  - Ogni componente varia in modo indipendente
  - Le tre componenti insieme formano uno spazio tridimensionale noto come **RGB** (Red Green Blue)
    - Quindi un'immagine è costituita da tre matrici R, G e B una per ogni colore base

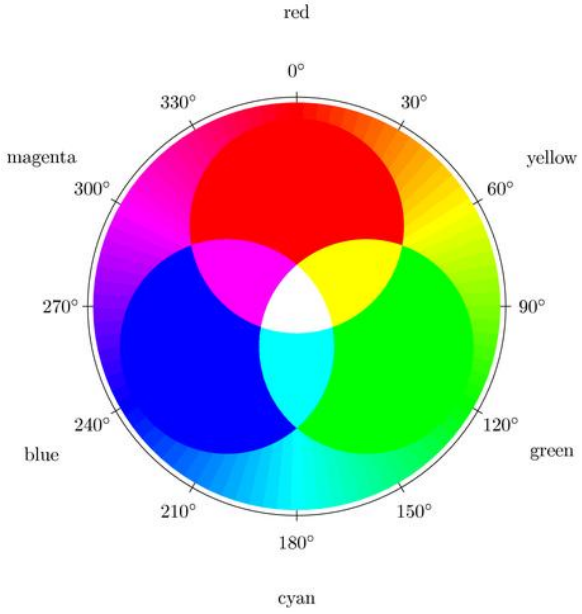


Image Watermarking

- Modello generale
  - Una rappresentazione alternativa è la **YUV**
    - Utilizza altre caratteristiche del colore: luminosità, tonalità e saturazione
    - La luminosità descrive l'intensità della luce (rivelando se il colore è bianco, grigio o nero)
    - La tonalità descrive la presenza del colore (rosso, verde, giallo, ecc.)
    - La saturazione descrive quanto è vivo il colore (molto forte, pastello, quasi bianco, ecc.)

Image Watermarking

- Modello generale
  - Una rappresentazione alternativa è la **YUV**
    - La luminosità di un pixel è ottenuta sommando i tre colori, del sistema RGB, nelle proporzioni approssimate di 30% per il rosso, 60% per il verde e 10% per il blu
      - Y = 0.3 R + 0.6 G + 0.1 B
    - La crominanza invece è definita come la differenza tra un colore e la luminosità
      - V = R – Y
      - U = B - Y

Image Watermarking

- Modello generale
  - I sistemi di spazi del colore, nei quali una componente è la luminosità e le altre due dipendono dalla tonalità e dalla saturazione, sono dette *rappresentazioni luminosità – crominanza*
  - La rappresentazione luminosità - crominanza è più utile rispetto a RGB per ottenere buone compressioni di immagini
    - È possibile scartare più informazione nelle componenti cromatiche, a cui l'occhio umano è meno sensibile, che in quella di luminosità

Dominio Spaziale

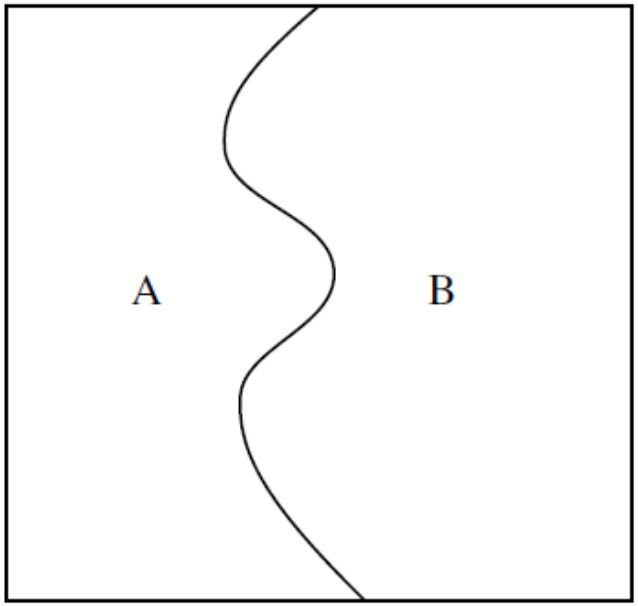
- Si agisce direttamente sui valori dei *pixel* che costituiscono l'immagine
- Le operazioni di gestione ed inserimento del watermark sono semplici ed immediate
- È stato però dimostrato essere non accettabile dal punto di vista della robustezza
- L'esempio più semplice è conosciuto con il nome di *LSB (Less Significative Bits)*
  - vengono modificati i bit meno significativi di ogni pixel dell'immagine

Dominio delle trasformate (frequenze)

- Questa metodo consiste nell'applicare una specifica *trasformata* all'immagine
- Successivamente vengono eseguite le operazioni necessarie al'inserimento/rilevazione del marchio
- Infine si inverte la trasformata per ottenere l'immagine marchiata o estrarre il marchio
- Le trasformate più utilizzate sono *DCT (Discrete Cosine Transform)*, *DFT (Discrete Fourier Transform)* e *DWT (Discrete Wavelet Transform)*

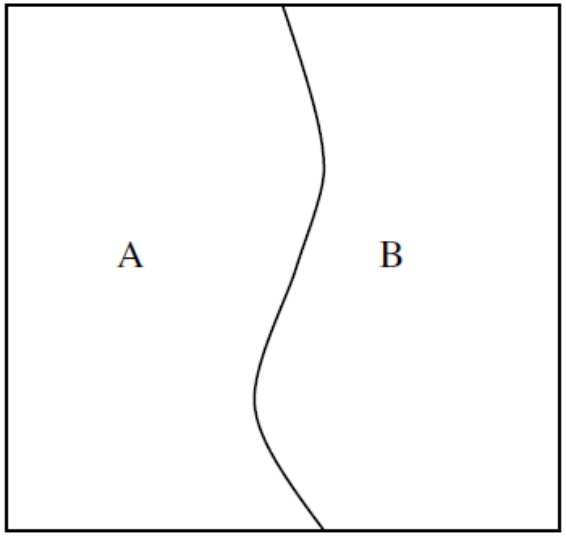
Tecniche del dominio spaziale

- Schema generale – Fase di inserimento
  - Si dividono i pixel dell'immagine in due partizioni (A, B)
  - Si suppone che i valori dell'intensità dei pixel siano uniformemente distribuiti e che $|A| \approx |B|$
  - Il partizionamento viene effettuato in accordo ad una *chiave segreta*



Tecniche del dominio spaziale

- Schema generale – Fase di inserimento
  - Il valore del watermark (W) viene aggiunto al valore dell'intensità di tutti i pixel di A e lo si sottrae a quelli di B
  - La caratteristica principale è che essendo W un intero piccolo abbastanza, la sua addizione/sottrazione non provoca una visibile degradazione dell'immagine



Tecniche del dominio spaziale

- Schema generale – Fase di estrazione
  - Si ricreano le partizioni in accordo alla *chiave segreta*
  - Si calcolano le intensità medie dei pixel di A e di B
  - Se nell'immagine è presente il watermark, la differenza sarà approssimativamente di 2W, vicina allo 0 altrimenti
    - Tale schema può essere migliorato dividendo l'immagine in blocchi e applicando ad ogni blocco lo schema visto

Tecniche del dominio spaziale

- Schema di Langelaar (2000)
  - Viene inserita una stringa di bit nel dominio spaziale dell'immagine
    - Per fare ciò viene manipolata la luminanza dei pixel in blocchi 8 x 8
      - Viene creato un pattern pseudocasuale delle stesse dimensioni del blocco

$$pat(x,y) \in \{0,1\}\ dove\ 0 \leq x,y < 8$$

Tecniche del dominio spaziale

- Schema di Langelaar
  - Per inserire il bit *s* del watermark nel blocco *B*
    - Si suddivide in due sottoinsiemi

$$B_0 = \{l(x + x_0, y + y_0)\ dove\ pat(x, y) = 0\}$$
$$B_1 = \{l(x + x_0, y + y_0)\ dove\ pat(x, y) = 1\}$$

    - Il valore della luminanza media è calcolata per entrambi
      - La differenza tra i due valori rappresenta la presenza del bit del watermark

$$l_0 - l_1 \geq +\alpha \quad se\ s = 1$$

$$l_0 - l_1 < -\alpha \quad se\ s = 0$$

Tecniche del dominio spaziale

- Schema di Langelaar
  - Per rendere l'algoritmo più robusto rispetto alla compressione JPEG
    - trasformazione DCT
    - quantizzazione dei coefficienti con un fattore di qualità Q
    - trasformazione DCT inversa
  - Il risultato è un blocco lievemente modificato
    - La luminanza media viene ricalcolata

$$\left.\begin{array}{l} l_0 - l_1 > +\alpha \\ \hat{l}_0 - \hat{l}_1 > +\alpha \end{array}\right\} \quad se\ s = 1$$

$$\left.\begin{array}{l} l_0 - l_1 < -\alpha \\ \hat{l}_0 - \hat{l}_1 < -\alpha \end{array}\right\} \quad se\ s = 0$$

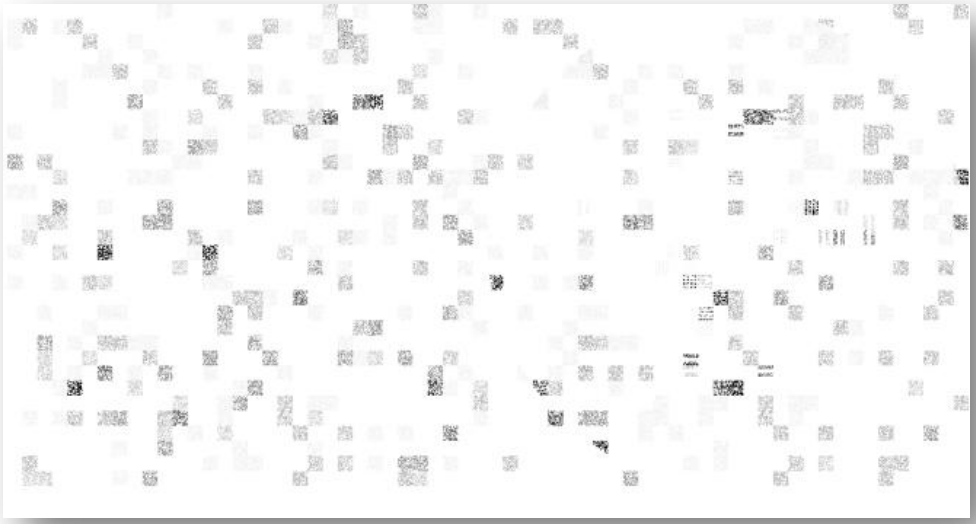
Tecniche del dominio spaziale

- Schema di Langelaar
  - L' algoritmo di estrazione non richiede l'immagine originale
    - Calcolo della luminanza media usando il pattern

$$s_k'' = \begin{cases} 0 & se\ l_0'' - l_1'' < 0 \\ 1 & se\ l_0'' - l_1'' > 0 \end{cases}$$

Tecniche del dominio spaziale

- Schema di Langelaar
  - La figura rappresenta la differenza tra l'immagine originale e quella con il watermark. Tale immagine è invertita e scalata in modo tale che il bianco indichi "nessuna differenza" ed il nero "differenza massima"
  -



Tecniche del dominio delle frequenze

- Discrete Cosine Trasform (DCT)
  - Formalmente, la DCT è una funzione lineare, invertibile

$$F : R^N \rightarrow R^N$$

  - Gli *N* numeri reali $x_0, x_1, \ldots, x_{N-1}$ sono trasformati nei numeri reali in accordo ad una delle seguenti formule:

$$X_k = \sum_{n=0}^{N-1} x_n \cos\left[\frac{\pi}{N}\left(n+\frac{1}{2}\right)k\right] \qquad k = 0, \ldots, N-1.$$

$$X_k = \frac{1}{2}x_0 + \sum_{n=1}^{N-1} x_n \cos\left[\frac{\pi}{N}n\left(k+\frac{1}{2}\right)\right] \qquad k = 0, \ldots, N-1.$$

  - I coefficienti DCT costituiscono una nuova rappresentazione dell'immagine rispetto ad una base *più comoda*

Tecniche del dominio delle frequenze

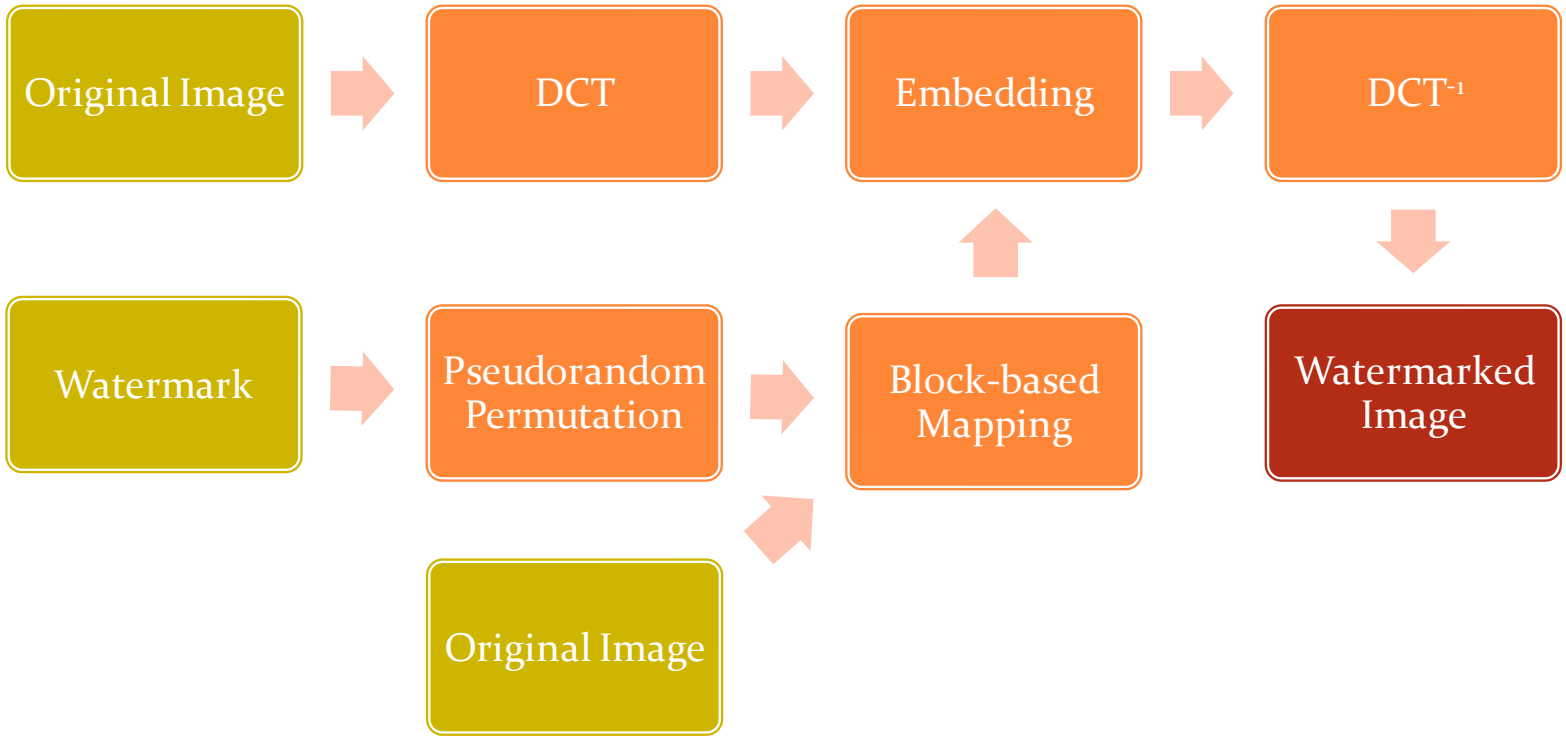
- Schema con Discrete Cosine Trasform (DCT)
  - Data un'immagine di partenza I viene divisa e considerata come una matrice di blocchi 8x8
  - Si scelgono in modo *pseudocasuale* un numero di blocchi pari alla lunghezza del watermark da inserire
  - Si effettua la *trasformazione DCT* e si considerano nella rappresentazione dello spazio delle frequenze i blocchi corrispondenti a quelli scelti
  - Si effettua l'operazione di quantizzazione, ottenendo una matrice di valori approssimati con coefficienti uniformati

Tecniche del dominio delle frequenze

- Schema con Discrete Cosine Trasform (DCT)
  - L'inserimento del watermark viene effettuato apportando modifiche ai valori dei coefficienti DCT ovviamente dei blocchi scelti precedentemente
    - Sia W il watermark
    - Consideriamo l' i-esimo bit di tale stringa che deve essere inserito nell'i-esimo blocco, del quale si hanno a disposizione i coefficienti DCT
    - Se tale bit è uno zero allora si modifica il primo coefficiente e la modifica viene fatta in modo tale che rappresenti uno zero
    - Alternativamente se tale bit è un uno

Tecniche del dominio delle frequenze

- Schema con Discrete Cosine Trasform (DCT)

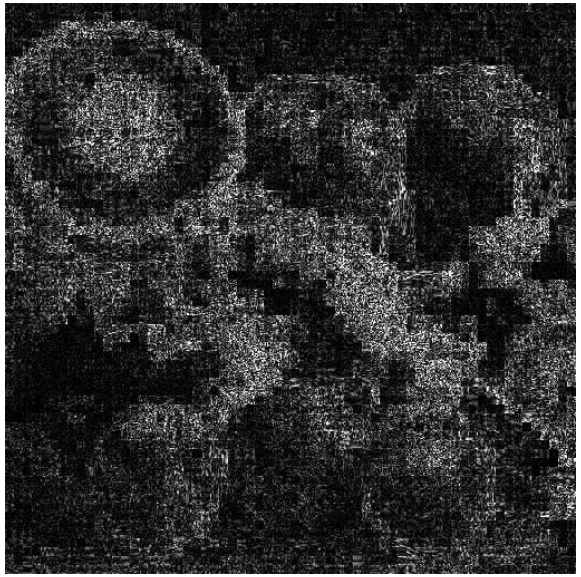


Tecniche del dominio delle frequenze

- Esempio di Watermark con DCT

Originale

Immagine marcata



Watermark

Tecniche del dominio delle frequenze

- Schema con DCT – Precisazioni
  - I coefficienti vengono conformati a tre relazioni che rappresentano rispettivamente 0, 1, *invalido*
    - Non sempre l'inserimento di un bit in un blocco va a buon esito
      - La trasformazione dei coefficienti può risultare pesante e provocare un degrado della qualità dell'immagine
      - In tali casi il coefficiente viene marcato *invalido* per indicare che nel rispettivo blocco non ci sono informazioni aggiunte

Tecniche del dominio delle frequenze

- Schema con DCT – Precisazioni
  - Per la verifica del marchio viene effettuata l'operazione inversa
    - Ottenuti i coefficienti DCT, vengono confrontati con quelli dell'immagine originale
    - Si identificano così i coefficienti modificati e si risale alla relazione che li ha conformati
    - A questo punto è facile ottenere il watermark

Tecniche del dominio delle frequenze

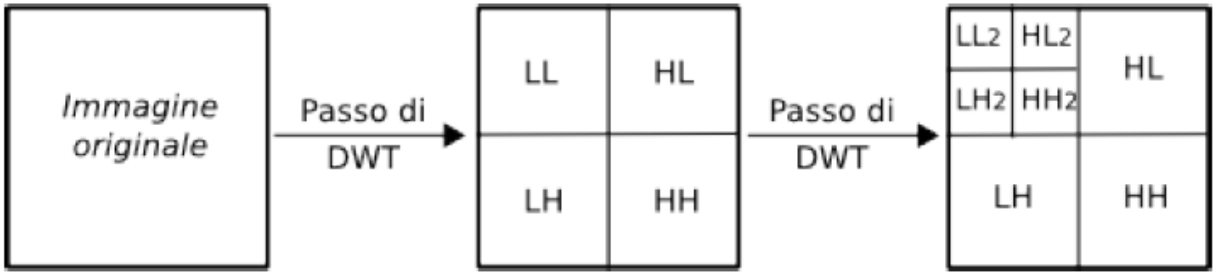
- Schema di Cox et al. (1997)
  - Non è altro che una variazione dell'algoritmo discusso in precedenza
  - L'osservazione su cui si basa questo schema è di scegliere i blocchi nei quali inserire l'informazione in modo oculato
  - Una volta applicata la DCT, la matrice risultante ha una struttura ben precisa
    - Ci sono aree in cui sono presenti coefficienti nulli (che non portano informazioni) e aree in cui sono presenti i coefficienti più significativi (quelli che hanno i valori più alti)
    - Occorre quindi, andare a marcare i valori più significativi
      - Se si decidesse di marcare i valori nulli, si otterrebbe una perdita dell'informazione
        - Tali valori vengono tagliati via dagli *algoritmi di compressione*

Tecniche del dominio delle frequenze

- Discrete Wavelet Transform (DWT)
  - È una trasformata *invertibile*
    - è possibile applicare il passo inverso e riottenere l'informazione iniziale
  - Scompone le componenti di un segnale ad alta frequenza da quelle a bassa frequenza
  - L'immagine considerata viene in quattro quadranti: LL, HL, LH e HH
    - Di dimensioni pari alla metà dell'originale (DWT di livello 1)

Tecniche del dominio delle frequenze

- Discrete Wavelet Transform (DWT)
  - La sottobanda LL (in alto a sinistra) contiene le basse frequenze *(coefficienti di approssimazione)*
    - può essere divisa ulteriormente, per ottenere un ulteriore livello di scomposizione
  - Le restanti tre sottobande contengono le alte frequenze
    - Sono quelle meno percepibili dall'occhio umano, nel caso di immagini
    - Sono le tre sottobande sulle quali agiscono gli algoritmi di watermarking



Tecniche del dominio delle frequenze

- Discrete Wavelet Transform (DWT) - Esempio

Originale

Prima applicazione della DWT

Seconda applicazione della DWT

Tecniche del dominio delle frequenze

- Schema di Dugad (1998)
  - È un algoritmo blind, che opera nel dominio delle *wavelet* e utilizza una tecnica additiva di embedding
    - ovvero i coefficienti modificati nella fase di inserimento del marchio sono un incremento, secondo determinati parametri, dei coefficienti originali
  - La trasformata wavelet utilizzata è di terzo livello
  - Nessun watermark viene inserito nella sottobanda LL

Tecniche del dominio delle frequenze

- Schema di Dugad
  - Effettua un mascheramento visivo implicito del watermark
    - Vengono selezionati per l'embedding solo i coefficienti wavelet con una frequenza abbastanza elevata
      - I coefficienti di elevata frequenza corrispondono alle regioni di texture e bordi di un'immagine
        - Questo ha l'effetto di rendere difficile, da parte dell'occhio umano, l'individuazione delle degradazioni dell'immagine dovute all'inserimento del watermark
      - Inoltre questi coefficienti sono altamente significativi dal punto di vista di percettibilità
        - Rendono molto difficile la rimozione del watermark senza degradare severamente l'immagine marchiata

Tecniche del dominio delle frequenze

- Schema di Dugad
  - Vantaggi
    - E' un algoritmo di watermarking blind
    - Incorpora un mascheramento visivo implicito
    - Utilizza un watermark di grandezza pari all'immagine
      - consente di mantenere una certa indipendenza dall'ordine dei coefficienti significativi nel processo di rilevazione

Tecniche del dominio delle frequenze

- Schema di Dugad
  - Svantaggi
    - Effettua l'inserimento utilizzando una tecnica additiva
      - i rilevatori per schemi che adottano tecniche additive devono correlare i coefficienti dell'immagine possibilmente marchiata con un watermark noto, in maniera da determinare se l'immagine è stata o meno marchiata
    - Effettua esclusivamente una rilevazione del watermark e non un'estrazione

Tecniche del dominio delle frequenze

- Schema di Inoue (1999)
  - A differenza dello schema presentato da Dugad effettua l'inserimento del watermark esclusivamente nelle sottobande del terzo livello della trasformata
  - Vengono modificati i coefficienti wavelet ad elevata frequenza
  - Ma in questo caso i coefficienti vengono *sostituiti* e non incrementati

Tecniche del dominio delle frequenze

- Schema di Inoue
  - Vantaggi
    - Utilizza un processo di quantizzazione per inserire il watermark
      - Permette di non essere influenzati da interferenze dell'immagine originale
    - Effettua l'estrazione del watermark, rendendone quindi possibile la visualizzazione
    - Incorpora anch'esso un mascheramento visivo implicito
  - Svantaggi
    - È uno schema semi-blind
    - Necessita dell'insieme di posizioni in cui è stato inserito il watermark

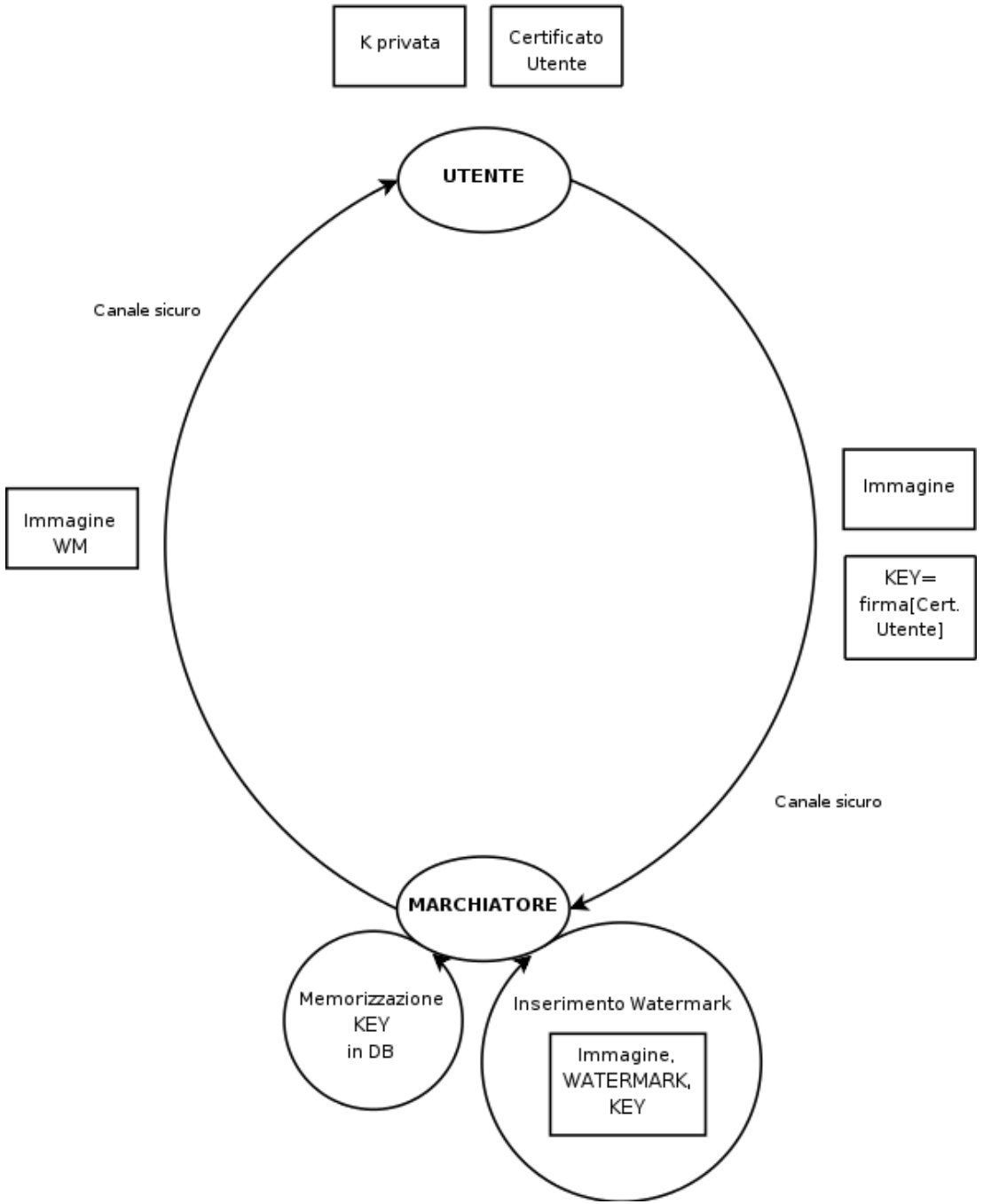
Tecniche del dominio delle frequenze

- Schema di Fraser (2003)
  - Vengono condivisi i vantaggi di Dugad e Inoue, rimuovendo la maggior parte degli svantaggi
    - Si utilizza l'idea di Dugad di un watermark di dimensioni pari all'immagine
      - In congiunzione con una versione adattata delle tecniche di inserimento e rilevazione basate su quantizzazione dello schema di Inoue
      - Lo schema risultante è quindi una sorta di versione di Dugad basata su quantizzazione

Schema di Watermarking con certificati digitali

- Processo di inserimento
  - L'utente invia in input al marchiatore l'immagine, ed una chiave KEY
  - Questa viene prodotta semplicemente firmando il proprio certificato a chiave pubblica
  - Ricevuta la coppia di input, il marchiatore verifica tramite il certificato dell'utente, la validità di KEY
  - In caso di verifica andata a buon fine memorizza KEY in un database
  - Il marchiatore si occupa di effettuare l'inserimento di un watermark nell'immagine
  - Allo scopo di aumentare la robustezza del sistema, il watermark è prodotto a partire da KEY
  - A questo punto l'immagine è stata marchiata ed è pronta per essere inviata nuovamente all'utente

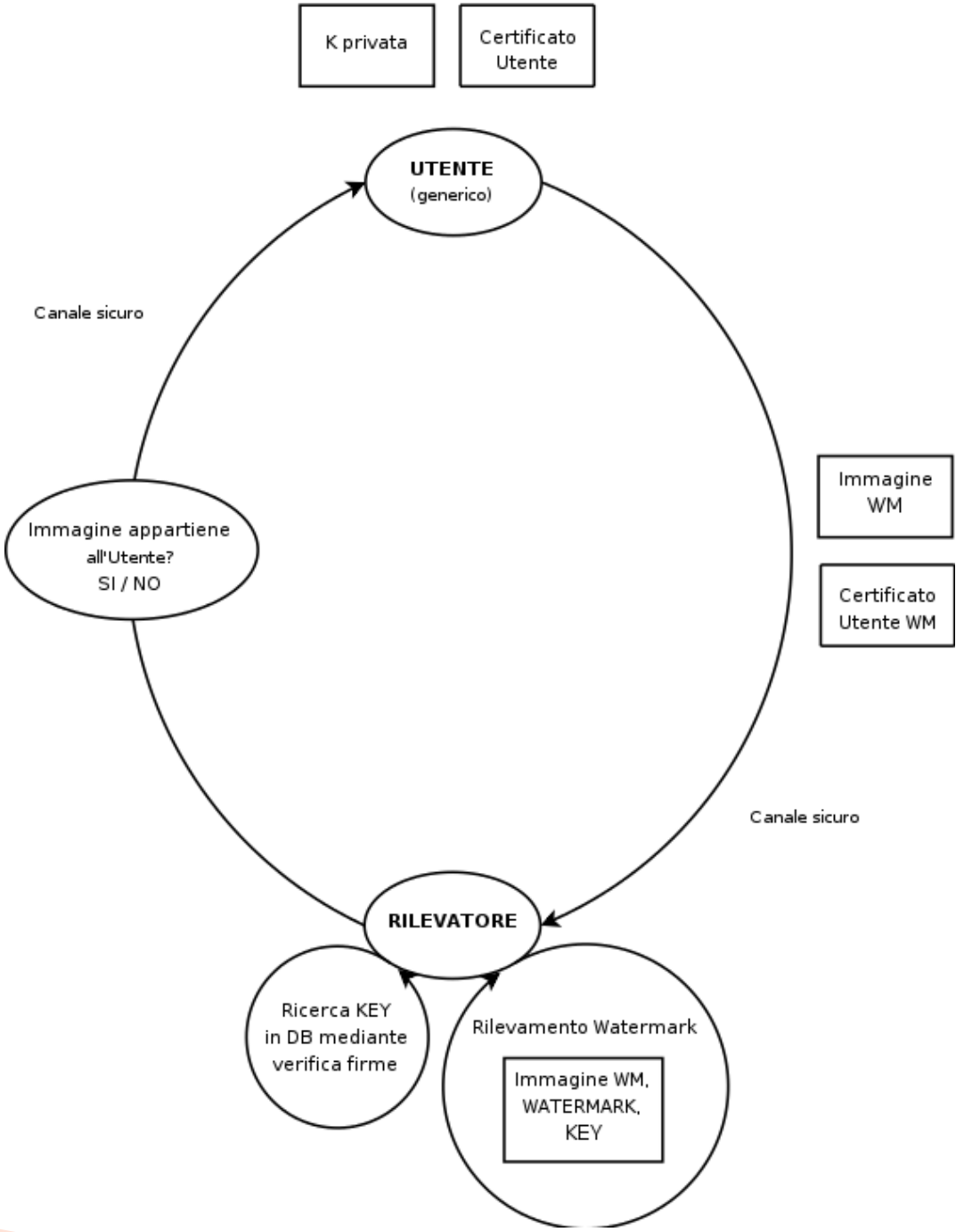
Schema di Watermarking con certificati digitali



Schema di Watermarking con certificati digitali

- Processo di verifica
  - Avendo l'immagine ed il certificato del presunto utente marchiatore il *validatore* procede con il reperimento della KEY dal database
  - In particolare, attraverso una verifica della firma del certificato è possibile recuperare la KEY relativa allo specifico certificato
  - In caso di ritrovamento della KEY, il *verificatore* utilizza tale chiave per il processo di estrazione del watermark dall'immagine ottenuta precedentemente

Schema di Watermarking con certificati digitali



Possibili attacchi

- Compressioni lossy
  - Molti schemi di compressione come JPEG, eliminano, allo scopo di comprimere l'immagine, le informazioni meno visibili all'occhio umano
    - Queste stesse informazioni vengono spesso utilizzate dagli schemi di watermarking per inserire il watermark
    - Quindi la perdita di queste informazioni potrebbe comportare il danneggiamento (o la perdita) del watermark
    - Gli schemi del dominio delle trasformate sono abbastanza robusti contro questa specie di attacchi

JPEG (Joint Photographic Experts Group)

- È il primo standard internazionale di compressione per immagini sia a livelli di grigio che a colori
  - È un formato gratuito e open-source
- Specifica come una immagine può essere trasformata in uno stream di byte
  - Non come questo può essere incapsulato in supporti di memorizzazione
  - Un ulteriore standard chiamato **JFIF** (JPEG File Interchange Format) specifica come produrre un file appropriato per la memorizzazione su computer di uno stream JPEG
    - creato da *Independent JPEG Group*

JPEG (Joint Photographic Experts Group)

- Lo standard JPEG definisce due metodi di compressione di base
  - Uno basato sull'uso della trasformata discreta in coseno (*DCT*) con compressione di tipo "*lossy*"
    - Con perdita di informazione
  - L'altro sull'uso di un metodo predittivo con compressione di tipo "*lossless*"
    - Senza perdita di informazione
  - Nuovi metodi lossy basati sulle *DWT* garantiscono migliori risultati in alcuni casi
    - Il comitato JPEG ha creato un nuovo standard basato su wavelet, **JPEG2000**, con la prospettiva di sostituire nel tempo lo standard JPEG

JPEG (Joint Photographic Experts Group)

- Passi dell'algoritmo
  - Partizionamento
    - L'immagine originale viene partizionata in blocchi B di dimensione 8 x 8
  - Rappresentazione con DCT
    - Ad ogni blocco viene applicata la trasformata DCT
  - Quantizzazione
    - I valori dei coefficienti ottenuti con la DCT vengono quantizzati
      - Il range dei valori viene suddiviso in sottointervalli
        - Tutti i valori che cadono all'intero del sottointervallo vengono sostituiti da un solo valore

JPEG (Joint Photographic Experts Group)

JPEG qualità 100% - 87,7 Kb

JPEG qualità 90% - 30,2 Kb

JPEG qualità 50% - 6,7 Kb

JPEG qualità 10% - 3,2 Kb

Possibili attacchi

- Distorsioni geometriche
  - La rotazione, la traslazione, il ridimensionamento e il ritaglio dell'immagine
- Comuni operazioni di Signal Processing
  - Conversioni A/D e D/A, resampling, requantization, riduzione di colori, aggiunta di un offset costante al valore dei pixel, aggiunta di rumore, etc.
- Stampa e scannerizzazione
  - Nel passaggio da digitale a cartaceo e viceversa è possibile che vengano perse le informazioni contenenti il marchio

Altri attacchi intenzionali

- Rewatermarking
  - Il watermarking di un'immagine marchiata. In questo caso lo scopo è di sovrascrivere il marchio o di ingannare il sistema di watermarking
- Collusione
  - I possessori autorizzati di una serie di copie della stessa immagine marchiata differentemente non dovrebbero essere in grado di generare l'immagine originale non marchiata.

Altri attacchi intenzionali

- Forgery
  - I possessori autorizzati di un'immagine non dovrebbero essere in grado di generare una copia marchiata con informazioni riguardanti una terza parte
- IBM Attack
  - Non dovrebbe essere possibile produrre un falso originale che sia uguale all'originale e inoltre permetta l'estrazione di un watermark come affermato dal possessore della falsa copia

Sommario

- Introduzione
  - Di cosa parliamo
  - Motivazioni
  - Un po' di storia
  - Ambienti applicativi
  - Proprietà e tipologie
  - Schemi generali
- Image Watermarking
  - Tecniche del dominio spaziale
  - Tecniche del dominio delle frequenze
  - Possibili attacchi
- **Audio Watermarking**
  - Schemi di Watermarking
  - *Audio Watermarking Technique Could Locate Movie Pirates*
- Curiosità

Audio Watermarking

- Effetto Mascheramento
  - La percezione audio umana non può determinare piccoli cambiamenti in certe componenti di frequenza del segnale audio
  - Un debole ma percettibile segnale audio diviene non percettibile in presenza di un forte segnale audio di frequenza leggermente superiore
    - Tale fenomeno si verifica quando nel segnale che viene inviato all'orecchio sono presenti più segnali
      - Ad esempio, se abbiamo un segnale ad una data frequenza ed un altro segnale, di livello più basso e ad una frequenza leggermente inferiore, quest'ultimo viene "mascherato" dal primo e l'orecchio non lo percepisce

An Affine Resistant Watermarking Scheme for Audio Signals

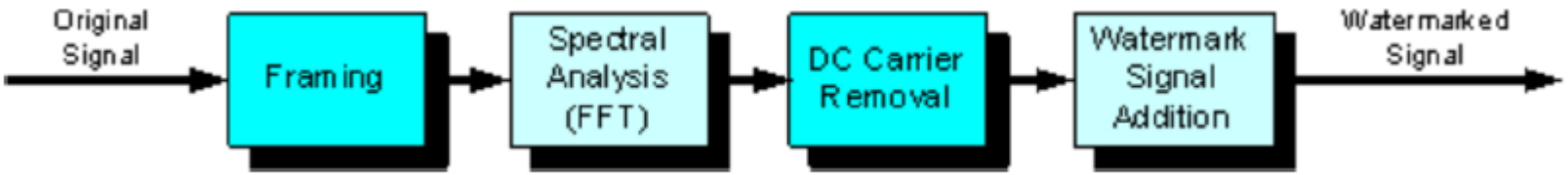
- L'idea (P.A. 60/907,778)
  - Codificare un watermark all'interno di una portante di un segnale audio, shiftando nel tempo blocchi del segnale portante in tutta la banda o in sottobande
  - Ci sono differenti tecniche, anche per quanto riguarda il watermarking analogico, che utilizzano lo shifting del segnale portante
  - Questo schema sfrutta la somiglianza dei dati audio nei due canali di un segnale stereo
    - È resistente ai più comuni attacchi: requantization, transcoding, manipolazioni dell'onda e dello spettro

An Affine Resistant Watermarking Scheme for Audio Signals

- Utilizzo
  - Non richiede il segnale sorgente per l'estrazione del watermark (blind)
  - Durante la fase di encoding viene creata una chiave privata di decodifica
    - L'estrazione del watermark non è possibile senza tale chiave
    - La chiave è univoca per ogni segnale sorgente e dipende solo dalla dimensione del watermark da inserire (non dal suo contenuto)

DC Watermarking Scheme

- L'idea
  - Questo schema nasconde il watermark nelle componenti a bassa frequenza di un segnale audio
    - Tali componenti sono al di sotto della soglia di percezione dell'orecchio umano
- Inserimento del watermark



DC Watermarking Scheme

- Framing
  - Il file audio è partizionato in frame della durata di 90ms
    - La dimensione del frame è scelta in modo tale da non introdurre distorsioni udibili nel file
    - Con un frame di questa dimensione possiamo inserire 1/0.09 = 11.1 bit/s di dati
- Spectral Analysis
  - Viene applicata la Fast Fourier Transform (FTT)
    - Permette di calcolare le componenti a bassa frequenza di ogni frame

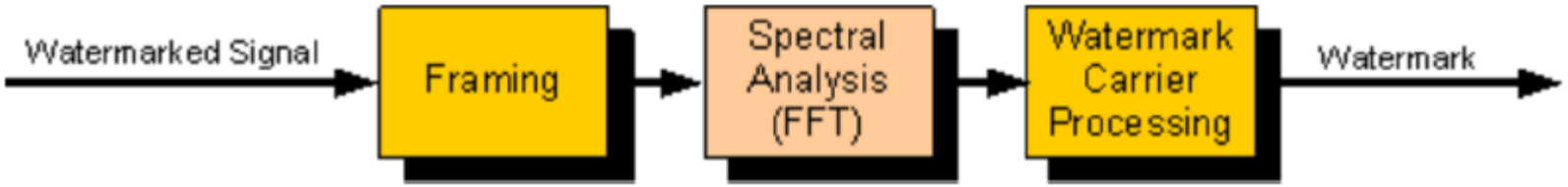
$$F(k) = \sum_{n=1}^{N} f(n) e^{\frac{-j2\pi(n-1)(k-1)}{N}} \qquad k = 1, 2, \ldots, N$$

DC Watermarking Scheme

- DC Removal
  - Dopo aver calcolato le componenti a bassa frequenza, queste possono essere rimosse
- Watermark signal addition
  - Dall'analisi dello spettro completata precedentemente è stato possibile calcolare la potenza per ogni frame
    - Che determina l'ampiezza del watermark che è possibile aggiungere

DC Watermarking Scheme

- Estrazione del watermark
  - Fasi analoghe a quelle di inserimento



- Limitazioni dello schema
  - Riguardano la robustezza e la densità dei dati
  - La robustezza migliora con file audio lunghi…
  - …e con l'inserimento multiplo del watermark
    - Aiuta anche l'estrazione e il riconoscimento delle manipolazioni

Altre tecniche

- Phase Encoding
  - Si sfrutta l'assoluta mancanza di sensibilità dell'orecchio umano rispetto ai cambiamenti di fase
- Spread Spectrum Watermarking
  - Si propaga il segnale del watermark sull'intero spettro delle frequenze udibili in modo tale da approssimare il rumore
    - Ad un livello di potenza tale da non essere udibile
- Echo Watermarking
  - Il segnale audio viene distorto in modo tale da essere percettivamente respinto dall'orecchio umano come distorsione ambientale

Audio Watermarking Technique Could Locate Movie Pirates

- Per individuare la *camcorder piracy* sono state sviluppate tecniche per aggiungere un messaggio segreto in un film
  - Per indicare in quale cinema e quando è stato visto
  - Una volta che viene condiviso su Internet, questo messaggio può essere estratto per capire da dove arriva la copia e quando è stata fatta
    - In congiunzione con un sistema di sorveglianza si può individuare il *pirata*
  - Queste tecniche non consentono di individuare automaticamente il *posto* dal quale è stata effettuata la ripresa

Audio Watermarking Technique Could Locate Movie Pirates

- Recentemente è stata proposta una nuova tecnica (Yuta Nakashima, Ryuki Tachibana, e Noboru Babaguchi dell'Università di Osaka)
  - Consiste in un sistema di stima della posizione basato sull'audio watermarking
  - Il watermark viene aggiunto alla colonna sonora per stimare la posizione del pirata
    - Il margine d'errore è compreso in mezzo metro
  - Come funziona? È più semplice di quanto si possa pensare

Audio Watermarking Technique Could Locate Movie Pirates

- Si sfruttano i diversi canali con cui viene trasmessa la colonna sonora *(host signals)*
- Un *embedder* genera un segnale di watermark diverso per ogni host signals, *watermarked host signals (WHS)*
  - Ogni altoparlante del cinema emette un WHS, e la camcorder del pirata registra un miscuglio di WHS
- In un segnale registrato mono, i WHS sono ritardati in proporzione alla distanza dall'altoparlante che li ha trasmessi

Audio Watermarking Technique Could Locate Movie Pirates

- Il *detector* del watermark può calcolare questi ritardi basandosi sulla lunghezza di ogni WHS
- Questo metodo può essere combinato con le tecniche convenzionali di watermarking e con impianti di videosorveglianza o sistemi avanzati di ticketing
  - Per determinare dove, quando e chi registra un film per poi metterlo in condivisione su Internet

Audio Watermarking Technique Could Locate Movie Pirates

- Prossime sfide
  - Rendere il sistema più robusto ed affidabile
  - Mantenere la qualità dell'audio elevata anche in presenza di fattori ambientali avversi e rumore di sottofondo
  - Investigare per trovare eventuali possibili attacchi che possono far stimare posizioni non rilevanti

Sommario

- Introduzione
  - Di cosa parliamo
  - Motivazioni
  - Un po' di storia
  - Ambienti applicativi
  - Proprietà e tipologie
  - Schemi generali
- Image Watermarking
  - Tecniche del dominio spaziale
  - Tecniche del dominio delle frequenze
  - Possibili attacchi
- Audio Watermarking
  - Schemi di Watermarking
  - *Audio Watermarking Technique Could Locate Movie Pirates*
- **Curiosità**

Curiosità

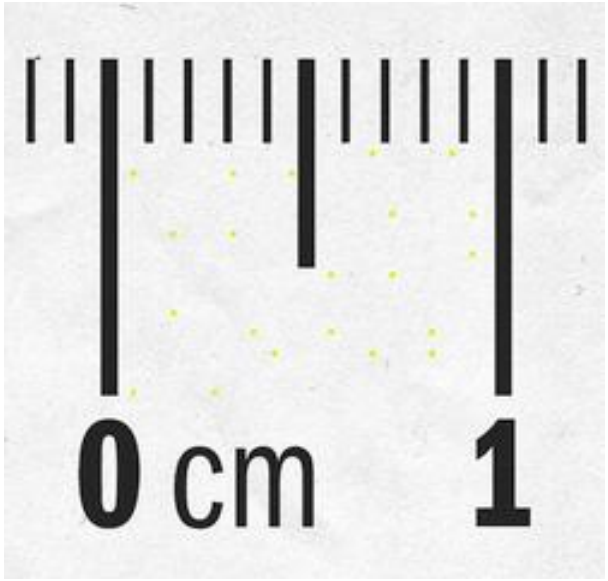
- Il caso dei Broadcast Flag per l'HDTV
  - A metà del 2005 è scoppiato un vero e proprio caso politico e legale riguardo all'uso di tecniche di digital watermarking nei flussi video della TV digitale
    - La FCC americana ha tentato di far passare abusivamente una regola che imponeva l'uso di appositi "Broadcast flags" nelle trasmissioni video digitali della televisione ad alta definizione (HDTV)

Curiosità

- Il caso dei Broadcast Flag per l'HDTV
  - Secondo questa regola, sarebbe diventato illegale vendere e distribuire sul territorio americano prodotti che non tenessero conto di questi broadcast flag e che non li usassero per limitare l'uso della trasmissione video (registrazione, copia, visualizzazione) in modo conforme a quanto deciso dal detentore dei diritti

Curiosità

- Nel 2003, la Electronic Frontier Foundation (EFF) ha scoperto che HP, Xerox ed altri produttori di stampanti usano delle tecniche di watermarking per marcare ogni foglio stampato in modo che sia sempre possibile tracciarne la provenienza.



Grazie per l'attenzione!