UNIVERSITÀ DEGLI STUDI DI TRIESTE

XXIV CICLO DEL DOTTORATO DI RICERCA IN INGEGNERIA DELL'INFORMAZIONE

PROGETTO ED APPLICAZIONI DI METODI PER L'ANALISI DELLA QUALITÀ DELLE IMMAGINI

Settore scientifico-disciplinare **ING – INF/01 Elettronica**

DOTTORANDA
FRANCESCA DARDI

COORDINATORE
PROF. WALTER UKOVICH

TUTORE
PROF. GIOVANNI RAMPONI

ANNO ACCADEMICO 2010 /2011

INDICE

I. INTRODUZIONE

L’analisi della qualità dell’immagine rappresenta attualmente un ambito di ricerca estremamente interessante ed innovativo per quanto riguarda molteplici settori di applicazione, dal multimediale, al biomedico, al forense. Trattare l’argomento in modo esaustivo è quindi particolarmente complesso, a seconda della particolare tipologia di applicazione possono essere definiti di conseguenza diversi criteri per esprimere la qualità dell’immagine in analisi e diverse categorie di algoritmi per la determinazione di criteri di valutazione.

Nel presente lavoro di tesi, si vuole cercare di dare un quadro sufficientemente ampio ed originale di quali possano essere stime efficaci della qualità di un’immagine, in particolare all’interno di una sequenza video ed in fase di studio di nuovi sistemi multimediali ad alta qualità, e con tale termine intendendo la qualità percepita dall’utente finale del sistema: parliamo quindi di qualità soggettiva, argomento ancora più circostanziato nel campo di studio individuato.

Sono stati finora implementati diversi metodi di valutazione e di miglioramento della qualità, soprattutto se stimata in termini oggettivi: più raramente in termini di percezione della persona – osservatore, tenendo in considerazione le proprietà ed i modelli del sistema visivo umano.

Tali metodi vengono particolarmente utilizzati nel settore dell'elaborazione dell'immagine: applicazione tipica è l’inserimento nella catena di trasmissione ed elaborazione che caratterizza il trattamento del segnale video in sistemi multimediali di varia natura. La fase di valutazione della qualità è particolarmente significativa in relazione ad un’altra operazione tipica di tali sistemi, ovvero la codifica del segnale atta a trasmettere il maggior numero di informazioni possibile in una banda relativamente contenuta. In fase di codifica, infatti, ciò che è desiderabile in primo luogo è l’ottenimento di un bit rate di trasmissione ragionevolmente basso in corrispondenza di una buona riproduzione del contenuto informativo inizialmente presente: la qualità conseguentemente percepita rappresenta quindi l’unico modo per formulare una stima in questo senso, ovvero di quanto fedelmente sia stata riprodotta l’informazione originaria.

La valutazione della qualità, in sistemi televisivi, viene tipicamente gestita in fase di ricezione: può essere necessario a tal fine introdurre l’utilizzo di un'immagine di riferimento [30]: proprio in situazione di compromissione del contenuto informativo, tale riferimento costituisce un reale supporto a metodi di ricostruzione delle informazioni compromesse, calibrando le diverse fasi del metodo.

In altre situazioni, invece, la valutazione della qualità viene gestita in assenza di riferimenti: tali metodi vengono tipicamente indicati con il termine di *"no-reference"* [35] [41][66] e costituiscono un settore di ricerca altamente innovativo in quanto ancora poco analizzati in confronto a quanto già proposto nello stato dell'arte riguardo a metodo basati su un riferimento.

Ancora una volta, il presente lavoro si propone di analizzare il caso meno banale, quindi il caso di valutazione della qualità percepita in assenza di immagini di riferimento per la valutazione della stessa. Ed ancora una volta, a seconda dell'ambito di applicazione, i criteri possono risultare estremamente diversificati a seconda dell'obiettivo da raggiungere: la qualità percepita dall'utente finale rimane in ogni caso il fine comune, che si tratti di un operatore forense in fase di riconoscimento di una traccia sensibile o che si tratti semplicemente di un utente di un sistema video di intrattenimento.

Principalmente, nel nostro lavoro, prenderemo in considerazione il caso in cui il dispositivo di visualizzazione dell'immagine in analisi sia costituito da un monitor televisivo, riferendoci in modo abbastanza dettagliato all'analisi della qualità in ambito di sistemi multimediali: ciò in modo tale da poter identificare, senza la specificità propria di determinati campi di applicazione, quali possano essere gli artefatti a danno della qualità stessa. In sostanza, tali situazioni possono richiedere il miglioramento del dettaglio senza accentuare di conseguenza gli artefatti prodotti da altre fasi di elaborazione della catena video, prima fra tutte la codifica e le relative cause concorrenti.

Altro elemento fondamentale per l'analisi di qualità di questi sistemi: il reale danneggiamento dell'informazione originaria va distinto da tutto ciò che viene artificialmente introdotto, o da altre operazioni di elaborazione o volutamente attuate dall'operatore stesso sia in fase di trasmissione che in ricezione, e ciò può essere realizzato solo attraverso un'attenta e differenziata analisi degli artefatti rilevanti in termini di qualità soggettiva percepita dall'osservatore – utente finale del sistema [29] [33].

Ad esempio, in una situazione in cui il segnale televisivo venga visualizzato su un monitor dalle caratteristiche note e riferibili allo stato dell'arte, si può incorrere in amplificazione di segnale che miri alla valorizzazione del dettaglio; l'operazione di taratura, tuttavia, non viene attualmente effettuata nel rispetto della qualità percepita, quanto della qualità oggettivamente misurata, e spesso non vi è congruenza con i termini di accettabilità dell'artefatto percepito in una valutazione soggettiva. Per questo motivo, dal punto di vista di un produttore del contenuto televisivo, è particolarmente importante superare canoni di valutazione oggettivi e basati su parametri standard come il rapporto segnale rumore, perché tali descrittori potrebbero non soddisfare l'osservatore, che rappresenta l'utente finale del prodotto nonché il cliente da un punto di vista strettamente commerciale.

Molte situazioni chiave nell'ambito dell'analisi della qualità sono state al momento trattate in maniera piuttosto riduttiva: sicuramente si tratta di migliorare ed accentuare le caratteristiche di dettaglio fine presente nell'immagine, e troppo spesso danneggiato dalla codifica, così come è necessario ridurre la presenza dei possibili artefatti che caratterizzano il segnale video a valle della catena di elaborazione.

Tuttavia, va considerato che il dettaglio, così come un dato artefatto, viene diversamente percepito dall'osservatore umano a seconda della situazione contingente e di diversi fattori, come ad esempio la presenza di moto nella sequenza video o di particolari cognitivamente significativi. Proprio per questi motivi si pone il problema, oltre alla necessità di valutazione della qualità in assenza di riferimenti, di definire dei criteri in un certo senso "adattativi" di stima, in modo tale da tenere presente come, in diverse condizioni, un artefatto oggettivamente della stessa entità può tradursi in valutazioni qualitative estremamente differenti da un punto di vista soggettivo. Sistemi a soglia e parametri staticamente definiti possono essere utili in una prima fase di misurazione oggettiva, ma tale misurazione deve essere integrata con modelli di comportamento del sistema visivo umano.

Si pone inoltre il problema di una valutazione differenziata anche in senso locale: sicuramente è necessaria una valutazione dell'immagine nella sua globalità, anche per fornire una risposta in termini di accettabilità o meno nell'ottica di una buona riproduzione del segnale. Tuttavia, sistemi di miglioramento della qualità, a valle di tale elaborazione, richiedono come informazione ulteriore, la determinazione delle aree dell'immagine in cui è particolarmente necessario intervenire in termini di attenzione soggettiva. Un deterioramento dell'immagine presente in una zona in cui si focalizza l'attenzione dell'osservatore, stimabile mediante vari metodi [27][28][36][68], andrà sicuramente trattato ai fini di un miglioramento, mentre un artefatto presente sullo sfondo potrà essere trascurato a seconda del livello di qualità percepita che si vuole raggiungere.

Diversi algoritmi di valutazione della qualità, presenti nello stato dell'arte, giungono ad una formulazione di stima in due diverse fasi: una prima fase comprende la rilevazione di specifici e noti artefatti, stimandone la loro entità. Una seconda fase, invece, prevede la ricerca di un collegamento, in termini di formulazione matematica, fra le diverse stime effettuate sui singoli artefatti con quella che è la valutazione della qualità complessiva di un'immagine in una sequenza video: in pratica, si cerca di definire come questa possa essere stimata in modo efficace, utilizzando un'appropriata combinazione delle diverse stime quantitative degli artefatti selezionati come significativi nella prima fase di studio.

Gli artefatti più frequentemente analizzati, in quanto ben noti e rilevanti dal punto di vista percettivo,

sono essenzialmente sfocatura e blocchettatura.

La blocchettatura conta sicuramente un gran numero di lavori nello stato dell'arte in termini di analisi [40]: questa situazione è, fra le altre cose, l'artefatto più tipicamente legato alla codifica. Ma fra gli artefatti imputabili alla compressione del video in termini di cause, nonché fra i più fastidiosi per l'osservatore umano, possiamo sicuramente individuare la sfocatura come uno fra i più importanti e meno banali da individuare, sia in termini di cause che di effetti, che rappresentano poi i presupposti per attuare una fase di miglioramento della qualità sufficientemente efficace.

Da queste considerazioni si deduce quella che è la motivazione del lavoro di tesi proposto: viene illustrato principalmente l'ambito di applicazione della multimedialità, con un breve accenno finale a quanto potrebbe essere importante l'estensione dei criteri proposti, con opportune specificità da considerare, ad altri ambiti specialistici come quello forense.

La possibilità di estendere le metodologie proposte ad altri ambiti, come appunto il forense, dipende essenzialmente da una caratteristica importante del lavoro svolto.

Vogliamo nuovamente sottolineare infatti come il lavoro presentato in questa sede si collochi a pieno titolo nella categoria di algoritmi di valutazione della qualità, o di artefatti fondamentali per la stima della stessa, implementati senza alcun ausilio di immagini di riferimento: queste ultime vengono frequentemente utilizzate in diversi metodi nello stato dell'arte, che di fatto sono basati su analisi differenziali di caratteristiche estratte dall'immagine in esame e da un riferimento noto, situazione decisamente poco plausibile in alcuni particolari settori di applicazione.

L'approccio "*no-reference*" diviene possibile grazie all'integrazione di diverse scelte, dalla selezione dell'artefatto significativo, alla definizione di opportune misurazioni oggettive, alla valutazione effettuata in termini percettivi grazie all'introduzione di determinate considerazioni, avanzate sulla base di modelli del sistema visivo umano già accettati dalla comunità scientifica.

La tesi viene organizzata quindi come segue: il capitolo II presenta lo stato dell'arte; nel capitolo III, vengono presentate le misurazioni proposte per le diverse tipologie di sfocatura, basate sia sulla misurazione della perdita di contrasto che sulla perdita di dettaglio, e di seguito, sono illustrati modelli del comportamento del sistema visivo umano selezionati come utili al nostro lavoro.

Quindi vengono considerate le diverse origini della sfocatura e le manifestazioni relative in corrispondenza dei valori delle metriche implementate, integrando di conseguenza i modelli oggettivo e soggettivo, in relazione con le diverse tipologie di misurazioni e con le cause dell'artefatto.

Il capitolo IV illustra ulteriori modalità di valutazione locale della sfocatura in relazione alla codifica,

con conseguenti risultati sempre in confronto a quanto percepito da osservatori umani.

Nel capitolo V vengono riportati i risultati sperimentali relativi sia a misurazioni di sfocatura a livello di causa che a livello di intensità locale da codifica e nel capitolo VI vengono tracciati quelli che potrebbero essere i possibili sviluppi futuri del lavoro di tesi in ambio di multimedialità.

Nel capitolo VII troviamo infine alcune considerazioni valide nelle applicazioni al settore forense e di seguito, nel capitolo VIII, le conclusioni.

II. STATO DELL'ARTE

Lo stato dell'arte da analizzare in termini di valutazione di artefatti presenti in immagini fisse ed in sequenze video è piuttosto variegato e difficile da sintetizzare. Facilmente, l'analisi della qualità viene riportata ad ambiti e ad applicazioni specifiche, mentre ben poca bibliografia viene rilevata nel caso dell'analisi delle cause di sfocatura in relazione alla perdita di qualità conseguente.

Più frequentemente, simili analisi di qualità, o più specificatamente di artefatti riconducibili a stime effettuate, vengono implementate in relazione a codifiche video proprie di standard avanzati: esempio tipico è rappresentato dalle applicazioni di algoritmi di qualità a possibili configurazioni di codifica previste dallo standard H264. Determinati tipi di codifiche, infatti, comprendono già nella catena di elaborazione del video la presenza di specifici stadi atti a ridurre determinati artefatti da codifica, fra cui citiamo la blocchettatura, che in qualche modo è possibile prevedere soprattutto in caso di "*bit rate*" di trasmissione particolarmente bassi e corrispondenti ad una maggiore perdita di informazione, almeno nella maggior parte dei casi. Tuttavia, il rischio che si corre in tali casi è quello di sostituire un artefatto all'altro, nel caso di riferimento introducendo sfocatura nell'immagine elaborata in termini di riduzione della blocchettatura presente.

Altri stadi nella catena di elaborazione del video che devono essere considerati come potenziali fonti di sfocatura possono essere: la fase di denoising; la fase di deinterlacciamento; la fase di conversione ad una risoluzione superiore. In tali casi, sicuramente va considerata la rilevazione degli artefatti presenti ed il modo in cui si manifestano a livello locale nell'immagine (o *"frame"*) è piuttosto differente, a seconda dello stadio di produzione, dal punto di vista della degradazione della qualità percepita.

Queste possono essere solo alcune delle situazioni che implicano la generazione di sfocatura in uno dei possibili stadi della catena di elaborazione del video: in alcuni casi vengono già previste delle compensazioni, in altri invece è particolarmente difficile la compensazione, soprattutto se l'immagine già in origine è particolarmente affetta da sfocatura. Questa è ad esempio la situazione di fuori - fuoco ottico, ovvero di artefatto già presente in fase di acquisizione dell'immagine.

Oppure, ulteriore causa di deterioramento del dettaglio, tipica di una sequenza video che già originariamente ne è affetta, può essere identificata nella ripresa di una scena contenente oggetti in movimento, percepiti dall'osservatore finale come carenti dal punto di vista informativo se considerati

singolarmente come immagine. Una simile situazione va distinta dalle precedenti in termini di analisi qualitativa, proprio perché un trattamento inopportuno in fase di miglioramento della qualità stessa potrebbe essere addirittura controproducente dal punto di vista percettivo.

In termini di effetti osservabili, e quindi in termini di azione successiva di miglioramento, le singole cause di sfocatura precedentemente illustrate vanno distinte: la codifica, ad esempio, agisce modificando gli oggetti nella scena; agendo sui contorni che divengono più sfumati e spessi; attenuando ulteriormente le tessiture fini e preservando invece ciò che è già netto e chiaramente visibile.

La sfocatura ottica, invece, proprio perché imputabile alla fase stessa di acquisizione, provoca una sfumatura dei contorni che è ancora più accentuata rispetto a quanto detto per il caso precedente e ne sono affetti anche i contorni di oggetti di una certa entità: ovviamente, a maggior ragione, ne sarà interessato il dettaglio fine e gli oggetti sullo sfondo, in maniera tale che il dettaglio stesso potrebbe non essere più visibile e quindi difficilmente ricostruibile anche con le procedure efficaci in casi differenti.

Ancora, l'operazione relativa ad un passaggio a risoluzione superiore dell'immagine, si manifesta in una generale perdita di attività e di contrasto ovviamente percepibile su tutta l'estensione dell'immagine stessa, anche se in determinati casi non in maniera uniforme, in quanto lo scalaggio può essere effettuato in maniera particolarmente accentuata sulle porzioni laterali dell'immagine, mentre la zona centrale mantiene quasi intatte le proprie caratteristiche di attività, in quanto l'area in primo piano va maggiormente preservata perché primaria dal punto di vista della percezione.

Proprio ragionando in termini di riduzione percepita del contrasto, possiamo far notare come la differenza di luminanza possa essere diversamente percepita a seconda delle caratteristiche della stessa nella regione in analisi: questo significa che i parametri di valutazione delle varie cause e dell'entità della sfocatura, imputabili o meno alla catena di elaborazione video, vanno definiti a livello locale.

In tutte le diverse situazioni, nondimeno va considerato l'insieme di fattori che impattano sulla valutazione della qualità soggettiva, considerando anche quando la stessa non rappresenta un degrado della qualità complessiva: ad esempio, nel caso di sfocatura volutamente introdotta dal cineoperatore per far risaltare maggiormente oggetti in primo piano rispetto allo sfondo della scena.

Essere in grado di distinguere questo tipo di sfocatura nel contenuto dell'immagine da quelle non desiderabili dovute a cattiva acquisizione, post-processing o codifica troppo restrittiva diviene un obiettivo estremamente interessante: le diverse situazioni finora citate sono state trattate singolarmente

in diversi lavori, come avviene in [2] [5] [7] [9] per es. in relazione a tecniche di *"autofocusing"*, ma mai prendendo in analisi la situazione nel suo complesso,

Ecco alcuni esempi dei lavori svolti nei diversi ambiti, alcuni basati sulla rilevazione della sola sfocatura imputabile a codifica, altri dal carattere più generale.

Gli autori Parvez, Sazzad, Kawayoke e Horita [3] [22], si rivolgono all'analisi dell'artefatto da codifica, individuando delle caratteristiche a livello spaziale per determinarne la presenza. Un merito di tale approccio consiste nel fatto che la qualità percepita dell'immagine viene considerata proprio nei casi maggiormente critici: vengono sottoposti a valutazioni diversi contenuti multimediali, trasmessi mediante Internet o dispositivi mobili di terza generazione, cercando di superare le valutazioni più tradizionalmente accettate in termini di qualità e di rilevazione di artefatti. Il metodo proposto utilizza determinate caratteristiche coerenti con la sensibilità del sistema visivo umano, sfruttando le informazioni relative, ad esempio, a bordi degli oggetti presenti nella scena ed a misurazioni della distorsione introdotta da artefatti eventualmente rilevabili. In fase di risultati sperimentali, i test vengono effettuati su immagini compresse JPEG2000 e ne viene fatta una verifica di coerenza rispetto a risultati ottenuti con misurazioni soggettive.

Oltre all'osservazione del deterioramento del dettaglio e della distorsione, è possibile riferirsi anche ad altre possibili metodologie di misurazione dell'attività presente. Infatti, vari metodi di rilevazione e misurazione della sfocatura propongono l'identificazione dell'artefatto in termini statistici e molto generali, ad esempio come valutazione della perdita di attività e di contrasto, utilizzando un approccio tipicamente basato su una quantificazione statistica della variazione locale a livello di pixel, tramite Kurtosis, o deviazione standard o misurazioni di entropia.

Citiamo gli autori Xia, Shi, Teunissen e Heynderickx [1] [21], che riferiscono la loro metrica di rilevazione della sfocatura alla misurazione della varianza, ed in particolare all'osservazione di come tale valore statistico venga alterato, a valle dell'operazione di compressione ed in confronto al valore calcolato nell'immagine originale e scevra da artefatto. La funzione di sensibilità al contrasto propria del sistema visivo umano viene introdotta per definire diverse bande di frequenza, in ciascuna di esse viene quindi separatamente calcolata la varianza, in modo da poter distinguere la variazione della stessa nelle diverse bande per selezionare le variazioni di maggior interesse e calibrare di conseguenza l'accuratezza della metrica. I risultati ottenuti in fase sperimentale vengono quindi confrontati con una serie di misurazioni soggettive di sfocatura, ottenendo un'elevata correlazione con esse a conferma della robustezza del metodo.

In modo maggiormente coerente con la percezione soggettiva, gli autori Caviedes e altri [23] [4], invece, propongono una metrica che valuta per prima cosa l'entità dei contorni presenti nell'immagine, per poi ricondursi ad una valutazione statistica. Viene elaborato lo spettro di frequenza dell'immagine, in quanto strettamente correlato alla visibilità del dettaglio insito nel contenuto informativo della sequenza. Tuttavia, la metrica proposta non si riferisce semplicemente al gradiente spaziale, troppo dipendente dal contenuto di quel particolare video e inconciliabile con l'approccio proprio di un metodo di qualità "senza riferimento": di fatto, viene creato un profilo di intensità dei bordi presenti, identificando i pixel appartenenti agli stessi e suddividendo l'immagine in blocchi di 8x8 pixel. Per ciascun blocco, l'entità di tali profili viene stimata facendo uso dell'approccio statistico, valutando la kurtosis della DCT, e la metrica finale viene elaborata effettuando l'operazione di media sui blocchi ed in termini di entità dei profili. La combinazione fra informazioni valutate sia nel dominio spaziale che nel dominio delle frequenza conferisce robustezza e flessibilità al metodo, la valutazione dei profili di intensità si dimostra abbastanza congruente con quanto poi effettivamente osservato ed i risultati sperimentali mostrano anche in questo caso una buona correlazione con misurazioni effettuate a livello soggettivo.

Ancora un metodo basato su approccio statistico viene proposto da Bovik e altri in [38] - altri lavori riconducibili allo stesso autore, molto attivo in tale ambito, sono ad esempio [24] [39]. Sicuramente un merito di tale lavoro è l'elaborazione di una stima di qualità senza alcun ausilio in termini di informazioni a disposizione relative all'immagine in origine, nonché sulla possibile distorsione presente nel video da analizzare. L'algoritmo elaborato viene inizialmente pensato e provato su video caratterizzati da statistiche di tipo naturale (NSS) ed essenzialmente viene attuato in due tempi, dei quali una prima fase di training atta alla taratura dell'indice proposto. Una volta effettuato il training del metodo con un dato insieme informazioni, non viene più richiesta alcuna conoscenza del contenuto in esame e della distorsione presente: inoltre, essendo il processo modulare, può facilmente essere esteso a diverse tipologie di distorsione. L'indice proposto viene infine calcolato per le immagini del database maggiormente utilizzato per algoritmi di rilevazione della qualità, ed accettato dalla comunità scientifica, ovvero il LIVE database.

Oltre all'incompletezza dell'analisi già menzionata, si vuole sottolineare, riferendosi ai precedenti casi, che approcci basati su valutazioni statistiche difficilmente potrebbero essere impiegati in ambito dell'elettronica di consumo né vengono considerati in questo lavoro: la richiesta di una grande quantità di campioni rappresenta un ostacolo in questo senso ed in determinati casi tali metodi mal si adattano

alla valutazione delle proprietà locali dell'immagine, specialmente in presenza di caratteristiche altamente variabili da una regione all'altra, situazione che invece intendiamo considerare in questa sede.

D'altronde, l'elaborazione di misurazioni statistiche di tipo complesso è difficile da integrare con modelli rappresentativi del sistema visivo umano, a livello algoritmico: generalmente le caratteristiche del sistema visivo vengono ricondotte alla valutazione statistica stessa e la congruenza viene ricercata solo in un riscontro sperimentale con misurazioni soggettive.

Vediamo ora altre possibili formulazioni rilevate in letteratura, basate essenzialmente sulla valutazione dei coefficienti della trasformata DCT o su Wavelet.

Gli autori Marichal, Ma e Zhang [8], ad esempio, propongono una metrica piuttosto semplice, ma non per questo meno robusta, elaborata a livello globale per la rilevazione della sfocatura presente in un'immagine. Il vantaggio della proposta è che viene costruita direttamente sull'istogramma dei coefficienti DCT non nulli, e quindi direttamente applicabile ad immagini e video in forma compressa (MPEG or JPEG) ed a tutti i tipi di *"frame"* previsti dalla codifica MPEG (I-, P- or B-frame). L'analisi in frequenza rappresenta di nuovo il collegamento con e proprietà del sistema visivo umano ed anche in questo caso, vengono effettuati test di confronto con misurazioni di tipo soggettivo.

Metodi come questo, tuttavia, se inseriti in una catena di elaborazione video, sono poi effettivamente meno vantaggiosi di quanto si possa inizialmente pensare: di fatto, essi si basano su informazioni generalmente non ancora disponibili allo stadio in cui la valutazione della qualità deve essere implementata. Va pensata quindi una soluzione contestualizzabile nella sequenza di elaborazione propria del canale video rivolto ad un monitor televisivo: ciò sarebbe sicuramente più interessante qualora la valutazione qualitativa venga finalizzata all'ambito multimedia.

Per quanto riguarda casi di sfocatura presente già in origine nella sequenza video, come ad esempio se essa è imputabile ad acquisizione, una valida stima può essere elaborata a partire dalla valutazione della point-spread-function: questo a meno che l'artefatto introdotto non si manifesti in modalità particolarmente accentuata, come in alcune delle situazioni che invece intendiamo prendere in considerazione nel presente lavoro.

In termini di semplicità ed efficacia, il miglior metodo attualmente elaborato, è tuttavia quanto variamente proposto da diversi autori e basato sull'analisi dello spessore dei bordi degli oggetti presenti nella scena. La media delle misurazioni di spessore effettuate viene considerata una stima della sfocatura presente: il criterio permette quindi di rilevare e correggere la sfocatura, poiché considerando

lo spessore dei bordi possono essere applicate diverse operazioni di "*enhancement*".

Gli autori Marziliano, Dufaux, Winkler e Ebrahimi [10] [25], presentano un lavoro sulla rilevazione della sfocatura percepita basato proprio sull'analisi dello spessore dei bordi: il metodo è estremamente conveniente dal punto di vista computazionale ed inseribile in contesti di realizzazione in tempo reale. La soluzione proposta viene quindi validata su varie tipologie di contenuto informativo, e come di consueto tali misurazioni vengono confrontate con criteri di congruenza con misurazioni soggettive effettuate, né altre considerazioni vengono avanzate riguardo a modelli del sistema visivo umano. Il vantaggio principale rimane, in ogni caso, la possibilità di applicazione a codifiche ottimizzate e ad una gestione efficace delle risorse di rete.

Ancora, gli autori Ong e altri [11], propongono una metrica per la valutazione percettiva della sfocatura, caratterizzata dallo '*spread*' valutato in direzione opposta a quella del gradiente relativamente al contorno in analisi. La valutazione della validità soggettiva viene riferita ad immagini JPEG-2000, rilevando una buona correlazione con risultati di valutazione soggettiva. Oppure in lavori successivi viene inserita l'analisi di ulteriori valutazioni statistiche[6].

Da sottolineare nuovamente come l'elemento soggettivo, in tutti i lavori visti finora, non venga direttamente integrato nel metodo ma solo introdotto in termini di confronto fra risultati ottenuti dall'algoritmo e valutazione soggettive di operatori, in modo assolutamente empirico. Inoltre questi metodi, seppur molto vantaggiosi dal punto di vista computazionale, sono basati su una misurazione estremamente semplificativa dell'artefatto. Infine, in casi di sfocatura estremamente accentuata, come nel caso di sfocatura da acquisizione, i bordi di oggetti e tessiture possono essere addirittura erosi o completamente assenti, non potendo essere in tal modo utilizzabili per costituire un indicatore.

Un lavoro che propone un primo esempio di integrazione fra misurazione oggettiva e modelli del sistema visivo umano viene presentato dagli autori Ferzli e Karam [12] [26]. L'articolo presenta una metrica basata su determinate proprietà del sistema visivo umano, considerandone la sensibilità in funzione delle proprietà locali dell'immagine ed introducendo il concetto inerente la minima variazione percepita dall'occhio in determinate condizioni di intensità della zona in osservazione. La metrica non utilizza alcun riferimento per la valutazione della qualità percepita ed incorpora misurazioni dell'entità dei bordi degli oggetti presenti nella scena in un modello probabilistico. A differenza di tutti gli altri metodi già visti e basati sulla valutazione dell'entità dei contorni, questa metrica è utilizzata per la predizione dell'entità della sfocatura presente in immagini di diverso contenuto informativo. I risultati ottenuti dal confronto con le misurazioni soggettive confermano la congruenza del metodo.

In questo lavoro di tesi, anche questo modello viene superato e vengono introdotti e integrati diversi modelli relativi a caratteristiche proprie del sistema visivo umano, affiancati da misurazioni oggettive dell'artefatto in analisi. Inoltre, vengono illustrati alcuni importanti risultati e vengono introdotte diverse tipologie di metriche che abilitano alla possibilità di distinguere fra le diverse manifestazioni della sfocatura in relazione alle cause che provocano la stessa. Di fatto, il problema della sfocatura viene analizzato nel suo complesso, come mai introdotto nello stato dell'arte: la sfocatura dovuta ad acquisizione, o a scalaggio a maggior risoluzione, viene differenziata dal caso di codifica scadente o intenzionalmente introdotta.

In un secondo momento, si propone un'analisi approfondita soprattutto per quanto riguarda l'artefatto dipendente da codifica, definendo stime sia a livello globale che locale e quantificando l'entità in funzione di una successiva applicazione di processi migliorativi della qualità.

Infine, gli algoritmi proposti in questa sede presentano varie possibilità di integrazione fra le misure oggettive parametriche individuate ed i modelli della percezione soggettiva già accettati dalla comunità scientifica: in ogni caso, al termine della procedura, verranno confrontati i risultati ottenuti, in termini di valutazione della qualità, da parte di osservatori umani e da parte dell'algoritmo, in modo tale da verificarne la congruenza, ottimizzandone la correlazione.

III. MISURAZIONI GLOBALI E CAUSE DI SFOCATURA

Misurazioni di sfocatura proposte

Per avviare uno studio più completo riguardo alla rilevazione dell'artefatto di sfocatura, come detto, diviene necessario analizzare le diverse manifestazioni dello stesso in relazione alle cause che ne sono all'origine, associando alle varie casistiche delle misurazioni che possano adeguatamente descrivere e rappresentare la situazione che si verifica innanzitutto a livello oggettivo.

La componente percettiva viene introdotta principalmente in un secondo momento, come vedremo, e sarà necessario identificare l'importanza soggettiva di quanto rilevato procedendo ad un'integrazione complessa con modelli selezionati alla descrizione del sistema visivo umano.

La misurazione di sfocatura che intendiamo proporre è costituita da due differenti componenti, in modo da poter ricondurre, grazie ad un'osservazione combinata di ciascun tipo di analisi, gli effetti della sfocatura a caratteristiche diverse ed imputabili a origini di natura differente.

La prima misurazione riguarda in particolare una stima in percentuale dell'area dell'immagine in esame interessata da un basso indice di contrasto, calcolato secondo formule note e nel rispetto di quanto percepito da un osservatore, come prima stima dell'informazione determinante per il nostro sistema visivo.

La seconda misurazione rileva invece il deterioramento o addirittura l'assenza di dettaglio fine, imputabile ad es. ad artefatto da codifica in modo abbastanza tipico, che di fatto introduce un danno nelle componenti che corrispondono ad elevate frequenze spaziali.

Come vengano definite tali misurazioni viene esaurientemente spiegato nel seguito [45] a partire da quanto riportato in più lavori pubblicati [43][44].

In questa sede si cerca infatti di raccogliere in modo completo tutti i dati, collegando i risultati ottenuti nelle varie tappe della nostra ricerca.

a) Prima misurazione: l’ area a contrasto ridotto

Inizialmente, sull’immagine di test, viene eseguita una misurazione volta a rilevare il contrasto percepito dall’osservatore, senza riferirsi a particolari effetti e situazioni imputabili a definite cause di sfocatura eventualmente presenti. Tale contrasto viene misurato localmente e con formule note, in modo da associarne un valore per ogni pixel dell’immagine in esame, che alla fine va analizzata nella sua globalità proprio per individuare, in questa prima misurazione, situazioni di sfocatura generalizzata. Vediamo prima in dettaglio quelle che sono le considerazioni a fondamento del calcolo effettuato per quanto riguarda il contrasto percepito a livello locale per il singolo pixel.

A questo proposito, devono essere essenzialmente considerati due fattori basilari dal punto di vista percettivo:

1. per prima cosa, la differenza in livelli di grigio: questa fornisce però solo un'informazione parziale del contrasto percepito fra gli stessi. È necessario valutare a questo scopo gli effettivi valori di luminanza dei pixel dello schermo, ovvero i livelli di grigio espressi come interi vanno mappati in valori di luminanza dalla funzione non lineare gamma;

2. il secondo elemento da valutare si deduce da studi sulla ricettività dei fotorecettori, che hanno dimostrato come la stessa variazione di luminanza venga differentemente percepita in dipendenza da un valore locale di luminanza L e dall'adattamento dell'occhio, indicato con S, alla luminanza della regione circostante.

In definitiva, la misura proposta per la valutazione del contrasto locale a livello di singolo pixel è rappresentata dal numero *Njnd(i, j)* di variazioni di luminanza percepite distintamente (o *Just Noticeable Differences*, JND) da un osservatore nella regione centrata nel pixel *(i, j)*.

Tale numero viene così calcolato.

Come primo passo, viene elaborato un modello di percezione di luminanza in base al quale è possibile calcolare le variazioni di luminanza relativa, chiamate soglie di contrasto, necessarie a produrre un JND [14], in dipendenza dal valore locale di luminanza L e dal grado di adattamento S, che per semplicità viene assunto come il valor medio della regione circostante centrata nel pixel; tale soglia di contrasto viene definita dalla formula:

$$c_t = c_{to} \cdot (L + S)^2 / (4 \cdot L \cdot S)$$

dove c_{to} è un parametro sperimentalmente determinato.

Una volta definito c_t, il numero di JND rilevato fra due valori di luminanza *Lmin* e *Lmax*, assumendo che il parametro c_t non subisca variazioni in tale intervallo, viene definito come il massimo numero intero *N* che verifica l'equazione:

$$Lmin \cdot (1 + c_t)^N < Lmax$$

Conseguentemente, il numero *Njnd(i, j)* viene definito come numero di JND percepiti fra il massimo ed il minimo valore di luminanza in una regione di area ridotta e centrata nel pixel *(i, j)*, scelta di struttura circolare e centrata nel pixel. Il diametro del disco da considerare varia in accordo con le dimensioni e con la risoluzione dello schermo: tale angolatura è determinata in base all'angolo coperto da una regione della retina detta foveola, in cui è situato il fuoco di attenzione dell'occhio. Ciò significa, per una dimensione dello schermo pari a 19" e 1440x900 pixel, osservato ad una distanza di 63 cm, un angolo corrispondente ad un diametro di 40 pixel nell'immagine digitale sotto test.

Per ottenere una stima globale, la metrica proposta viene calcolata come frazione del *"frame"* in cui tale livello di contrasto viene percepito come basso relativamente al massimo percepito nell'immagine in analisi.

La misurazione proposta per la sfocatura si basa quindi sul contrasto così definito e rappresentato dalla percentuale di "*frame"* costituita da pixel con contrasto dal valore appartenente ad un intervallo determinato. Gli estremi di tale intervallo sono stabiliti come percentuale del massimo valore di contrasto indicato come $Njnd_{max}$.

Formalmente, la metrica Fsm_k è data da [44][45]:

$$Fsm_k = 100 \cdot (\#A_k / N_t)$$

dove:

- N_t è il numero complessivo di pixel nell'immagine,
- $A_k = \{(i,j) | Njnd(i,j) \in R_k\}$,

- $R_k = [Njnd_{k,low}; Njnd_{k,high}]$

e:

- $Njnd_{k,low} = (low_{\%,k}/100) \cdot Njnd_{max}$
- $Njnd_{k,high} = (high_{\%,k}/100) \cdot Njnd_{max}$

Diverse metriche possono essere definite in tal modo, ovvero ogni possibile metrica Fsm_k viene definita attraverso l'assegnazione di una coppia di valori in percentuale del massimo numero di livelli $Njnd_{max}$, ovvero

$$C_k = (low_{\%,k}, high_{\%,k})$$

b) Seconda misurazione: il livello di dettaglio preservato

Questa seconda misurazione viene selezionata in modo maggiormente mirato ad individuare particolari situazioni di interesse, soprattutto nell'ambito di studio del multimedia, come la sfocatura da codifica. Come principio di base, è ragionevole assumere che il dettaglio fine e le tessiture siano individuabili come parte del contenuto dell'immagine più probabilmente assoggettato all'artefatto di sfocatura, in particolar modo alla sfocatura introdotta da codifica scadente.

Contorni di oggetti già sfumati o limitati in estensione tendono a divenire quasi impercettibili e regioni caratterizzate da tessiture poco estese tendono a perdere il contrasto in modo tale da apparire quasi uniformi; in contrapposizione, la codifica non interferisce con bordi di oggetti ben evidenti, che rimangono altamente visibili.

A fondamento della definizione di un parametro oggettivo utile a rilevare la situazione appena descritta, vi è la scelta di un operatore matematico abile a riprodurre artificialmente un simile effetto, ovvero l'introduzione di sfocatura sul dettaglio già poco evidente ma non su caratteristiche visive ben

accentuate: applicare tale operatore all'immagine di test rende possibile decidere se l'immagine in esame era già stata soggetta o meno a perdita di dettaglio derivante da codifica scadente.

Infatti, a questo proposito, è possibile confrontare semplicemente il contenuto dell'immagine di test con quello dell'immagine stessa elaborata secondo l'operatore scelto per riprodurre l'effetto voluto: se il confronto fra valutazioni del dettaglio presente, utilizzando ad esempio il valore medio dei gradienti calcolati sull'immagine, è molto diverso nei due casi, ciò indica che l'operatore ha agito eliminando molto dettaglio dall'immagine di test, che ne era quindi in partenza ricca. In caso contrario, situazioni confrontabili in termini di gradienti, a monte e a valle dell'applicazione dell'operatore prescelto, possono indicare una situazione critica già in partenza, in quanto assenti dettaglio fine e tessiture, effetto potenzialmente imputabile all'azione della codifica.

Un operatore sfruttabile per una simile procedura è ad esempio la diffusione anisotropa [15], nota come operatore che preserva i bordi principali degli oggetti mentre ne riduce il rumore presente a livello spaziale.

Quanto detto finora converge nella definizione di una metrica per la determinazione della sfocatura, in termini di dettaglio preservato presente nell'immagine, così come segue:

- l'immagine di test *I* viene elaborata con la diffusione anisotropa, in modo da ottenere l'immagine *Id*.
- il gradiente morfologico viene valutato per ogni pixel delle immagini *I* e *Id*, ottenendo due mappe indicate rispettivamente con *Ig* e *Idg*.
- la metrica proposta per la valutazione della perdita di dettaglio viene determinata quindi come la differenza relativa delle medie spaziali valutate sulle mappe di gradienti così ottenute, secondo la formula:

$$MGR_{rel} = (mean\{Ig\} - mean\{Idg\}) / mean\{Ig\}$$

Selezione delle regioni a rilevanza percettiva

Alle misurazioni di sfocatura precedentemente illustrate, è a questo punto necessario affiancare altre considerazioni di tipo percettivo, riconducibili a modelli del sistema visivo umano, per ottenere valutazioni coerenti dell'impatto della sfocatura eventualmente presente in termini di qualità percepita da un osservatore. Vari studi sono stati implementati in questo senso e la bibliografia nello stato dell'arte è piuttosto ricca di riferimenti in proposito, ovvero riguardo alla possibilità di individuare all'interno di una scena quali elementi possano catturare l'attenzione in modo più immediato.

Partendo dall'immagine in analisi, quindi, è possibile determinare, a partire dai diversi modelli corrispondenti agli studi effettuati, quale sia la regione e/o gli oggetti focalizzati immediatamente dall'attenzione di un osservatore umano.

In questa sede ci prefiggiamo di combinare due diversi tipi di approccio al problema, ovvero si cerca di abbinare procedure di determinazione di aree di "*spot of attention*" a procedure di estrazione degli oggetti presenti in una scena, in modo da distinguerne quelli così detti "in primo piano". A tale scopo si procede all'integrazione di modelli del sistema visivo umano con opportuni algoritmi di segmentazione, sperimentalmente coerenti con quanto percepito in termini di regioni estratte all'interno della scena in esame.

L'integrazione di questi processi porta al procedimento proposto ai fini dell'estrazione delle regioni in primo piano ("*foreground"*), che consiste essenzialmente in due passi:

1. il "*frame"* viene segmentato in regioni dalle caratteristiche omogenee e coerenti con la cognizione dell'osservatore, prendendo come riferimento l'algoritmo di segmentazione successivamente descritto e già approvato dalla comunità scientifica.

2. vengono selezionate fra le regioni individuate quelle che soddisfano determinati criteri di rilevanza percettiva, secondo il modello scelto per descrivere le proprietà del sistema visivo umano, già approvato dalla comunità scientifica, e descritto nel seguito.

Una volta individuate le regioni appartenenti al "*foreground"*, il "*background"* viene identificato dalle regioni rimanenti. Artefatti presenti nelle due zone così definite assumeranno ovviamente significati molto diversi dal punto di vista della nostra analisi.

a) L’algoritmo di segmentazione adottato

Per quanto riguarda la segmentazione in regioni, viene utilizzato l’algoritmo di Felzenswalb e Huttenlocher descritto in [16]: questo algoritmo di segmentazione non è stato implementato tenendo in considerazione le peculiarità cognitive della visione percettiva, ma alla luce degli esperimenti fatti, si è dimostrato coerente con le osservazioni soggettive.

Questo algoritmo, ma non è l’unico, ha la particolarità di essere già stato collaudato in termini di efficacia nella descrizione percettiva degli oggetti componenti una scena, secondo una serie di esperimenti eseguiti e illustrati in [42].

L’importanza di questo tipo di valutazione, ovvero di analisi in termini di regioni componenti la scena, si evince da quanto emerge dagli esperimenti, volti a mettere in relazione il tempo medio impiegato dall’osservatore umano nel focalizzare un determinato elemento dell’immagine ed il numero di regioni in uscita dall’algoritmo di segmentazione applicato all’immagine stessa.

Possiamo affermare quindi, che, all’aumentare del numero di regioni rilevate dall’algoritmo di segmentazione scelto, aumenterà in relazione il tempo medio impiegato dall’utente finale umano nel focalizzare un determinato dettaglio, e di conseguenza, eventuali artefatti presenti in esso.

Un numero di regioni dell’immagine piuttosto contenuto significherà, quindi, che qualunque artefatto presente verrà immediatamente focalizzato e, conseguentemente, l’impatto qualitativo sull’osservatore ne sarà notevolmente influenzato.

Al contrario, la presenza di un gran numero di regioni in una scena, sarà indice, mediamente, di un maggior intervallo di tempo necessario per rilevare ed identificare un artefatto, con una percezione in generale di maggior qualità.

In sintesi, la procedura si compone di un primo passo in cui l’estrazione riguarda un gran numero di regioni di piccole dimensioni, in base a criteri locali di omogeneità. Quindi, si prosegue in modo iterativo accorpando regioni adiacenti identificate da caratteristiche soddisfacenti determinati criteri di similarità in termini di intensità.

Uno dei parametri richiesti dalla procedura applicata, a livello di elaborazione dei dati, riguarda proprio una stima del numero di regioni finali che viene utilizzato per fermare il processo di accorpamento. Nel caso specifico di analisi, siamo interessati a mantenere le regioni di segmentazione in un numero ragionevolmente limitato, effettuando una segmentazione piuttosto grezza ma in ogni caso funzionale ai nostri scopi.

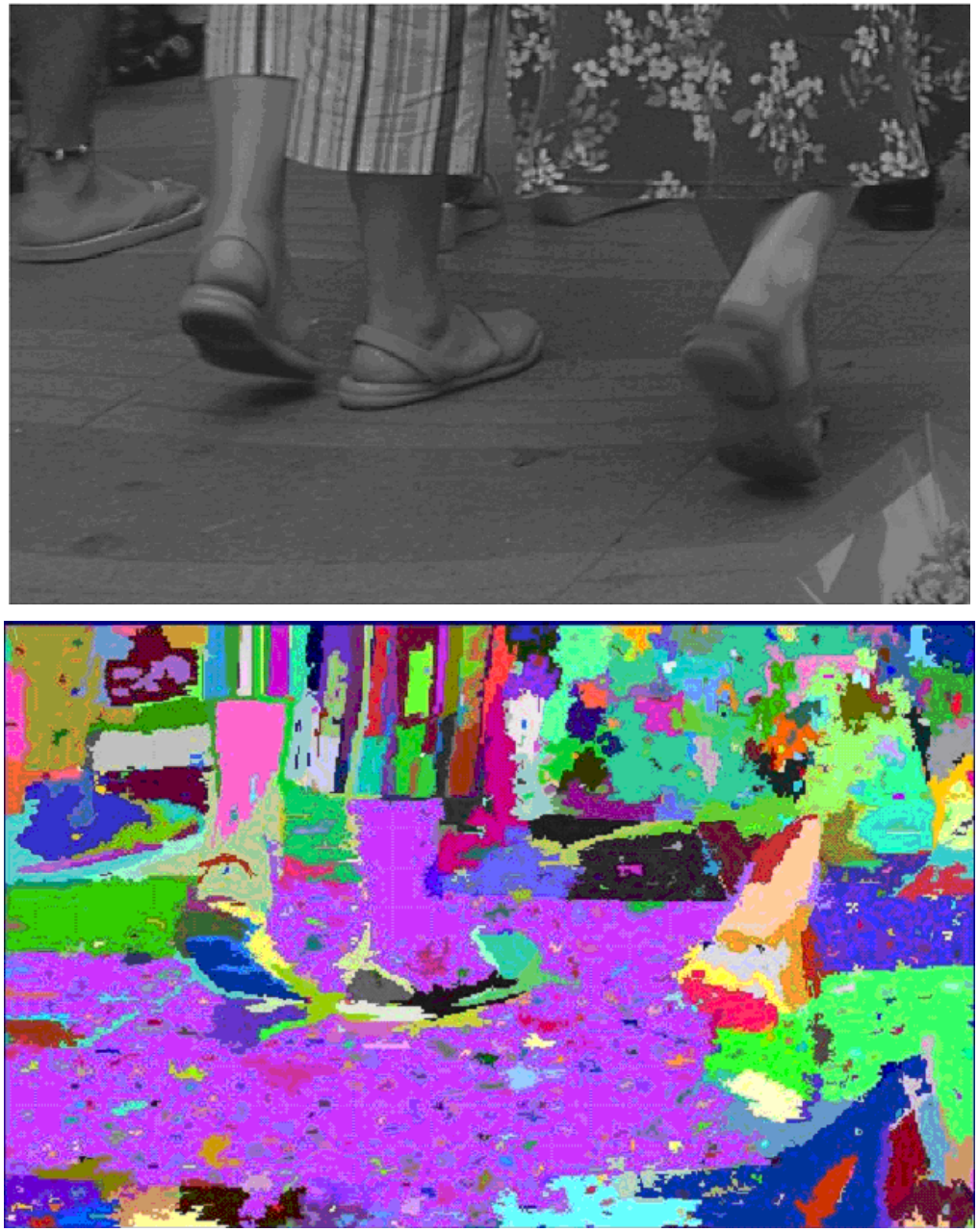

Figura 1: Esempi di segmentazione in un particolare di una sequenza di test

b) Il modello scelto per la rappresentazione del sistema visivo umano

Il metodo proposto per la selezione delle regioni di rilevanza percettiva si basa invece sul modello proposto da Koch e Ullman già negli anni Ottanta: il modello del sistema visivo da loro elaborato è stato in seguito ampliato ed è alla base di numerosi lavori come ad esempio [18] e [19].

La scelta di prendere a supporto un modello di questo tipo può essere interpretata come segue: a prescindere dalle regioni componenti l'immagine, è necessario individuare su quale fra queste cadrà in prevalenza l'attenzione dell'osservatore, proprio per individuare quale area dell'immagine, se affetta da artefatti, sarà determinante per la valutazione soggettiva della qualità. La presenza di sfocatura in questa area implicherà necessariamente un trattamento a livello di miglioramento del dettaglio presente, se affetta da sfocatura.

Il procedimento e la teoria di base vengono dettagliatamente descritti in [17] e [18], l'algoritmo implementato risulta molto complesso, per questo cerchiamo di riassumerlo in modo estremamente sintetico.

In primo luogo, diverse mappe di "*feature*" vengono estratte dall'immagine in analisi, divise in gruppi relativi a intensità, colore e orientazione. Conseguentemente, vengono accorpate con operazioni di fusione le mappe relative ad uno stesso gruppo di "*feature*" per arrivare ad una sola mappa di scalari, designata come mappa complessiva di salienza $S(i; j)$, in cui ad ogni pixel viene assegnato un valore corrispondente in termini di salienza, valutata nei termini delle caratteristiche selezionate inizialmente. Nell'algoritmo di Koch e Ullman, la zona di "*spot of attention*" ovvero quella immediatamente focalizzata dall'osservatore, viene quindi rilevata con la tecnica del "*winner-take-all*" a partire dal massimo della mappa di salienza, inibendo le zone adiacenti e ripetendo il procedimento per il secondo valor massimo rilevato.

Semplificando questa procedura, per ottenere una stima di quanto possa essere significativo per l'osservatore, l'estrazione delle regioni che noi consideriamo in "primo piano" viene quindi completata come segue, considerando una combinazione delle due diverse tipologie di analisi date dalla segmentazione e dalla mappa $S(i; j)$ su cui si basa il modello finora esposto.

In sintesi, decidiamo di considerare una regione, derivante dalla segmentazione, rilevante ovvero in "primo piano", se una percentuale prevalente dei pixel dell'area ad essa appartenente corrisponde a valori elevati della mappa di salienza [45].

Più precisamente si definiscono queste condizioni:

1. viene determinata una soglia di salienza S_{th} , definita come il valor medio valutato sull'intera mappa $S(i, j)$

2. in ogni regione X_k viene calcolata la percentuale di pixel dell'area $F_{VA,k}$ al di sopra S_{th}

3. per $F_{VA,k} > F_{VA,th}$, ovvero superiore ad una percentuale sufficientemente elevata e sperimentalmente scelta, la regione viene considerata percettivamente rilevante ed assegnata al "*foreground*", al "*background*" altrimenti.

I parametri di seguito specificati potranno essere valutati sull'immagine complessiva o separatamente su "*foreground*" e "*background*" in modo tale da poter distinguere determinate situazioni di sfocatura correlate a diverse origini, trattate diversamente in termini di valutazione della qualità ed opportunità di miglioramento della stessa.

Misurazioni a livello globale e relazione con le cause di sfocatura

Giunti a questo punto, cerchiamo di sintetizzare le misurazioni proposte a livello globale, in modo da collegare le considerazioni effettuate per rilevare la presenza di artefatto con la tipologia di sfocatura e la causa predominante della stessa nell'immagine in esame.

Per prima cosa, si tratta di completare la descrizione delle diverse tipologie di sfocatura di interesse per il presente lavoro, in modo tale che le caratteristiche imputabili alle diverse cause possano essere riconosciute e quantificate seguendo la traccia di analisi finora descritta e rappresentata dalle valutazioni oggettive e soggettive proposte [45].

Partendo dai parametri individuati a livello oggettivo, definiamo in termini di soglie le misurazioni della riduzione di contrasto, e con riferimento a quanto spiegato precedentemente, prendiamo in considerazione parametri così definiti:

- Fsm_1 con C_1 = (0%; 10%)
- Fsm_2 con C_2 = (10%; 30%)

La metrica finale utilizzata per la decisione della presenza di sfocatura appartenente alle categorie di interesse è data da due valutazioni, ovvero:

- il parametro Fsm_1 che rappresenta la percentuale di area probabilmente affetta da sfocatura molto accentuata, e verosimilmente dovuta ad acquisizione, come sarà ancor più evidente nel seguito;

- il rapporto Fsm_2/Fsm_1 il rapporto fra l'area contenente invece dettaglio definito e l'area stimata al passo precedente.

Queste valutazioni vanno associate alla metrica utilizzata per valutare la perdita di dettaglio *MGR_rel* che potrebbe invece corrispondere ad un indicatore della sfocatura introdotta dalla codifica.

Di seguito la descrizione dettagliata delle tipologie di sfocatura considerate in questa sede, per rendere ulteriore chiarezza nell'associazione fra tipologie di misurazione e cause di sfocatura.

a) Sfocatura da codifica

La codifica può introdurre sfocatura per effetto di una quantizzazione eccessiva e può attenuare fortemente il dettaglio fine ed i bordi contenuti, mentre contorni relativi a oggetti principali non ne risentono in modo eccessivo ma divengono meno accentuati ed aumentano di spessore. Conseguentemente, gli effetti di una codifica scadente impattano sicuramente sui valori del parametro *MGR_rel* che decresce all'aumentare della sfocatura introdotta ovvero al diminuire del contenuto informativo preservato dalla codifica, come emerge in modo significativo nei test eseguiti ed illustrati nel capitolo inerente i risultati sperimentali del metodo.

Il parametro Fsm_1 assume invece valori diversi principalmente a seconda del contenuto informativo originale del "*frame*" in esame, e può essere utilizzato come informazione propedeutica per distinguere casi in cui il dettaglio già in origine era presente in modo contenuto.

Un intervallo di interesse per i valori di tali parametri può essere individuato sperimentalmente in corrispondenza a questa particolare tipologia di artefatto.

b) Sfocatura nativa

Questa tipologia di sfocatura può essere imputata a cause intenzionali o accidentali, come un fuori fuoco dell'obbiettivo della telecamera, o ai limiti della strumentazione di acquisizione o di elaborazione, ed è comunque indipendente dalla codifica. Poiché tutte queste situazioni dipendono in realtà da caratteristiche proprie del video, appunto indipendenti dalla codifica effettuata, possono essere complessivamente assegnate alla categoria qui definita come "sfocatura nativa".

In tutti questi casi ritroviamo le caratteristiche seguenti: aree uniformi estremamente estese in termini percentuali, risultanti in valori elevati del parametro Fsm_1 e contemporaneamente valori molto bassi di *MGR_rel*, poiché non soltanto i contorni meno importanti ma gli stessi bordi principali degli oggetti presenti nella scena vengono attenuati in modo consistente dall'artefatto.

Importante è la distinzione fra il caso di sfocatura accidentale, da trattare in termini migliorativi di qualità, dal caso invece di sfocatura intenzionalmente introdotta dall'operatore esperto in una sequenza video: questo ultimo caso infatti non deve essere considerato come elemento di degrado qualitativo nell'immagine e non va trattato di conseguenza.

Vediamo esempi nei due casi.

La sfocatura da inadeguata acquisizione raggruppa in sé le situazioni di fuori fuoco della camera accidentalmente generate e le operazioni di *upscaling* di un video avente una risoluzione ed un contrasto molto limitati già in origine, ulteriormente accentuati poi dall'esigenza di visualizzazione su schermo ad alta definizione. Un esempio abbastanza tipico si può ottenere estraendo un "*frame*" da una qualsivoglia sequenza video scaricabile da YouTube, di risoluzione tipica pari a 320x240, e conseguentemente scalata a full HD. Ciò causa una decisa perdita di contrasto diffusa in tutta l'immagine. Come si potrà osservare nella sezione relativa ai risultati sperimentali, implementando una segmentazione in "*foreground*" e "*background*", secondo i principi decritti in precedenza, il parametro Fsm_2/Fsm_1 assume valori analoghi in entrambe le regioni individuate.

Contrariamente a quanto visto per la sfocatura da acquisizione, nel caso della sfocatura intenzionalmente introdotta in regioni di "*background*" si presenta un effetto caratteristico dato dalla elevata variazione del parametro dato dal rapporto Fsm_2/Fsm_1 valutato nel "*foreground*" e nel "*background*", con valori estremamente più elevati nel primo rispetto al secondo. Di fatto, Fsm_2 rappresenta la parte di area di una regione ad attività moderata e propria di oggetti della scena interessati da dettaglio fine, come volti o tessiture non affette da sfocatura. Conseguentemente, se il dettaglio in regioni dalle caratteristiche di questo tipo viene preservato in quanto "*foreground*", il rapporto Fsm_2/Fsm_1 è realmente molto più elevato che in regioni dello sfondo volutamente sfumato dal cineoperatore esperto in relazione ad un effetto voluto e non incidente sulla qualità dell'immagine così percepita dall'osservatore finale.

Al contrario, come già detto, sfocature che affliggono così gli oggetti in primo piano quanto lo sfondo, si manifestano in parametri di valore similare nelle due diverse aree dell'immagine individuate dai passi di segmentazione ed analisi di importanza percettiva.

Algoritmo di rilevazione a livello globale

Sintetizzando le precedenti considerazioni, diviene quindi possibile unificare le valutazioni provenienti dall'applicazioni delle metriche proposte, integrando con i modelli di percezione analizzati. È possibile giungere alla formulazione di una procedura complessiva per discriminare le cause e quantificare l'entità della sfocatura presente [45]:

1) Prima decisione sulla presenza e causa della sfocatura, da codifica o nativa

Un prima operazione avviene in termini di distinzione fra sfocatura nativa e sfocatura da codifica. Fra i parametri precedentemente definiti, vengono calcolati Fsm_1 e *MGR_rel* sull'intera immagine e rapportati a soglie sperimentalmente definite in modo tale da distinguere aree dell'immagine molto estese e caratterizzate da basso livello di contrasto, tipico di sfocatura nativa. Queste le condizioni verificate per riconoscere la presenza di sfocatura nativa:

a. $Fsm_1 > Fsm_{1th}$

b. $MGR\_rel < MGR\_rel_{th}$,

con Fsm_{1th} e MGR_rel_{th} soglie sperimentalmente determinate. Queste condizioni corrispondono ad una situazione di area molto estesa del "*frame*" con caratteristiche di uniformità e scarsità di dettaglio già in origine ed il "*frame*" viene riconosciuto affetto da sfocatura nativa e non da codifica; in tal caso si prosegue con la fase successiva. In caso contrario, si passa direttamente alla valutazione di un eventuale danno da codifica.

2) Distinzione fra sfocatura accidentale da acquisizione ed intenzionalità

Nel caso in cui una sfocatura nativa sia riconosciuta, è importante distinguere quale dei due casi menzionati si verifichi, in termini di azione efficace nel miglioramento della qualità da

introdurre eventualmente in fase di *"post processing"*. La sfocatura intenzionale di "*background"* preserva il dettaglio in primo piano mentre la sfocatura da acquisizione si estende in modo globale sull'immagine in esame. La segmentazione nelle regioni di "*foreground"* e "*background"* precedentemente descritta viene applicata in modo tale da poter valutare i parametri Fsm_1 e Fsm_2 separatamente nelle due aree di interesse. Se il rapporto Fsm_2/Fsm_1 varia in modo deciso fra la regione di "*foreground"* e quella di "*background"*, l'intenzionalità della sfocatura introdotta viene riconosciuta e non viene segnalato alcun deterioramento della qualità visiva, in caso contrario, l'artefatto viene riconosciuto come accidentale e uniformemente molto dannoso in termini di qualità percepita.

3) Misure di sfocatura da codifica

Se il "*frame"* di test non è affetto da sfocatura nativa, la presenza e l'intensità di artefatti dovuti a codifica scadente possono essere individuati sempre mediante le metriche proposte *MGR_rel* e Fsm_1. I valori del parametro *MGR_rel* possono essere considerati una funzione inversamente crescente rispetto alla sfocatura eventualmente presente. In particolare, valori elevati relativamente ad un intervallo sperimentalmente definito corrispondono a situazioni di codifica di buona qualità, mentre valori intermedi o bassi indicano la presenza di degrado nel dettaglio. Come precedentemente formulato, bassi valori di *MGR_rel* indicano essenzialmente assenza di dettaglio: tale può essere una proprietà intrinseca del "*frame"* oppure una conseguenza di una codifica non adatta. Per chiarire tale situazione, si ricorrerà all'utilizzo di altri parametri, ovvero nuovamente al valore valutato per Fsm_1, come meglio illustrato nel capitolo dedicato ai risultati sperimentali. Una volta riconosciuta la presenza di sfocatura da codifica, si potrà procedere alla stima più puntuale degli effetti dell'artefatto, non uniformi all'interno della scena e dipendenti in modo non trascurabile dal contenuto informativo della stessa, in modo da poter in un secondo momento agire in termini di miglioramento della qualità proprio a livello locale.

IV. STIMA LOCALE DELLA SFOCATURA DA CODIFICA

Come precedentemente menzionato, l'artefatto di sfocatura presenta caratteristiche peculiari a seconda della motivazione che ne determina la presenza. Nel precedente capitolo, identificando le diverse possibili cause, abbiamo posto in essere le condizioni per una più efficace determinazione dei parametri di rilevazione più opportuni, al fine di identificare la presenza dell'artefatto a partire dalla considerazione, con cognizione di causa, degli effetti ad esso attribuibili.

Abbiamo anche già visto come, in presenza di sfocatura da codifica, ne siano particolarmente affetti i bordi degli oggetti presenti nella scena e non particolarmente netti, così come le tessiture ed altri dettagli fini. Tessiture e dettaglio fine corrispondono infatti a rapide variazioni di intensità ma di ampiezza contenuta, per questo la trasformata DCT e la quantizzazione successiva propria della codifica penalizzano soprattutto tali frequenze più elevate. Proprio per tali motivi, la tecnica proposta per la stima locale della sfocatura si basa sull'analisi dei bordi e del dettaglio nella componente di luminanza dell'immagine.

Supponendo che questo particolare tipo di informazione dell'immagine sia soggetto all' artefatto da codifica, i risultati ottenuti sperimentalmente dal metodo descritto nel seguito, mostreranno gli effetti delle diverse situazioni di codifica in termini di sfocatura a livello locale, e specialmente sulle tessiture e sul dettaglio fine, e come l'assenza di tale dettaglio possa essere un indizio del danno di sfocatura introdotto.

Vediamo ora in dettaglio il metodo proposto e pubblicato in [13].

Partendo da un'immagine da analizzare di dimensione *mxn*, dividiamo l'area della stessa in blocchi di dimensione *pxq*, ottenendo come risultato *M* blocchi in direzione orizzontale ed *N* blocchi in direzione verticale, che indichiamo come $X_1,\ldots, X_{MN}$. I blocchi dell'immagine così ottenuta vengono sottoposti all'operatore già precedentemente descritto, ovvero la diffusione anisotropa, proprio per riuscire a dare una prima stima, a livello locale, dell'entità del contenuto soggetto a sfocatura, in ciascuno dei blocchi definiti. In diverse zone dell'immagine l'effetto può essere diversamente visibile ed è proprio questo il nostro obiettivo, ovvero individuare le zone del "*frame*" in esame maggiormente soggette e danneggiate.

Come già illustrato, la diffusione anisotropa può essere una valida simulazione del danno introdotto dalla codifica, proprio in quanto agisce in modo analogo sul dettaglio dell'immagine: ora l'applicazione avviene in senso locale per ciascuno dei blocchi in cui l'immagine è stata suddivisa. Con $X'_{1,}$ ..., X'_{MN} indichiamo i blocchi sottoposti all'operatore di diffusione anisotropa, quindi valutiamo il valor medio del gradiente morfologico m_k per ogni blocco X_k. Tale gradiente morfologico nel pixel (i_0,j_0) viene calcolato secondo le formule:

$$m_k(i_0,j_0) = \max M_k(i_0,j_0) - \min M_k(i_0,j_0)$$

$$M_k(i_0,j_0) = \{ X_k(i_0 + i, j_0 + j) \mid -1 < i, j < +1 \}$$

A questo punto, vengono ripetute le stesse elaborazioni per valutare la media del gradiente morfologico m'_k al fine di valutare su ogni blocco l'indice seguente:

$$MGR_k = m_k / m'_k$$

Il valore dell'indice così definito dipende dalla significatività del dettaglio non più presente nell'immagine a valle della diffusione anisotropa, ovvero dall'entità della sfocatura presente nel caso locale sul blocco in esame: infatti il dettaglio fine, se presente, sarà cancellato dalla diffusione anisotropa, in modo tale da determinare un rapporto elevato fra i gradienti medi valutati a monte e a valle dell'applicazione dell'operatore. Questo sarà il caso di un blocco ben codificato e con dettaglio ben preservato. Al contrario, blocchi codificati con notevole perdita di informazione e conseguentemente di dettaglio, non subiranno alterazioni significative del contenuto dopo l'applicazione dell'operatore e presenteranno di conseguenza un valore del parametro MGR_k molto prossimo all'unità.

In sintesi possiamo affermare che la qualità di codifica del singolo blocco, in modo sommario, può essere descritta dal parametro appena definito: c'è ancora da considerare, infatti la dipendenza dal contenuto informativo peculiare dell'area in analisi. A livello globale, è possibile liberarsi da tale

dipendenza effettuando una media dei valori del parametro sui blocchi dell'immagine ed ottenendo il valore *MGR_av*: in ogni caso, un valore del parametro appartenente ad un intervallo sperimentalmente determinato può indicare la necessità di intervenire con operatori migliorativi della qualità in termini di sfocatura presente.

Le soglie, presentate nella sezione relativa alla parte sperimentale, vengono definite come [*MGR_low*, *MGR_high*] e determinano l'intervallo utile di intervento per quanto riguarda processi di miglioramento della qualità; valori del parametro superiori ad *MGR_high* sono infatti corrispondenti a situazioni di codifica adeguata, mentre valori del parametro inferiori a *MGR_low* rappresentano situazioni di codifica talmente scadente da richiedere un intervento più radicale nella scelta delle opzioni in fase di trasmissione.

Questa prima valutazione rappresenta appunto una stima grezza, da affinare identificando bordi e zone di rilevanza percettiva nei blocchi dell'immagine selezionati da questo primo passo.

Misurazione oggettiva a livello locale

Procediamo quindi in direzione di una maggiore precisione a livello locale e di una quantificazione oggettiva della sfocatura sui bordi degli oggetti di interesse identificati nel singolo blocco. Una prima fase in tal senso riguarda la classificazione degli stessi contorni presenti nel blocco e l'individuazione delle aree interessate da tessitura. Tale classificazione viene effettuata come segue.

All'interno del blocco, i pixel vengono suddivisi in tre possibili categorie, in dipendenza del valore precedentemente valutato per il gradiente morfologico dopo la diffusione anisotropa:

- regione A_k: include pixel $X_k\ (i, j)$ corrispondenti a valori di gradiente morfologico che soddisfano la condizione:

$$m'_k(i,j) > m'_k + \Delta$$

dove Δ è un valore positivo e sperimentalmente scelto in modo tale che a questa condizione corrispondano contorni di oggetti non affetti dall'artefatto in analisi.

- regione B_k: include pixel $X_k\ (i, j)$ corrispondenti a valori di gradiente morfologico che soddisfano la condizione:

$$m'_k < m'_k(i,j) < m'_k + \Delta$$

e supponiamo che questa regione contenga pixel soggetti a sfocatura in quanto valori intermedi di gradiente potenzialmente soggetti all'effetto della codifica.

- regione C_k: include pixel $X_k\ (i, j)$ corrispondenti a valori di gradiente morfologico che soddisfano la condizione:

$$m'_k(i,j) < m'_k$$

tale regione comprende tessiture e ed aree sfumate o a dettaglio fine.

L'estensione della regione composta da pixel con valori intermedi del gradiente, quindi potenzialmente affetti da sfocatura, viene messa in relazione alla presenza complessiva dei contorni nella scena, in modo da quantificare in termini percentuali l'impatto della codifica stessa sul contenuto informativo. La regione definita come B_k viene confrontata in termini di estensione all'unione delle regioni B_k e A_k, comprendenti la totalità dei contorni degli oggetti presenti, in modo tale che i bordi soggetti a sfocatura vengano rapportati in termini di occupazione percentuale ai bordi complessivamente presenti. Ne consegue un rapporto correlato alla sfocatura riconducibile ai bordi degradati, definito come segue per ogni blocco X_k:

$$DEP_k = dimensione(B_k)/dimensione\ (A_k\ U\ B_k),\ k=1,..,MN$$

Valori bassi di tale parametro corrispondono a qualità elevata, ovvero a una bassa percentuale di bordi degradati sul totale. Viceversa, un valore elevato di DEP_k corrisponde ad una situazione di esteso degrado, sicuramente di impatto percettivo. Sintetizzando l'informazione contenuta nei parametri appena definiti, si può determinare un parametro comprensivo di tutte le informazioni del caso e dato dal rapporto:

$$BE_k = MGR_k / DEP_k\ ,\ k=1,..,MN$$

Il valore di tale indice aumenta in modo sensibile all'aumentare della qualità del blocco in termini di sfocatura presente e rappresenta un criterio oggettivo per discriminare blocchi degradati nel contenuto a causa della codifica.

A questo punto, il parametro così definito a livello oggettivo, va integrato con valutazioni di tipo soggettivo basate su modelli del sistema visivo umano e con considerazioni di tipo cognitivo riguardo all'impatto percettivo degli oggetti presenti sulla scena: vi sono infatti particolari situazioni, come ad esempio la presenza di volti, che sono estremamente critiche in termini di giudizio qualitativo da parte di un osservatore umano, molto più sensibile ad artefatti presenti in tali casi.

Identificazione di eventi di rilevanza cognitiva

Le performance dell'algoritmo sicuramente possono essere migliorate identificando contenuti importanti per il nostro sistema visivo e cognitivo. Questi contenuti possono variare con il soggetto delle immagini e con le aspettative dell'osservatore, tuttavia se ne possono determinare con chiarezza alcune categorie: fra queste, a titolo esemplificativo, verrà considerata come evento significativo la presenza di volti sulla scena. Integrare tale evento nella valutazione della qualità rende sicuramente una maggiore coerenza con quanto espresso in termini di giudizio soggettivo, come già dimostrato in [31].

La rilevazione di volti in un'immagine diviene peraltro un argomento affrontato in molti contesti e ciò consente di affrontare il problema con una buona base di partenza ed una serie di considerazioni preliminari già accettate dalla comunità scientifica.

Una di esse riguarda il colore: la pelle umana è caratterizzata da valori di crominanza compresi in un intervallo piuttosto limitato [32], la loro distribuzione si presenta piuttosto stabile [34] e molti algoritmi si basano su questi assunti, dimostrandosi efficaci. Un secondo tipo di considerazione che è possibile fare, riguarda invece il livello di attività presente in un'area dell'immagine occupata da un volto, caratterizzato da un elevato livello di complessità e da un gran numero di regioni: bordi netti in corrispondenza dei tratti somatici si alternano a zone ad elevata correlazione ed a basso livello di dettaglio.

Questi principi possono essere utilizzati nel lavoro attuale per rilevare la presenza di volti nel blocco X_k in analisi.

Come primo passo, regioni distinte vengono identificate nel blocco ed estratte grazie all'applicazione dell'algoritmo di segmentazione proposto nei precedenti capitoli. Blocchi corrispondenti ad un numero troppo basso di regioni non vengono presi in considerazione, proprio perché la complessità dell'oggetto rappresentato da un volto corrisponde ad un elevato grado di dettaglio e di bordi, ovvero un numero piuttosto elevato di regioni in uscita dall'algoritmo di segmentazione. Per i blocchi che soddisfano questo primo criterio viene quindi effettuata un'analisi secondo le considerazioni fatte sulle caratteristiche di colore e viene tracciata una mappa corrispondente a pixel di determinati valori di crominanza. In tali zone viene valutata l'attività presente, sempre tramite la valutazione del gradiente morfologico e la classificazione in regioni secondo i criteri già visti.

Le regioni di dettaglio fine sono caratterizzate dalla condizione di gradiente morfologico pari a quanto già visto per la condizione su C_k:

$$m'_k(i,j) < m'_k$$

mentre la misura di attività viene data da un indice descritto in [37] e calcolato per ogni pixel (i_0, j_0) nel dato intervallo di crominanza:

$$MAG(i_0, j_0) = \sum_R |I(i,j) - I(i_0, j_0)| / (N-1)$$

con $I(i, j)$ che denota il valore di intensità del pixel alle coordinate (i, j), e nella somma consideriamo gli N pixel appartenenti all'intorno prescelto R e centrato in (i_0, j_0).

Il valore dell'indice *MAG* viene calcolato per ogni pixel del blocco nell'intervallo di crominanza, la media viene valutata sulla regione corrispondente a basso valore di dettaglio risultante nel parametro MAG_av1_k, poi ricalcolando il valor medio sull'intera regione di crominanza appropriata e risultante nel valor medio MAG_av2_k. Il blocco X_k contiene un volto se sono soddisfatte le seguenti condizioni:

1. $MAG\_av2_k > N_{fd} \cdot MAG\_av1_k$
2. $MAG\_av1_k < FD_{th}$

con FD_{th} soglia sperimentalmente definita ed N_{fd} fattore moltiplicativo sperimentale. Tali condizioni traducono formalmente quanto già considerato in merito alle caratteristiche della situazione da identificare, ovvero una differenza notevole di gradiente, presente su bordi significativi propri dei volti, in raffronto a valori di bassa attività in zone di dettaglio fine. Se tali condizioni sono entrambe soddisfatte, viene settato ad 1 un flag che denominiamo FD_k e che viene conseguentemente definito per ogni blocco in analisi, appunto ad indicare la presenza di un volto. In caso di assenza di volti, ovviamente, tale flag è posto a 0.

Sintesi dei parametri di qualità locale

A questo punto, diviene necessario sintetizzare le informazioni a disposizione finora, date dai parametri definiti e corrispondenti alle diverse situazioni determinanti per la valutazione di qualità a livello locale di blocco.

Si tratta di definire un set di parametri, rilevanti per la valutazione della qualità del blocco in esame, con il vantaggio di non richiedere alcun tipo di riferimento per effettuare tale stima. Per ogni blocco selezionato come sospetto in termini di degrado di qualità, in base alla prima stima grezza data dal parametro MGR_k possiamo estrarre le seguenti informazioni parametriche:

- a partire dal modello prescelto per descrivere il sistema visivo umano [17][18], possiamo tracciare una mappa di salienza in termini di focalizzazione dell'attenzione. Possiamo decidere se analizzare o meno un blocco a seconda della sua importanza percettiva, stimata come percentuale della sua stessa area che appare significativa secondo tale mappatura [13]. Di conseguenza, il primo parametro da considerare è un flag, indicato con RB_k e settato ad 1 se una sufficiente percentuale di pixel del blocco corrisponde a valori di salienza oltre la media valutata sulla mappa di salienza;

- a partire dalla misurazione oggettiva della sfocatura presente, selezioniamo come significativo il parametro BE_k, che dipende da MGR_k e dalla percentuale di bordi degradati rispetto al totale dei bordi presenti nel blocco;

- altra informazione determinante consiste nel numero di regioni presenti nel blocco in analisi, calcolate in uscita dall'algoritmo di segmentazione selezionato. Il numero di regioni fornisce infatti una stima della complessità della scena, ovvero della rapidità di identificazione degli artefatti presenti all'interno della stessa da parte di un osservatore. Tale numero di regioni viene memorizzato nel parametro che definiamo come NoR_k e sempre associato al blocco in analisi;

- la presenza di volti come situazione cognitivamente importante, data dal flag FD_k appena definito.

Queste informazioni possono essere utilizzate in termini di decisione, sia per quanto riguarda la stima della qualità sia nel senso della definizione dei processi migliorativi opportuni a livello locale. A partire dai quattro parametri, RB_k , BE_k, NoR_k e FD_k siamo in grado di tracciare una mappa dell'incidenza della sfocatura in termini percettivi, guidando un'eventuale azione adattativa di miglioramento.

Sicuramente si tratterà di intervenire in presenza di volti degradati, con elevate percentuali di bordi potenzialmente sfocati presenti in termini di misurazione oggettiva. Così come, in presenza di blocchi percettivamente importanti, si può decidere di intervenire soprattutto in situazioni di un numero di regioni di segmentazione contenuto, proprio perché in tali casi gli artefatti presenti vengono rapidamente focalizzati dall'osservatore. Oppure, anche in corrispondenza di un numero di regioni maggiormente elevato, si può decidere di effettuare o meno procedure migliorative di qualità a seconda della stima del degrado dei contorni presenti nel blocco.

Tuttavia si pone ora il problema di unificare tali informazioni, per ovvi motivi di praticità decisionale: i flag definiti possono comunque essere utilizzati solo nella prima fase di selezione dei blocchi di interesse, prendendo in considerazione blocchi che abbiano settato ad 1 almeno uno dei due flag. Ciò che ora ci prefiggiamo è condensare le informazioni riconducibili al numero di regioni ed al degrado dei contorni, anche nell'ottica di un confronto qualitativo con stime soggettive. Cerchiamo quindi di sintetizzare in un solo parametro I_k quanto ottenuto tramite il nostro studio per il blocco X_k:

$$I_k = \alpha \cdot log(BE_k) + \beta \cdot log(NoR_k) + \gamma$$

Dove:

$$(\alpha, \beta, \gamma) = f(MGR\_av)$$

dove questa espressione deriva dall'assunto, giustificato dalla precedenti considerazioni, che la qualità percepita aumenti in modo monotono con il risultato della misurazione oggettiva e con il numero di regioni. Il logaritmo viene utilizzato per tenere conto di fenomeni di saturazione di giudizi

soggettivi per valori molto elevati di qualità così come un offset viene introdotto per riprodurre in modo più fedele la valutazione soggettiva.

Tuttavia, l'influenza di queste due quantità, misurazione oggettiva e numero di regioni, varia a seconda del livello di qualità percepito: se il blocco è pesantemente degradato, tale effetto difficilmente può essere mascherato dal fatto che il numero di regioni segmentate è elevato; d'altro canto, in casi di qualità elevata, la percezione è buona indipendentemente dal numero di regioni provenienti dalla segmentazione del blocco. La valutazione proposta si dimostra efficace quindi soprattutto in casi di qualità percepita di livello intermedio, in modo tale da poter supporre pesi dello stesso ordine di grandezza. Coerentemente con quanto appena detto, un diverso set di coefficienti va utilizzato a seconda della stima iniziale basata sul valore del parametro MGR_k globalmente mediato sull'immagine e risultante in *MGR_av*. Il set di coefficienti viene determinato in fase sperimentale, in modo da fornire valori di qualità in uscita dall'algoritmo massimamente coerenti alle valutazioni degli osservatori umani interrogati come campione. I dettagli vengono forniti nel seguito e in [13], mediante un confronto fra valutazione soggettiva e stima di qualità ottenuta mediante l'algoritmo proposto, sulla base di un campione sufficientemente esteso in termini di contenuto informativo delle immagini e di codifiche esaminate.

V. RISULTATI SPERIMENTALI

Misurazioni a livello globale e rilevazione delle cause

I dati video usati per gli esperimenti vengono da diverse fonti, hanno contenuti molto variabili e sono stati sottoposti a diverse operazioni di codifica e di manipolazione [45].

I dati originali sono di risoluzione 1080p o inferiore, scalati a risoluzione pari a Full-HD (1920x1080p) o similari, come meglio spiegato nel seguito, usando funzioni Matlab o *"scaler"* gentilmente concessi in uso per questo studio da Philips Consumer Electronics.

Aumentando il fattore di scala, proporzionalmente diminuisce l'ampiezza dell'intervallo di frequenze occupato dal segnale video. In altre parole, l'immagine appare più sfocata. L'operazione di scalaggio è tipicamente necessaria quando un segnale SD o HD deve essere visualizzato su uno schermo Full-HD, ma anche se un segnale 4:3 viene visualizzato su uno schermo 16:9, o ancora un segnale 16:9 su uno schermo 21:9. nel caso di schermi in *'panorama mode'* viene applicato lo scalaggio non lineare con fattori variabili in senso orizzontale in modo parabolico, in modo tale che il fattore di scala è ai bordi dello schermo pari al doppio di quanto applicato nella parte centrale, presentando di conseguenza gradi diversi di sfocatura.

Figura 2: Frame 'Shuttle920'

Figura 3: Esempio di immagine della sequenza 'Football'

L'algoritmo viene applicato con parametri sperimentalmente assegnati come segue: $F_{VA,th}$= 60% per la segmentazione del *"foreground"*, Fsm_{1th} = 60%, e MGR_rel_{th} = 0.2 per la rilevazione della sfocatura nativa.

L'efficacia della metrica proposta nella stima della sfocatura viene analizzata in due fasi. Per prima cosa i *"frame"* vengono raggruppati secondo criteri oggettivi, come ad esempio la presenza di sfocatura e la sua origine, esaminando la corrispondenza fra il valore calcolato della metrica e la categoria oggettiva del *"frame"* in esame. Successivamente, viene studiata la relazione fra metrica e stima complessiva di qualità, confrontata con giudizi di osservatori umani, sempre in relazione alla tipologia di sfocatura corrispondente. I risultati ottenuti vengono presentati nel seguito.

a) Distinzione delle cause di sfocatura

Ogni "*frame*" analizzato, viene assegnato a una delle categorie identificate come segue, a seconda delle caratteristiche di sfocatura:

1) Immagini ben codificate e ricche di dettaglio:
   Gli esempi riportati riguardano immagini ricche di dettaglio già in origine, in particolare si tratta dei "*frame*" descritti nel seguito ed appartenenti alle sequenze "Barcelona" in corrispondenza alle codifiche di buona qualità e "Shuttle" ad eccezione di 'Shuttle810'.

2) Video affetti da sfocatura da codifica:
   Gli esempi riportati riguardano i "*frame*" descritti nel seguito ed appartenenti alla sequenza "Barcelona" in corrispondenza alle codifiche di qualità intermedia o scadente.

3) Video correttamente codificati ma con sfocatura intenzionalmente introdotta in "*background*", come di seguito descritto in merito al materiale di test.

4) "*Frame*" correttamente codificati ma con sfocatura nativa introdotta ad esempio da operazioni di scalaggio a risoluzione di molto superiore: un esempio può essere un "*frame*" a bassa risoluzione, proveniente ad es. da YouTube e scalato a full HD.

5) Video codificati correttamente ma con estese aree soggette a movimenti rapidi: un esempio può essere un "*frame*" contenente immagini di una gara sportiva, come in Fig. 3.

La relazione che sussiste fra tipologia di "*frame*" ed i valori delle metriche viene analizzata in due grafici distinti. Il primo mostra la coppia di metriche (Fsm_1, *MGR_rel*) corrispondente ad ogni "*frame*" rappresentato da un punto del piano. Diversi colori corrispondono alle diverse situazioni rappresentate: verde per immagini ben codificate e dettagliate (Tipo 1), rosso per sfocatura da codifica (Tipo 2), blu per "*frame*" affetti da sfocatura nativa o intenzionale (Tipi 3 e 4), magenta per "*frame*" contenenti moto intenso.

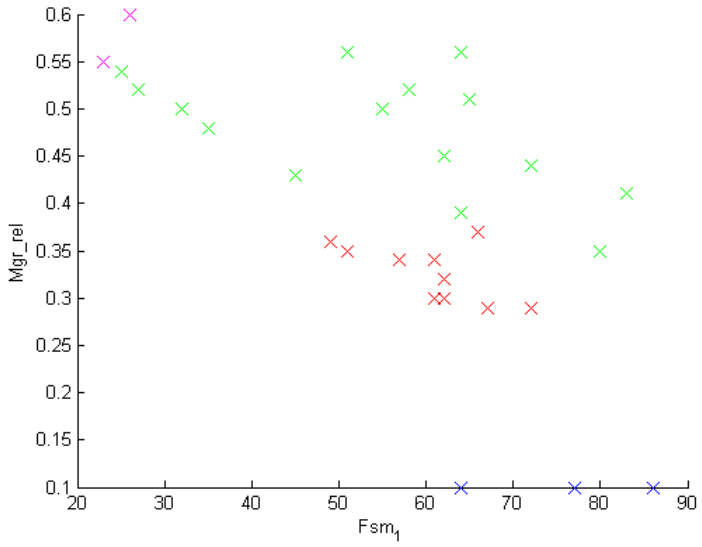

Figura 4: Valori dei parametri per immagini di test affette da diverse cause di sfocatura (in colori diversi)

Una corrispondenza fra le caratteristiche del *"frame"* e la posizione dello stesso nel piano delle metriche può essere facilmente osservata, provando l'efficacia della misurazione scelta. Più precisamente:

- Un buon livello di qualità corrisponde ad elevati valori della metrica *MGR_rel*, o a valori intermedi qualora il parametro Fsm_1 sia elevato ed originariamente corrispondente a scarso dettaglio in origine come nel caso del frame 'Shuttle920'.

- Valori di metrica corrispondenti a sfocatura da codifica costituiscono un raggruppamento compatto nella parte centrale del piano.

- *"Frame"* propriamente codificati contenenti sfocatura intenzionalmente introdotta o accidentale dovuta ad acquisizione o ad operazioni di scalaggio (Tipi 3 e 4) sono contraddistinti da valori molto bassi del parametro *MGR_ rel*.

- "*Frame"* interessati da movimenti rapidi su zone estese (Tipo 5) hanno metriche associate a zone del piano corrispondenti a buona qualità.

Il secondo grafico permette di effettuare una distinzione fra i due casi di sfocatura nativa non causata da codifica, indicati con i nomi di sfocatura intenzionale (Tipo 3) e sfocatura di acquisizione/ scalaggio (Tipo 4), entrambi caratterizzati da valori molto bassi per la metrica *MGR_rel*. Fsm_1 e Fsm_2 vengono calcolati sia su "*foreground*" che su "*background*" per ciascuno dei "*frame*" di test, separatamente. Nel grafico, le coppie (Fsm_1, Fsm_2/Fsm_1) corrispondenti a "*foreground*" e "*background*" per ogni "*frame*" sono rappresentate in colori diversi. Per i due "*frame*" di test di tipo 3, riportati graficamente in colori blu e ciano, il valore del rapporto Fsm_2/Fsm_1 calcolato nel "*foreground*" è il doppio di quanto valutato per il "*background*". Invece, per il test di tipo 4 (magenta), caratterizzato da una estesa sfocatura sia in zone di "*foreground*" che di "*background*", il rapporto Fsm_2/Fsm_1 assume valori molto simili se calcolato separatamente sulle due regioni.

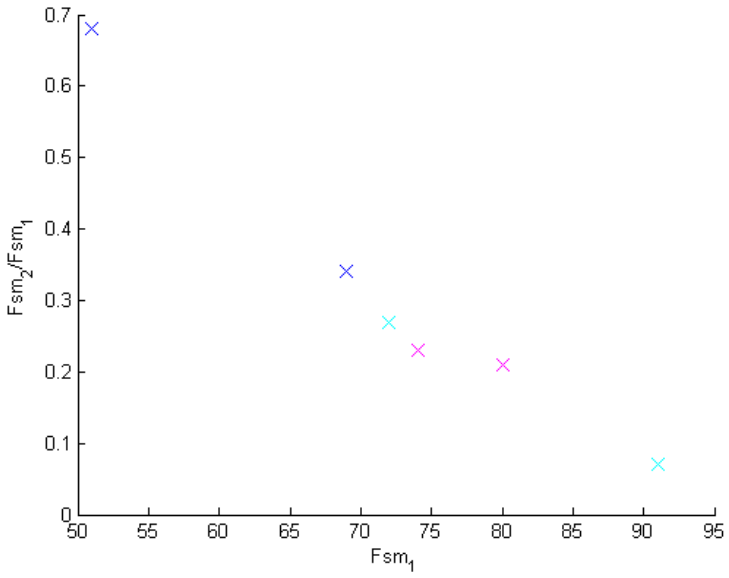

Figura 5: Metriche valutate su foreground e background per sfocatura nativa: immagini Tipo 4 in magenta, Tipo 3 ciano e blu.

b) Valutazione della qualità soggettiva

La corrispondenza fra la metrica proposta e la qualità percepita viene valutata attraverso una serie di test fatti con la collaborazione di osservatori umani. Viene usato un metodo coerente con quanto definito in ITU-R BT.500, mostrando l'immagine all'osservatore in un unico e breve intervallo di tempo ai fini del giudizio [20]. Ogni "*frame*" di test viene mostrato separatamente, senza alcun riferimento di merito, e valutato secondo il seguente punteggio di qualità: 5. Eccellente 4. buono 3. intermedio 2. scadente 1. pessimo. Il test comprende 30 "*frame*" per i quali i valori delle metriche oggettive sono già stati precedentemente valutati secondo l'algoritmo in esame. Ogni "*frame*" viene presentato per 3 secondi, scaduti i quali all'osservatore viene richiesto di formulare un giudizio, senza più avere di fronte il "*frame*" di test bensì uno schermo grigio.

Per contenere errori dovuti a distrazione o incertezza dell'operatore, l'intero set di test viene sottoposto alla persona per due volte di seguito, in ordine casuale. Il test viene sottoposto a dieci soggetti, sia maschi che femmine, senza alcuna esperienza di artefatti video.

Gli osservatori sono istruiti in modo da valutare la qualità di ogni "*frame*" di test in accordo con il grado di sfocatura, sottolineando però che determinate situazioni corrispondono a sfocatura volutamente introdotta che come tale non va considerata come artefatto. Vengono quindi mostrati loro esempi di immagini di ottima e pessima qualità, in ordine di calibrazione del giudizio da formulare. Dopo il test, ad ogni "*frame*" viene assegnato un punteggio, con numero di serie $Nser = 2$ per numero di osservatori pari a $Nobs = 10$ osservatori. Nella valutazione del Mean Opinion Score (*MOS*) su ogni "*frame*", sarebbe auspicabile dare meno rilevanza ai giudizi sui quali l'osservatore ha dimostrato una maggiore incertezza.

Quindi, valutata la media, viene dato minor peso a coppie di punteggi provenienti da uno stesso osservatore ma differenti nelle due serie di test.

Formalmente, la definizione è come segue, con

$$MOS_k = \left(\sum_{o=1,..,Nobs} \sum_{r=1,..,Nser} S_{k,o,r} \cdot p_{k,o} \right) / \left(Nser \cdot \sum_{o=1,..,Nobs} p_{k,o} \right)$$

dove in relazione al punteggio $S_{k,o,r}$ dato al "*frame*" k dall'osservatore o alla serie $Nser = 2$, i parametri

assumono i seguenti valori:

- $p_{k,o} = 1$ se $S_{k,o,1} = S_{k,o,2}$
- $p_{k,o} = 0.75$ se $|S_{k,o,1} - S_{k,o,2}| = 1$
- $p_{k,o} = 0$ se $|S_{k,o,1} - S_{k,o,2}| > 1$

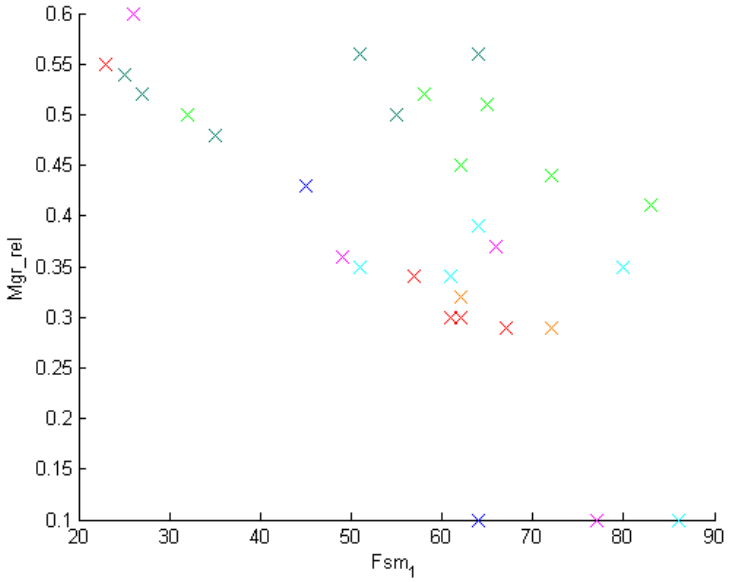

Figura 6: Mean Opinion Score per le immagini di test, marcature dello stesso colore corrispondono allo stesso intervallo di valori

Per osservare la relazione fra le due metriche, date dai parametri *Mgr_rel* e Fsm_1, e il *MOS*, coppie di metriche corrispondenti a "*frame*" diversi sono rappresentate come *Mgr_rel* vs Fsm_1 con tratto in colore dipendente dal valore del *MOS* del "*frame*". La corrispondenza colore-*MOS* è la seguente:

- 'verde scuro' per MOS > 4.5
- 'verde chiaro' per 4 < MOS < 4.5
- 'ciano' per 3.5 < MOS < 4
- 'blu' per 3 < MOS < 3.5
- 'magenta' per 2 < MOS < 3

- ’rosso’ per 1.5 < MOS < 2
- ’arancio’ per 1 < MOS < 1.5

Una corrispondenza può essere osservata fra coppie di metriche e locazione nel piano, in cui a diverse zone viene associata una diversa percezione di qualità, così come un secondo tipo di associazione può essere fatto nei confronti della tipologia di sfocatura. In dettaglio:

- *“Frame”* che corrispondono a punteggi da 4 a 5 hanno valori distribuiti in zone del piano caratterizzate da elevato *MGR_rel*, indice di abbondanza di dettaglio, sempre associabile a buona qualità percepita ed in particolar modo nei casi in cui il valore di Fsm_1 non è particolarmente basso. Basso valore di *MGR_rel* è maggiormente tollerato infatti in *“frame”* che originariamente sono costituiti da estese aree uniformi che si manifestano come elevati valori di Fsm_1. questo è ad esempio il caso del *“frame”* ’Shuttle920’, la cui qualità viene giudicata intermedia sebbene il valore di *MGR_rel* sia piuttosto basso, poiché spesso ampie zone uniformi possono far percepire una lieve presenza di sfocatura.

- Tutti i *“frame”* considerati di bassa o pessima qualità, con punteggi da 1 a 3, corrispondono a valori di metriche della parte centrale del piano. Ciò implica che valori medio-bassi di *MGR_rel*, per valori di Fsm_1 non troppo elevati - quindi la scarsità di dettaglio senza casi di estese aree uniformi – possono sempre essere identificati come *“frame”* a qualità percettivamente degradata. La considerazione inversa è ugualmente valida con poche eccezioni, ovvero *“frame”* corrispondenti alla parte centrale del piano ottengono dagli osservatori un giudizio che nel migliore dei casi si colloca a livello intermedio di qualità.

- Due *“frame”* sono stati valutati di bassa qualità seppur corrispondenti a valori di metrica associati a zone del piano caratterizzate da alta qualità e tali *“frame”* sono identificati da aree estese in movimento rapido. La metrica rileva la presenza di dettaglio nelle zone non interessate dal moto ed interpreta la situazione percettiva coerentemente con quella che è la percezione della visione poi complessiva della scena: di fatto all’osservatore viene sottoposto il singolo *“frame”*.

- *“Frame”* sfocati da operazioni indipendenti dalla codifica corrispondono alle situazioni percettivamente più fastidiose per gli osservatori, mentre la sfocatura intenzionalmente introdotta corrisponde in ogni caso ad una qualità comunque intermedia. Questo prova l’utilità del metodo e l’abilità nel distinguere le due diverse situazioni (immagine di Tipo 4, giudicata ’scadente’, in magenta; immagini di Tipo 3, giudicate ’intermedie’, in blu e ciano).

c) Descrizione del materiale di test

La prima sequenza denominata “Barcelona” è già in origine in alta qualità video 1920x1080 e gentilmente fornita da Philips Consumer Electronics. Questa sequenza viene utilizzata per analizzare gli effetti di codifica e scalaggio. Per prima cosa, cinque codifiche H264 vengono applicate con un diverso set di parametri come segue:

- 1Pass intermediate (1Pint): bassa qualità e medio bitrate (1000 kbps).
- iPod: bitrate molto basso e bassa qualità (600 kbps).
- Common Encoding (CE): qualità intermedia e parametri standard (1000 kbps).
- BlueRay (BR): bitrate e qualità elevata (8000 kbps).
- Constant Quantization (CQ): step di quantizzazione costante e qualità molto elevata.

Lo stesso *“frame”* viene estratto dalla sequenza e suddiviso in quattro quadranti, ciascuno dei quali viene riportato ad alta qualità tramite scalaggio in *'panorama mode'*, che introduce maggior sfocatura lateralmente rispetto alle regioni centrali.

Figura 7: Frame originale della sequenza 'Barcelona'

Figura 8: Frame con codifica 1 Pass Intermediate

Figura 9: Frame con codifica iPod

Figura 10: Frame con codifica Common Encoding

Figura 11: Frame con codifica BlueRay

Figura 12: Frame con codifica Constant Quantization

La seconda sequenza è denominata “Shuttle”, scaricabile a fini di ricerca da [47]: rappresenta scene di contenuto vario nei nove *“frame”* estratti alla risoluzione (1280x720), codificati e usati con fattore di scala pari a 1.5.

Figura 13: Frame 3 della sequenza 'Shuttle'

Figura 14: Frame 170 della sequenza 'Shuttle'

Figura 15: Frame 280 della sequenza 'Shuttle'

Figura 16: Frame 360 della sequenza 'Shuttle'

Figura 17: Frame 560 della sequenza 'Shuttle'

Figura 18: Frame 810 della sequenza 'Shuttle'

Figura 19: Frame 920 della sequenza 'Shuttle'

Figura 20: Frame 1050 della sequenza 'Shuttle'

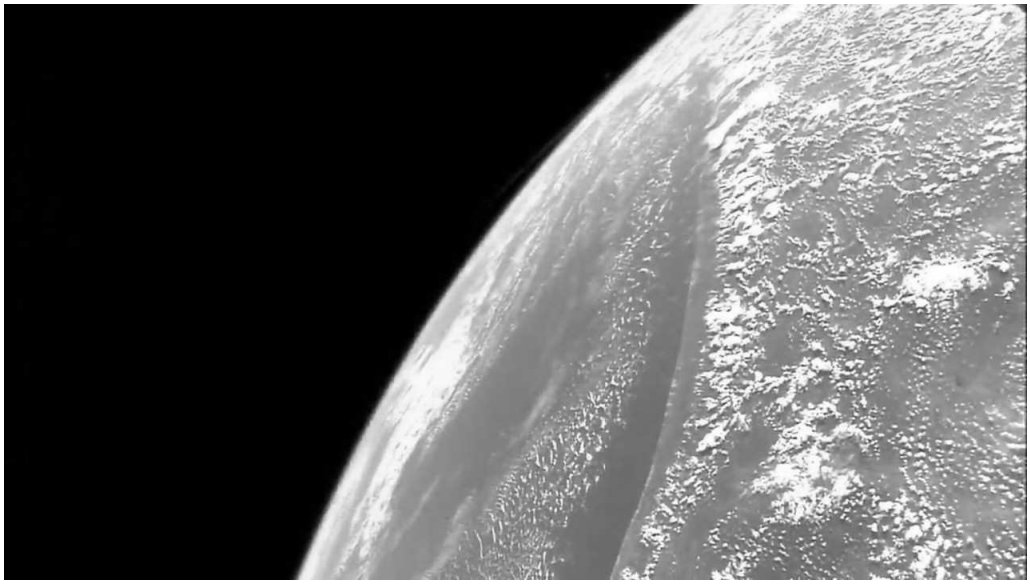

Figura 21: Frame 1400 della sequenza 'Shuttle'

Per la terza sequenza utilizzata (“Chips”), scaricabile a fini di ricerca dal sito VQEG, il *“frame”* viene estratto alla risoluzione originaria di 720x576. Questo è un esempio di una sequenza ricca di dettaglio e ben codificata, viene scalata con fattore 2 mediante una funzione Matlab, introducendo una sfocatura moderata.

Figura 22: Versione originale della sequenza

Figura 23: Versione ridimensionata

Altre tipologie di sequenze vengono quindi utilizzate per verificare i risultati su diverse risoluzioni.

Il *"frame"* di una sequenza proveniente da YouTube, fortemente compressa e di risoluzione originaria di 320x240, è stato utilizzato per simulare la presenza di sfocatura nativa. Questo esempio viene infatti introdotto ed utilizzato per analizzare l'effetto di uno scalaggio ad alta definizione a partire da situazioni di bassa risoluzione utilizzando il *'panorama mode'*.

Altro esempio riguarda due immagini di una sequenza con caratteristiche tipiche dell'effetto da analizzare, ovvero l'introduzione di una sfocatura volutamente prodotta su aree dell'immagine che sono in secondo piano anche dal punto di vista dell'importanza del contenuto informativo.

Un esempio caratteristico di ciò che si vuole studiare viene riportato nella successiva figura, tipica immagine di test per diverse tipologie di algoritmi di elaborazione, scaricabile da siti di laboratori di ricerca. Le due immagini di test in questione sono caratterizzate per contro da una risoluzione medio-bassa (720x304) ma da una buona qualità. Quindi i due "*frame"* sottoposti a verifica sono selezionati per analizzare l'effetto dato da sfocatura intenzionalmente introdotta dopo essere stati scalati in full HD.

Figura 24: Esempio di sfondo volutamente sfocato rispetto alla figura in primo piano

Infine, l’ultima sequenza analizzata è liberamente scaricabile dai siti di diversi istituti di ricerca, e costituisce un caso di movimenti estremamente rapidi in aree estese di*“frame”* di risoluzione originaria pari a 720x486. Due *“frame”* sono estratti e scalati con fattore 2 usando la funzione Matlab '*imresize*', come si vede dalle figure seguenti. La qualità nelle zone fisse è preservata, non vi è codifica, la sfocatura introdotta dal movimento non viene rilevata dall’osservatore se le immagini vengono visualizzate in sequenza.

Figura 25: Esempio di frame originale della sequenza 'Football'

Figura 26: Frame precedente ridimensionato

Figura 27: Esempio di frame originale della sequenza 'Football'

Figura 28: Frame precedente ridimensionato

Misurazioni e valutazioni a livello locale per sfocatura da codifica

a) La valutazione algoritmica

Nella prima parte di questo capitolo, abbiamo visto che il valore del parametro *MGR_rel* valutato sull'intero *"frame"* può essere utilizzato per evidenziare situazioni di sfocatura da codifica. Ora cerchiamo di confrontare i valori dei parametri presentati a livello locale con le opinioni degli osservatori.

Di nuovo diviene utile in qualità di test il *"frame"* selezionato di risoluzione 1080x1920 dalla sequenza "Barcelona". I *"frame"*sono caratterizzati da contenuto ricco di dettaglio ed estremamente variabile in termini di informazioni contenute, per questo l'esempio diviene estremamente interessante per analizzare gli effetti della codifica per quanto riguarda la sfocatura introdotta a livello locale. Di nuovo selezioniamo le stesse tipologie di codifica per confermare la robustezza del metodo in dipendenza dal set di parametri e dal bit rate scelto, oltre che dal contenuto informativo:

- 1Pass intermediate (1Pint): bassa qualità e medio bit rate (1000 kbps).
- iPod: bit rate molto basso e bassa qualità (600 kbps).
- Common Encoding (CE): qualità intermedia e parametri standard (1000 kbps).
- BlueRay (BR): bit rate e qualità elevata (8000 kbps).
- Constant Quantization (CQ): step di quantizzazione costante e qualità molto elevata.

Il *"frame"* viene sottoposto alle cinque diverse codifiche, cercando di mettere in evidenza in tal modo gli artefatti reali della compressione, dati da movimento, basso bit rate, quantizzazione grezza dei coefficienti DCT. La valutazione a livello locale avviene dividendo il *"frame"* in 5x6 blocchi di dimensione 216x320 ed indicati con X_k, per $k = 1,.., 30$. Il calcolo del parametro MGR_k nei diversi casi e conseguenti considerazioni vengono illustrati di seguito e sono pubblicate in [13]: in sintesi, possiamo affermare che il valore del parametro MGR_k varia in funzione dell'indice k a seconda del contenuto del blocco selezionato, mentre, a parità di contenuto ovvero a parità di indice k, il parametro MGR_k subisce un incremento dipendente dalla qualità della codifica. Valutando la media

sul *"frame" MGR_av* e per codifica in funzione dell'indice *k*, otteniamo valori molto diversi, dai valori più elevati corrispondenti alle migliori codifiche (Constant Quantization, *MGR_av* = 3:7; Blu-Ray, *MGR_av* = 3:2;), ai valori più bassi per bassi bitrate (1 P-Intermediate, *MGR_av* = 1:9, and iPod, *MGR_av* = 1:8). In caso di codifiche di basso livello, si osservano alterazioni a livello di oggetti e di bordi principali, come ad esempio per volti e altre parti delle figure in movimento. Volti a media distanza sono poco distinguibili. Nel caso di codifica di livello intermedio, si presentano valori medi del parametro ed attorno a 2.4: a questo livello si possono osservare imperfezioni riguardanti i bordi degli oggetti. Quindi, come per il parametro *MGR_rel*, anche *MGR_av* può essere equivalentemente utilizzato come una prima stima della qualità globale, confermando la necessità o meno di proseguire l'analisi a livello locale.

Il parametro MGR_k invece non può essere utilizzato in modo preciso come stima locale, perché fortemente influenzato dal contenuto informativo del blocco X_k, mentre può fornire un termine di confronto a parità di contenuto ed in funzione della qualità della codifica: come è possibile osservare graficamente dai successivi esempi, infatti, rimanendo fisso l'indice del blocco in esame, si rileva un aumento del valore del parametro MGR_k al migliorare delle caratteristiche della codifica applicata alla sequenza di test.

Per verificare la validità di tali affermazioni, indipendentemente dalla risoluzione delle immagini di partenza, riportiamo un breve test relativo a quanto ottenuto sperimentando su sequenze a risoluzione inferiore a quella appena considerata.

Esaminiamo ad esempio la sequenza "Mit", tipica sequenza di test per diverse tipologie di algoritmi di elaborazione, scaricabile da siti di laboratori di ricerca ed utilizzata alla risoluzione di 480x800: viene suddivisa in 16 blocchi da 120x200 ai fini dell'analisi. Immagini esemplificative della sequenza per entrambe le codifiche proposte, sono riportate nelle successive figure. Tale sequenza viene sottoposta in particolare a due delle codifiche già analizzate, ovvero:

- 1Pass intermediate (1Pint): bassa qualità e medio bit rate (1000 kbps).
- Common Encoding (CE): qualità intermedia e parametri standard (1000 kbps).

Per la codifica CE, essendo la risoluzione inferiore rispetto al caso precedente, il valore del parametro *MGR_av* si attesta attorno al valore 3.1, confermando la buona qualità della codifica,

valutata come intermedia invece nel caso della sequenza “Barcelona”. La qualità della codifica diviene più scadente per quanto riguarda l’algoritmo 1PassIntermediate, che nel caso della sequenza “Mit” rimane comunque ad un livello qualitativo accettabile, per i motivi appena illustrati.

Figura 29: Sequenza 'Mit' codificata CE

Figura 30: Sequenza 'Mit' codificata 1Pint

I grafici seguenti illustrano la variazione del parametro MGR_k in funzione dell'indice di blocco, sia per quanto riguarda la sequenza "Barcelona" che la sequenza "Mit": da osservare in entrambi i casi lo spostamento verso il basso del grafico relativo al diminuire della qualità di codifica.

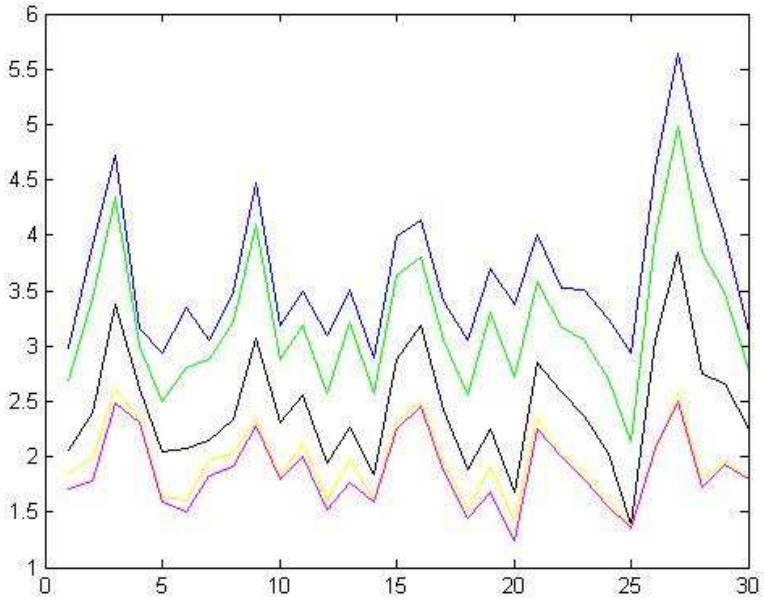

Figura 31: MGR_k per 'Barcelona' codificata a qualità crescente, da iPod (in rosso) a CQ (in blu)

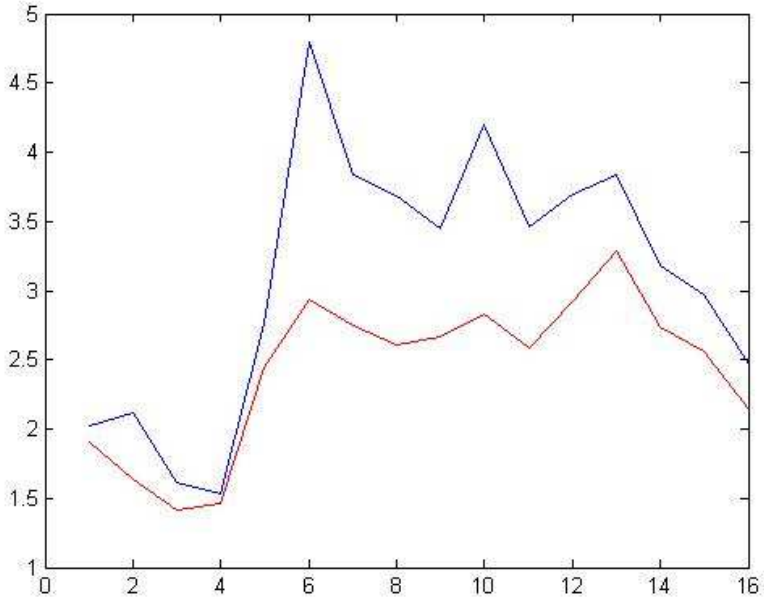

Figura 32: MGR_k per 'Mit' codificata a qualità crescente, da 1Pint (in rosso) a CE (in blu)

In termini di analisi della sfocatura locale, tuttavia, lo studio della sequenza “Barcelona” è maggiormente rappresentativo, grazie alle notevoli variazioni presenti all’interno della sequenza stessa: ciò permette di analizzare una grande varietà di situazioni nei diversi blocchi di una stessa immagine.

Ritroviamo infatti situazioni relative alla presenza di volti, di movimento, di dettaglio fine e, contemporaneamente, zone prevalentemente uniformi. Su ciascuno dei blocchi determinati viene valutata la mappa di salienza ed effettuata la segmentazione, così come vengono applicati i criteri per la rilevazione della presenza di volti.

Solo i blocchi che soddisfano le condizioni stabilite, ovvero un’area sufficientemente estesa nella zona di focalizzazione dell’attenzione, oppure che contengono oggetti cognitivamente importanti, come volti, potrebbero essere sottoposti all’algoritmo di rilevazione della sfocatura in termini oggettivi, ai fini di successive procedure di miglioramento della qualità. Ciò essenzialmente per due motivi.

Per prima cosa, le risorse di calcolo sono limitate ed a tal fine è buon principio limitare l’analisi alle zone dell’immagine effettivamente rilevanti dal punto di vista della percezione. In secondo luogo, è intuitivo pensare che il giudizio degli osservatori sia maggiormente influenzato dalla qualità degli oggetti della scena che colpiscono maggiormente la sua attenzione.

Negli esperimenti riportati, le soglie sperimentali sono determinate come segue: FA_{th} = 25%, la rilevazione dei volti viene applicata su blocchi con un numero di regioni di segmentazione superiori a 15 e con parametri fissati in modo tale che N_{fd} = 2, FD_{th} = 1.2.

Il flag risultante FD_k è uguale a 1 solo per il blocco mostrato in figura 33 e contenente effettivamente dei volti. In tale caso la qualità stimata è comunque scadente, in quanto deterioramento è presente su situazioni significative da un punto di vista cognitivo, come confermano i parametri oggettivi di seguito calcolati.

Gli indici DEP_k (ovvero la frazione di contorni degradati) e BE_k (stima oggettiva della sfocatura) sono comunque calcolati per ogni blocco di interesse X_k. Il valore del parametro Δ viene posto sperimentalmente pari a 4 per la rilevazione dei contorni degradati.

Le successive figure mostrano esempi di blocchi stimati di buona o cattiva qualità in funzione delle misurazioni oggettive effettuate tramite la stima di BE_k e sono riportati i bordi marcati come oggettivamente degradati per diversi blocchi dell’immagine codificata CE.

Per tutti i blocchi selezionati, viene calcolato il numero di regioni in uscita dall'algoritmo di segmentazione, con parametri scelti in funzione dell'attività del blocco, stimata attraverso l'indice *MAG*. Ciò che viene ottenuto è visibile negli esempi successivi, che illustrano vari casi di blocchi con numero di regioni estremamente differente e con valutazioni oggettive di sfocatura molto diverse fra loro.

È possibile osservare l'effetto della sfocatura a livello percettivo in funzione del numero di regioni presenti nella scena: di fatto, a parità di misurazione di sfocatura, questa è molto più evidente all'osservatore in corrispondenza di una scena caratterizzata da un basso numero di oggetti.

In particolare, in figura 38 viene raffigurata una situazione piuttosto interessante da analizzare, in quanto si rileva un'area di attività contenuta ma piuttosto estesa, individuata grazie all'algoritmo di segmentazione: questo fornisce in uscita un numero di regioni particolarmente elevato, di conseguenza l'osservatore, grazie alla sovrabbondanza di dettaglio, è meno influenzato dalla sfocatura presente sullo stesso, sebbene questa risulti oggettivamente significativa.

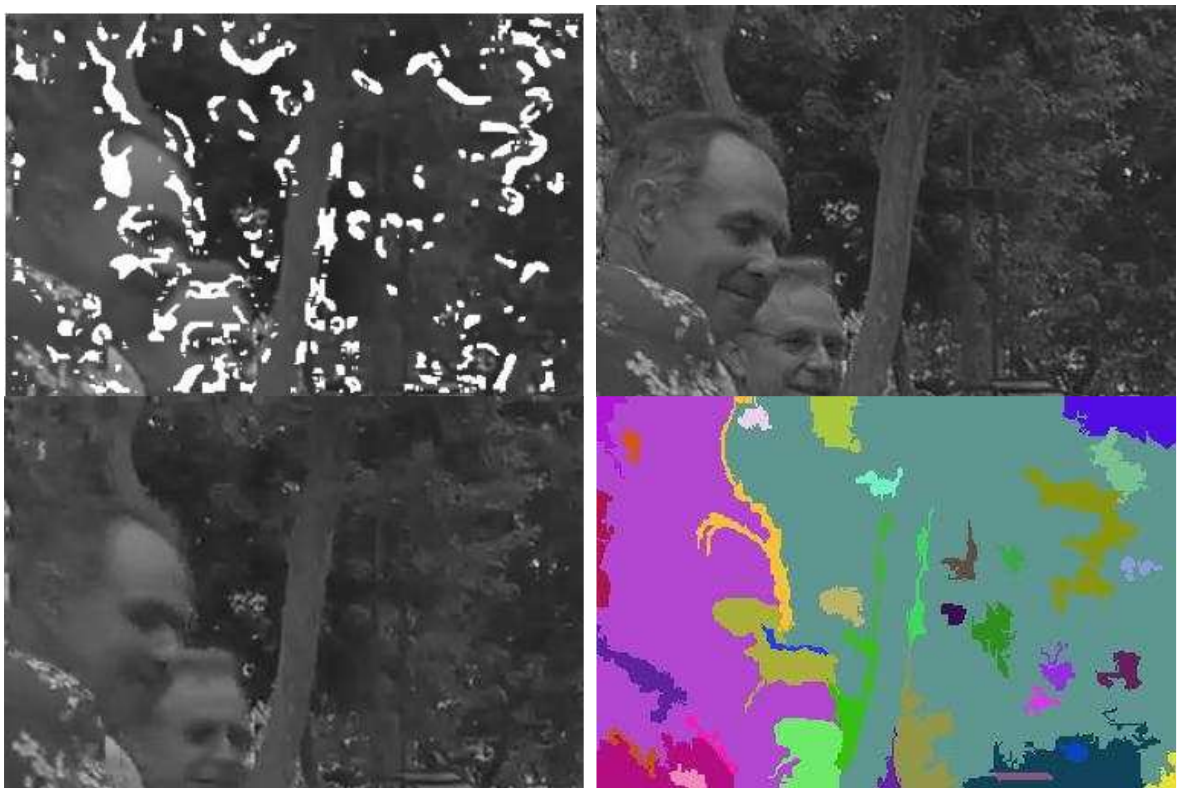

Figura 33: Blocco in originale e codificato CE, con bordi degradati evidenziati e regioni di segmentazione in diversi colori

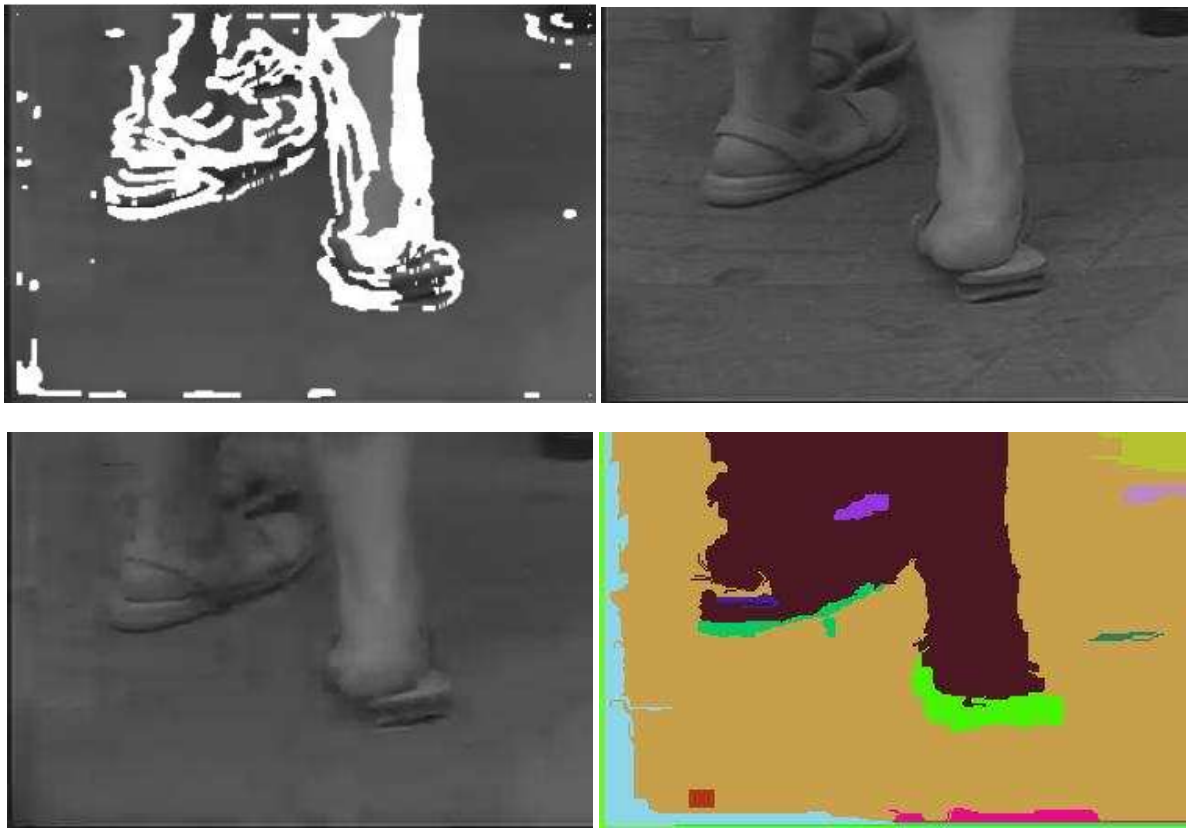

Figura 34: Blocco in originale e codificato CE, con bordi degradati evidenziati e regioni di segmentazione in diversi colori

Figura 35: Blocchi con buona qualità oggettiva ed elevato numero di regioni

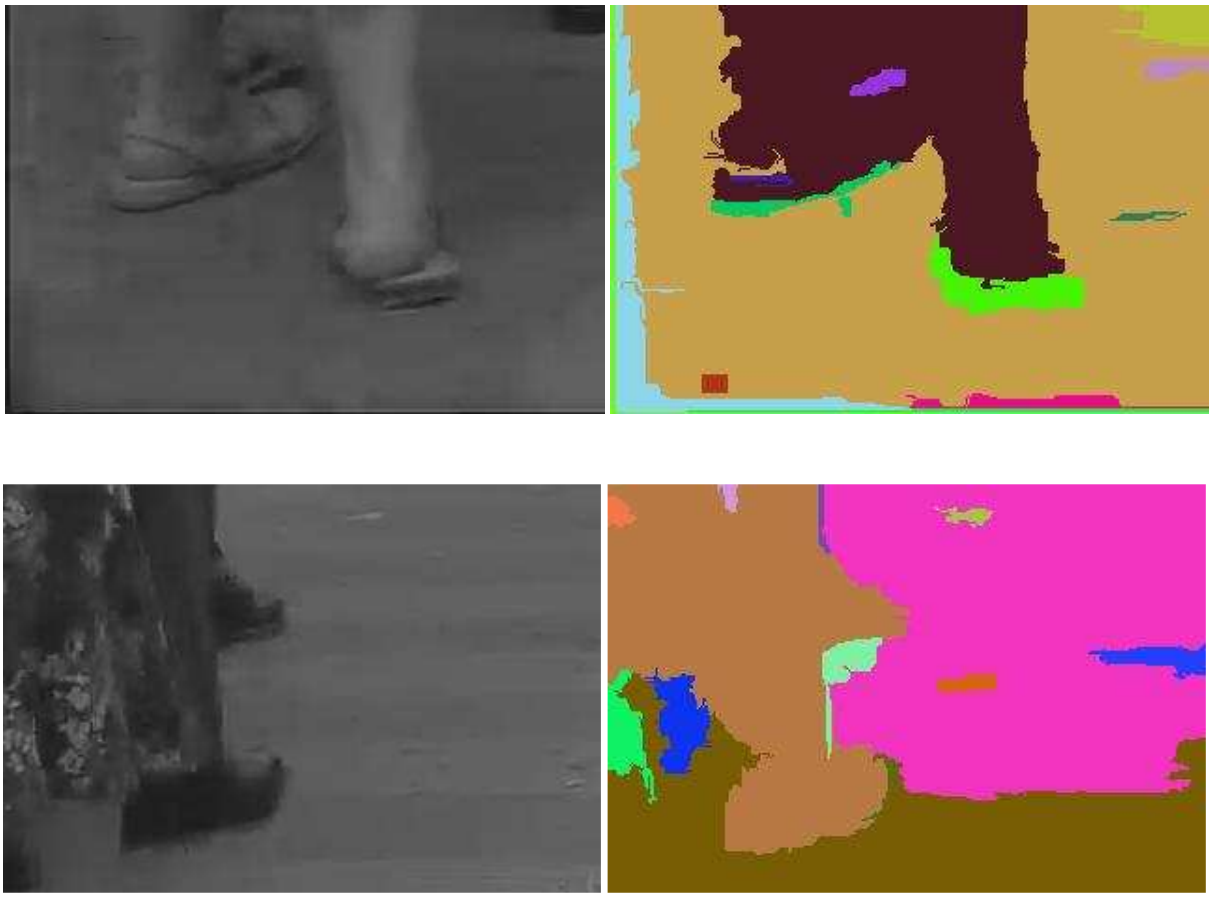

Figura 36: Blocchi con scadente qualità oggettiva e basso numero di regioni

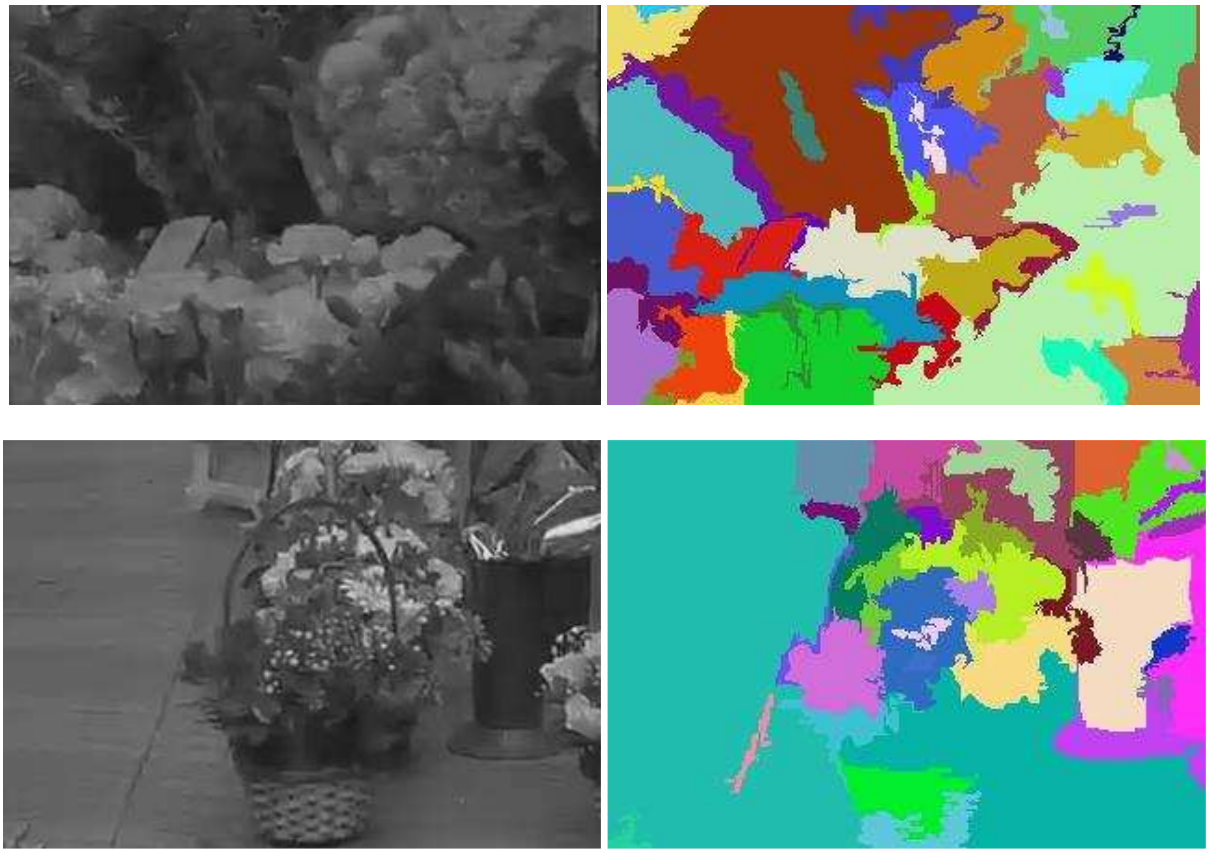

Figura 37: Blocchi con valori intermedi di qualità oggettiva e di numero di regioni

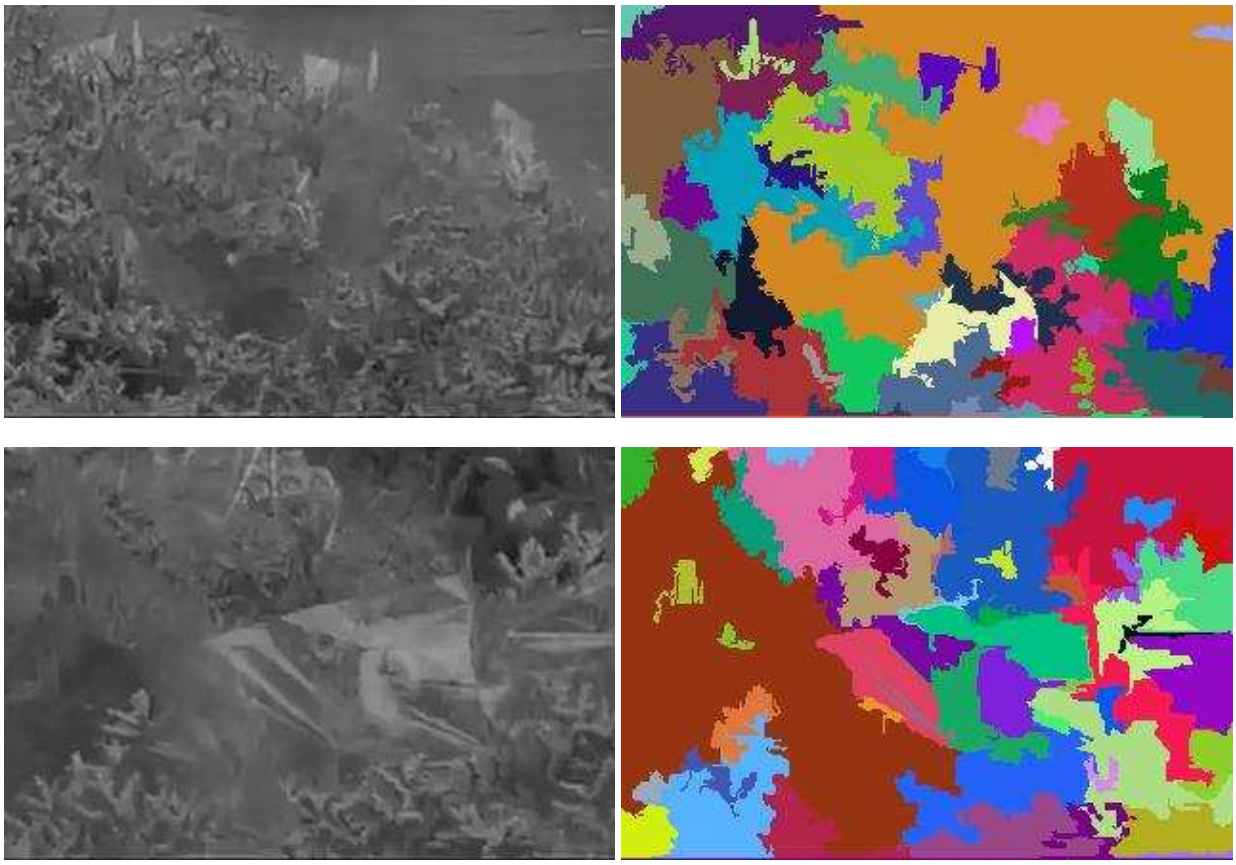

Figura 38: Blocchi con scadente qualità oggettiva ed elevato numero di regioni

b) Valutazione della qualità soggettiva

A questo punto si tratta di confrontare le considerazioni fatte a livello oggettivo con le valutazioni percettive degli osservatori umani che hanno partecipato al test. Avendo valutato la metrica proposta su blocchi, anche gli osservatori vengono chiamati a valutare gli stessi blocchi analizzati. Infine, in ogni *"frame"*, la distribuzione in uscita dall'algoritmo proposto viene comparata con il punteggio dato dagli osservatori.

La sessione di test con gli osservatori viene organizzata come segue [20]: gli osservatori vengono per prima cosa chiamati a considerare *"frame"* ad elevate qualità ed alla loro risoluzione nativa, in modo tale da calibrare la valutazione rispetto al contenuto del *"frame"* in esame. Segue una breve sessione di training, in cui vengono sottoposti agli osservatori esempi diversi di blocchi corrispondenti a casi di codifica ottima e pessima, con obiettivo di tarare le risposte sugli estremi di qualità analizzata. Quindi, tutti e 30 i blocchi, per ciascuna delle cinque codifiche, vengono sottoposti al giudizio, visualizzando in ordine casuale blocchi corrispondenti ai diversi casi.

Ogni blocco viene mostrato all'osservatore per un intervallo di tempo pari a 3 secondi e su sfondo nero, quindi all'osservatore viene richiesto di esprimere un giudizio da 1 a 5, dove 1 corrisponde alla qualità peggiore e 5 a qualità eccellente, come negli esperimenti effettuati sulla valutazione globale. Il set completo di test viene eseguito due volte per minimizzare errori di giudizio dovuti a distrazione.

Il test è stato sottoposto a 9 soggetti, sia maschi che femmine, con diverso grado di esperienza nell'ambito di analisi. Il calcolo del *Mean Opinion Score* per ogni blocco viene quindi stimato, come spiegato in [13], dando al punteggio attribuito da ogni osservatore pesi variabili in funzione della coerenza delle risposte nelle due serie di test.

Come anticipato nelle precedenti sezioni, il set di parametri prescelti come pesi nella valutazione dell'indice complessivo varia a seconda: delle stime di sfocatura oggettiva; del numero di regioni in uscita dall'algoritmo di segmentazione; ciò in modo da ottenere la migliore approssimazione di quanto può essere il *Mean Opinion Score* per il blocco *k*. Questa ottimizzazione viene eseguita nel senso del *Least Mean Square Error*, ed in accordo con quanto riportato in VQEG, con pesi più bassi assegnati a errori residui su *MOS* di valore corrispondente a maggiore dispersione.

Per definire I_k in dipendenza dagli indici BE_k e NoR_k, calcolati per ogni blocco *k*, le ottimizzazioni vengono fatte in modo distinto per tre differenti intervalli di qualità, come già precedentemente

spiegato e discriminando sulla base dell'indice *MGR_av* secondo la Tabella 1 di seguito riportata al fine di sintetizzare quanto analizzato.

Qualità	*MGR_av*	Codifica	α	β	γ
Alta	[2.6, + ∞]	ConstQuant, Blu-Ray	0.33	0	5.20
Media	[2, 2.6]	CE	0.62	0.43	2.81
Scadente	[- ∞, 2]	1Pint, iPod	0.78	0	3.93

Tabella 1

La tabella conferma, in termini di valori scelti, come l'importanza del numero di regioni, in uscita dall'algoritmo di segmentazione, valga ai fini della stima globale solo in presenza di qualità intermedia: in casi di qualità ottima o pessima, il numero di oggetti presenti nella scena non influisce in modo significativo sulla percezione degli osservatori. Al fine di calcolare la correlazione fra stime di qualità, valutata in modo automatico mediante l'algoritmo, e il *MOS,* valutato su ogni blocco k, si procede alla rilevazione di una serie di valori prescelti a tale verifica. Ulteriori rappresentazioni grafiche vengono riportate in [13] al fine di verificare la congruenza dei dati ottenuti.

Nella successiva tabella vengono sintetizzate, nell'ordine, le seguenti valutazioni:

- i coefficienti di Pearson ρ di correlazione fra *I* e *MOS*
- le misure di Spearman r_s di correlazione fra punteggi attribuiti
- *MAE* (mean absolute error) e *RMS* (root mean square error)
- *OR,* ovvero il numero di outlier rapportato al numero di blocchi, con valori di I_k considerati outlier se $| I_k - MOS_k | > 2^* \sigma_k$, con la deviazione standard σ_k valutata introducendo pesi differenziati per tener conto dell'incertezza dell'osservatore [13].

Dato il ridotto numero di soggetti sottoposti al test e la necessità di fare una stima della robustezza dei dati ottenuti, un ulteriore indice viene stimato per effettuare il confronto, ovvero il parametro denominato Observer Standard Deviation (*OSD*), che è la deviazione standard del punteggio soggettivo per quanto riguarda il blocco *k*, mediato sull'insieme dei blocchi N_{blocks}.

Nella seguente tabella riportiamo quindi tutti gli indici di correlazione prescelti per il confronto fra metriche e giudizi degli osservatori.

Qualità	ρ	r_s	MAE	RMS	OSD	OR
Alta	0.46	0.51	0.31	0.43	0.59	0.05
Media	0.80	0.77	0.28	0.34	0.59	0
Scadente	0.86	0.81	0.20	0.24	0.51	0.01

Tabella 2

Va sottolineato che includere il numero di regioni di segmentazione nella valutazione della metrica è particolarmente utile in situazioni di qualità intermedia.

Come si può notare dai valori riportati nella precedente tabella, le migliori rispondenze fra valutazioni degli osservatori e valutazioni algoritmiche si hanno in casi di qualità scadente, casi in cui il metodo si dimostra particolarmente efficace per identificare le aree degradate. Al contrario, in situazioni di qualità ottimale, come nella situazione di codifica CQ, si presentano le maggiori divergenze, dovute principalmente a giudizi soggettivi non accurati e particolarmente affetti da dispersione, dovuta ad esempio a distrazioni.

Si può sottolineare il fatto che, indici come il *MAE* e *RMS*, entrambi misure della distanza media fra metrica e *MOS*, assumono in ogni caso valori più bassi rispetto all'indice *OSD*, ovvero la deviazione standard dei valori soggettivi osservati. Ciò significa che l'errore medio commesso dal metodo nella stima del giudizio percettivo degli osservatori risulta di minore entità rispetto all'incertezza di ogni stima soggettiva richiesta ai partecipanti all'esperimento. Da sottolineare infine, il minor numero di outlier ottenuti rispetto ad altri metodi proposti nello stato dell'arte.

In conclusione, riferendosi all'analisi qui riportata e dettagliata in [13], si può affermare che il metodo proposto è in grado di rispondere in modo piuttosto efficace, nell'intento di riprodurre la

distribuzione delle valutazioni attribuite da osservatori umani e complessivamente su tutti gli esempi di codifica esaminati.

Tale considerazione non viene sminuita dalle precedenti osservazioni sulla maggiore dispersione in ambito di giudizio soggettivo: tale maggiore dispersione è di fatto prevedibile nelle valutazioni soggettive, se messe a confronto con quanto automaticamente calcolato e quindi esattamente riproducibile, esente dal grado di aleatorietà dato dall'elemento soggettivo.

L'algoritmo selezionato come modello di sistema visivo umano si dimostra inoltre adatto allo scopo, come emerge dal test e dalla coerenza delle valutazioni effettuate sull'intera immagine e sul solo sottoinsieme di blocchi rilevanti in termini di focalizzazione dell'attenzione [13], sfruttando il modello definito in [18] ed applicato come qui illustrato.

VI. CONSIDERAZIONI COMPLESSIVE E SVILUPPI FUTURI IN AMBITO MULTIMEDIA

Al termine di questa esposizione, possono essere fatte diverse considerazioni complessive e significative, da un punto di vista del superamento dei punti di criticità presenti nei lavori spesso rilevati nello stato dell'arte, sia in termini di valutazione della qualità che di misurazione degli artefatti principalmente ad essa correlati.

Una prima osservazione può essere, ad esempio, la proposta di integrazione fra modelli di misurazione della qualità sia in termini di percezione che in termini di valutazione oggettiva, tramite l'opportuna considerazione dei modelli del sistema visivo umano e la contestualizzazione dell'artefatto presente in termini di caratteristica locale ad esso correlata.

La quantificazione oggettiva presente viene quindi soppesata con l'importanza dell'artefatto dal punto di vista dell'osservatore, tenendo conto tanto delle cause che hanno determinato la presenza di sfocatura, quanto dell'importanza dell'area in cui si presenta, sia in termini di focalizzazione dell'attenzione che in virtù di considerazioni di carattere cognitivo.

L'importanza della identificazione della causa di sfocatura è determinante non solo in un simile contesto: al fine di ottimizzare un sistema di miglioramento della qualità percepita, tenendo contemporaneamente contenute le risorse richieste in termini di calcolo, è possibile determinare quali artefatti ed in quale zona dell'immagine andranno prevalentemente trattati.

Altre situazioni potranno essere trascurate ed altre ancora, invece, verranno ignorate, come nel caso in cui la sfocatura venga volutamente introdotta in una sequenza video al fine di mettere ulteriormente in rilievo l'informazione fondamentale della scena. Diverse cause dell'artefatto potrebbero essere, in lavori futuri, collegate a diverse metodologie di intervento e di miglioramento, evitando di introdurre ulteriori artefatti come in casi di enfasi introdotta su parti in movimento.

L'osservatore umano ed i modelli di percezione rimangono sempre e comunque in primo piano in questo lavoro, sempre a fianco di parametri di misurazione dell'entità oggettiva.

Proprio in tale senso, tutta una serie di metriche è stata definita per rilevare e misurare la sfocatura imputabile alle diverse cause: in particolare è stata quantificata e localizzata la sfocatura da codifica, distinguendo la stessa da situazioni di sfocatura nativa, da acquisizione o intenzionalmente introdotta.

La definizione di parametri e metriche così soppesati potrebbe essere poi di esempio per rapportare alla percezione soggettiva le diverse tipologie di artefatti, come ad esempio la blocchettatura stessa, per arrivare alla fine ad un unico indice di qualità soggettiva.

Da sottolineare, nella procedura proposta, l'importanza delle operazioni di segmentazione della scena e di determinazione delle aree di focalizzazione dell'attenzione, tramite algoritmi e modelli già riconosciuti dalla comunità scientifica: tali metodologie potrebbero poi essere conseguentemente applicate per individuare zone omogenee di trattamento nell'ottica del miglioramento percettivo adattativo.

L'efficacia dell'algoritmo così elaborato viene dimostrata mediante una selezione di test opportunamente scelti, tali da coprire i casi più significativi in modo sufficientemente esaustivo: tutto ciò avviene classificando le diverse tipologie di artefatto grazie alle metriche proposte, e dimostrandone la coincidenza con le valutazioni eseguite su un campione di osservatori umani.

Gli osservatori vengono chiamati a valutare l'artefatto sia in termini di danno globale che localmente, considerando che i danni introdotti da sfocatura possono influire in maniera estremamente varia, a seconda del contenuto informativo della sequenza codificata.

Infine, viene proposta una metodologia per unificare differenti misurazioni, significative in termini di sfocatura, in una sola stima, in modo consistente con le osservazioni soggettive di degrado a livello locale e riutilizzabile anche nel caso si trattasse di considerare ulteriori parametri per la definizione di un indice complessivo.

Sicuramente il lavoro così come si presenta non si presta agevolmente all'esecuzione in tempo reale su piattaforme commerciali: si può pensare quindi, come ulteriore sviluppo, di proseguire con una fase di ottimizzazione delle procedure intermedie utilizzate. La diffusione anisotropa, ad esempio, si presta certamente ad un'ottimizzazione in termini computazionali ed ai fini dell'implementazione di elettronica di consumo. Tale passo successivo costituirebbe uno sviluppo importante su opportune piattaforme, sufficientemente potenti ed adatte ad operare a bit rate di trasmissione caratteristici del video ad alta definizione.

Una ulteriore opzione potrebbe riguardare l'implementazione non in linea, e quindi senza le conseguenti problematiche temporali di esecuzione, per quanto può riguardare contenuti video da immagazzinare prima di avviare la visualizzazione. Questo caso si presenta come caratteristico per quanto riguarda la visualizzazione di video su web: uno degli obiettivi potrebbe essere proprio quello di migliorare la qualità di immagini trasmesse a risoluzioni estremamente basse e distribuite, ad

esempio, via YouTube o tramite cellulari. Gli algoritmi proposti potrebbero essere quindi utilizzati all'interno di tali catene di elaborazione.

I possibili sviluppi elencati finora, ovviamente in modo estremamente semplificativo, potrebbero essere ulteriormente espansi per quanto riguarda l'ambito della multimedialità, ma andrebbero trattati in un nuovo lavoro dedicato se si intende ampliare il raggio di azione al di fuori di questo ambito e rivolgendosi ad altri ambiti molto specialistici, quali ad esempio il forense.

Il successivo capitolo vuole introdurre l'argomento delle applicazioni forensi in modo estremamente intuitivo, semplice e sintetico, vista la complessità della trattazione e la necessità di rielaborare i criteri stabiliti centrandoli sullo specifico caso di utilizzo.

VII. POSSIBILI SVILUPPI IN SPECIFICI SETTORI DI APPLICAZIONE: IL FORENSE

Il caso di studio

In questa sezione si cerca di analizzare altre possibili applicazioni di algoritmi di valutazione della qualità dell'immagine, diversi dal principale ambito di sviluppo riguardante la multimedialità.

In particolare, uno dei settori che maggiormente sta acquisendo importanza in termini di ricerca si riferisce proprio alle applicazioni per le scienze forensi, con le problematiche ad essa relative e prevalentemente inerenti una stima della qualità dell'immagine, finalizzata ad un miglioramento e/o preventiva al riconoscimento di volti e tracce contenute nell'immagine in esame, da sottoporre ad un operatore esperto o ad un algoritmo di riconoscimento automatico.

Si può ben capire come una prima analisi di qualità dell'immagine da sottoporre a riconoscimento possa rappresentare quindi un importante passo da eseguire preventivamente ad altri tipi di operazioni, anche per ottenere una stima di attendibilità di quanto poi effettuabile in termini di miglioramento e riconoscimento.

Data l'estensione del problema in esame, è necessario limitarsi all'analisi di un determinato tipo di tracce in termini di esempio, poiché a seconda della tipologia selezionata potrebbero verificarsi necessità di trattamenti differenti al fine dell'ottimizzazione e riconoscimento. In questa sede, abbiamo scelto una tipologia di tracce che presenta particolare interesse in quanto ancora non sufficientemente sfruttata in termini investigativi, ovvero l'analisi delle tracce di calzatura rilevate sulla scena del crimine [64].

L'attività degli operatori forensi in una prima fase, direttamente sulla scena del crimine, è determinante per risolvere il caso ed individuare il colpevole: principalmente tale fase si riferisce alla ricerca di prove e tracce rilevanti al fine dell'identificazione. In particolare, le tracce di calzatura lasciate sulla scena del crimine possono permettere alle squadre di polizia di fare luce sulla dinamica del fatto. Con pochi elementi a disposizione e nessun sospettato, la conoscenza di marca e modello della calzatura che corrisponde alla traccia lasciata sulla scena del delitto è una informazione importante al fine della soluzione del caso.

Come detto, il settore rappresenta un ambito di ricerca di estremo interesse ed attualmente in crescita, con problematiche di: gestione di database di immagini riferite ad impronte di calzatura note; e di implementazione di sistemi di riconoscimento automatici o semiautomatici a partire da una traccia di input realmente rilevata. L'analisi di qualità della traccia di calzatura da riconoscere può costituire in tal senso un preventivo esame di attendibilità del responso dell'algoritmo di riconoscimento successivamente applicato.

Ciò è tanto più vero considerando il fatto che, per la maggior parte dei casi, i sistemi di riconoscimento attualmente proposti vengono prevalentemente validati su impronte prodotte sinteticamente in laboratorio, con aggiunta appunto di rumore sinteticamente generato, e non su tracce realmente rilevate sulla scena del crimine, con tutti i relativi problemi che ne conseguono.

In questa sezione, si vuole essenzialmente effettuare un confronto delle prestazioni dei diversi algoritmi proposti in relazione alla tipologia di traccia presentata all'ingresso del sistema di riconoscimento, sicuramente corrispondente a livelli diversi di qualità a seconda della caratteristica di rumore presente ed evidentemente più accentuata nel caso reale.

Nell'ambito dell'attività di tesi, oltre all'analisi della qualità già esposta, è stata anche svolta una fase di studio relativa allo sviluppo di un algoritmo di riconoscimento delle tracce di calzatura [48, 49, 50],[52, 53, 54], in modo orientato ai casi maggiormente interessati da rumore, ovvero le tracce realmente rilevate sulla scena del crimine: tale caso è di natura molto distante dai lavori proposti nello stato dell'arte, che ora andiamo a esaminare.

L'approccio più tradizionale per gli esperti del forense, preposti a rintracciare marca e modello della calzatura che produce la traccia, è attualmente quello di confrontare tale traccia mediante l'utilizzo di dati relativi a modelli di calzatura noti e memorizzati a livello informatico, oppure con modelli contenuti in un catalogo cartaceo.

Ovviamente, il modello attuale in fase di sperimentazione e ricerca tenta invece di superare l'approccio tradizionale: l'obiettivo è infatti quello di ricondurre il problema alla gestione di un database di riferimento, contenente una serie di immagini di impronte corrispondenti a modelli di calzatura noti e correttamente riprodotte. Tali impronte vanno poste a confronto con il dato da identificare, che va a costituire l'input di un algoritmo di riconoscimento sufficientemente robusto, in modo da selezionare e sottoporre ad osservazione diretta dell'operatore forense solo quegli elementi del database che ottengono un punteggio di correlazione più elevato nei confronti della traccia reale in ingresso.

La valutazione può essere attuata in modo semi-automatico che in maniera completamente automatica, senza ulteriore valutazione o altro intervento da parte dell'operatore umano.

Per quanto riguarda l'approccio semi-automatico possiamo citare alcuni riferimenti come [56, 63, 59]: i modelli che vengono proposti in questi lavori hanno caratteristiche basate di fatto sull'esperienza dell'operatore umano, grazie all'intervento del quale è possibile definire un determinato vocabolario costituito da una serie di *"pattern"*geometrici selezionati e sui quali vengono definiti dei criteri di riconoscimento.

Tuttavia, l'interesse principale dell'ambito di studio, come è ben possibile comprendere, si riferisce all'elaborazione di sistemi quanto più possibilmente indipendenti dall'intervento dell'operatore, proprio per svincolarsi dal grado di soggettività nel procedimento decisionale che caratterizza l'operato dell'esperto forense.

Recentemente, tale ambito di interesse è stato sviluppato ed arricchito dall'elaborazione di nuove proposte in termini di sistemi automatici di riconoscimento: può essere utile, a questo punto, effettuare una panoramica degli stessi, in modo da individuarne vantaggi e svantaggi mediante test di confronto, sia per quanto riguarda tracce sinteticamente prodotte in laboratorio che tracce realmente rilevate. Tutto questo può dimostrarsi estremamente utile, nell'ottica di poter selezionare a priori, a partire da una valutazione di qualità della traccia specifica, il procedimento più adatto per effettuare il riconoscimento della stessa.

Lo stato dell'arte

Cominciamo con la descrizione di alcuni sistemi, in termini di metodologie algoritmiche di riconoscimento e di test effettuati per la validazione del metodo, proprio per stabilire la tipologia di input e di qualità richiesta per lo stesso ai fini del buon funzionamento dell'algoritmo in fase di comparazione. Questo tipo di comparazione fra le varie problematiche analizzate, è illustrato prevalentemente in [53] e di seguito riportiamo, in sintesi, lo studio effettuato.

In un primo approccio di tipo automatico [62], le impronte di calzatura vengono raccolte in collaborazione con le forze di polizia, essenzialmente fotografando suole di calzatura o imprimendo impronte in modo tale da simulare la pressione su diverse tipologie di materiali, per poi effettuare la foto. Le immagini così ottenute vengono scalate e memorizzate in formato compresso, i bordi impressi e rilevati per ciascuna impronta vengono classificati ed indicizzati attraverso binarizzazione, come primo passo, per poi essere segmentati in funzione di proprietà di connessione dell'immagine binaria. I bordi estratti vengono quindi classificati con la trasformata di Fourier, in termini sia di ampiezza che di fase, le analisi successive fanno uso principalmente delle proprietà rilevanti delle "*feature*" principali della trasformata per descrivere ogni bordo. Ai fini del riconoscimento viene implementata una rete neurale, con training effettuato tramite "*back propagation*": la descrizione del sistema così costituito è molto dettagliata in termini di architettura ma non di performance, e per tale metodo riesce difficile identificare quali possano essere i principali ambiti di utilizzo, proprio per mancanza di documentazioni ulteriori sulle prestazioni in relazione a casi di utilizzo.

Maggiore accuratezza riguardo la fase di analisi sperimentale si ha ad esempio in [57], pur rimanendo nell'ambito di situazioni sinteticamente riprodotte. Ai fini del riconoscimento, vengono utilizzati in questo caso sia frattali che errore quadratico medio del rumore presente, in quanto ad esempio la decomposizione in termini frattali produce una lista di trasformazioni spaziali che rigenerano l'immagine di partenza se applicate ricorsivamente all'immagine stessa. Da ciò ne consegue che la trasformazione frattale, se applicata all'immagine in esame, non dovrebbe implicare grandi cambiamenti alla stessa, se relativa ad un'immagine simile del database di riferimento. Se applicata invece ad un'impronta molto differente, dovrebbe produrre variazioni consistenti nel "*pattern*" geometrico caratteristico dell'impronta. L'errore quadratico medio del rumore presente viene usato invece come metrica di similarità. Il database di riferimento sul quale vengono effettuati i confronti è composto all'incirca da 150 immagini a livelli di grigio, di dimensione contenuta, il

rumore sinteticamente introdotto è di tipo gaussiano: possono essere introdotte anche rotazioni e traslazioni contenute. I risultati sperimentali presentano una percentuale di riconoscimento non particolarmente elevata, soprattutto nei casi di massima dissimilarità a partire dal campione di riferimento.

Migliori prestazioni vengono ottenute dal metodo descritto in [60] e basato sull'implementazione e sull'analisi della trasformata di Fourier in termini di densità spettrale di potenza (PSD) al fine di caratterizzare l'immagine ed ottenere per la descrizione la proprietà di invarianza alla traslazione. La metrica di similarità è basata su un coefficiente di correlazione bidimensionale, valutato fra la densità spettrale dell'immagine da riconoscere e quella dell'immagine proveniente dal database di riferimento. L'invarianza alla rotazione viene sperimentalmente provata in un intervallo limitato, il database di riferimento viene costituito grazie all'aiuto di volontari, acquisendo numero piuttosto elevato di dati, all'incirca 500 impronte completamente riprodotte a risoluzione elevata e di seguito sottocampionate a diverse risoluzioni. Per provare la robustezza del metodo, tale database viene interrogato anche con impronte parzialmente riprodotte ed i valori di correlazione, calcolati in termini di punteggio medio o cumulativo, vengono utilizzati per ottimizzare la scelta dei parametri del sistema di riconoscimento. I risultati sono soddisfacenti, in quanto l'impronta corretta viene rilevata nel primo 5% del campione del database riordinato con una probabilità pari a 85%. Limite di questo approccio è che non vengono utilizzate nei test immagini rumorose, per le quali non è possibile stimare la robustezza del sistema.

La trasformata di Fourier (FFT) delle impronte viene nuovamente proposta in [65], ma questa volta in termini di correlazione di fase, proprio perché l'informazione di tale fase è di maggiore importanza rispetto a quella di ampiezza e trattiene l'informazione relativa alla geometria caratteristica del *pattern* della suola. Il database di riferimento è costituito da un centinaio di elementi, ovvero da immagini a livelli di grigio di risoluzione intermedia e prefissata, corrispondenti ad impronte di calzatura sinteticamente generate ed organizzate nei seguenti sottoinsiemi: il primo contiene 400 esempi di impronte ottenute dividendo in quarti ogni impronta originale; il secondo contiene 2000 esempi di impronte parziali e rumorose, ovvero il precedente insieme con aggiunta di rumore gaussiano; il terzo gruppo di 2000 elementi viene generato sfocando le impronte del primo insieme, con sfocatura introdotta tramite movimento; l'ultimo gruppo di 2000 impronte viene ottenuto sempre dal primo insieme mediante sovrapposizione di tessiture ricavate dall'album Brodatz [58]. Con queste categorie viene eseguita l'interrogazione del database, dimostrando una casistica di successo

pari al 100%, ovvero la corrispondente impronta viene riconosciuta al primo posto in uscita dall'algoritmo di riconoscimento. Al momento, per quanto riguarda le impronte sintetiche, tali risultai corrispondono alle migliori prestazioni di sistemi automatici.

Altri metodi più recenti proposti, ma che non superano le prestazioni precedentemente descritte, si basano su caratteristiche di invarianza spaziale in corrispondenza di regioni dell'immagine identificate secondo diverse metodologie. Ad esempio, in [55] vengono utilizzati i momenti di Hu, invarianti per rotazione, traslazione e scalaggio, in relazione ad un database contenente 500 immagini di impronte di calzatura. Le immagini del database sono di nuovo generate sinteticamente e addizionate a rumore gaussiano a media nulla e varianza massima pari al 20%: il database completo viene quindi interrogato con test a risoluzione variabile e rotazioni a diverse angolazioni, la distanza euclidea e di Canberra ed altri tipi di correlazione vengono utilizzati come misure di similarità, ma i risultati si mostrano piuttosto variabili in funzione dei valori di varianza applicati.

Ancora partendo da considerazioni di invarianza spaziale, si cerca la definizione di una regione a massima stabilità (Maximally Stable Extremal Region - MSER) rilevata con metodi descritti in [67] per identificare le caratteristiche discriminanti della geometria di una data impronta: quindi l'algoritmo Scale Invariant Feature Transform (SIFT) viene utilizzato per la descrizione, proprio perché si rivela particolarmente efficace in fase di confronto con dati parziali. Il metodo si basa sulla localizzazione e sull'orientazione del gradiente valutato su tutta l'immagine e rappresentato come istogramma, procedendo poi ad una quantizzazione in un numero definito di bin. Il database di riferimento nel test è costituito all'incirca da 400 impronte, ognuna corrispondente sia alla parte sinistra che alla parte destra, il test viene eseguito su un input costituito da un'impronta intera a due diverse risoluzioni. I risultati dimostrano che in una lista di uscita comprendente i primi 8 elementi del database selezionati, in termini di similarità, il corrispondente corretto viene individuato al 91% dei casi.

A partire da quanto detto finora, cerchiamo quindi di ricavare per i successivi confronti gli algoritmi maggiormente significativi, ovvero quanto descritto in [60] e in [65], applicando i metodi proposti sia su tracce di calzatura sinteticamente ottenute sia su tracce reali, con evidente divergenza in termini di qualità di partenza dell'immagine da trattare. Questa comparazione diviene particolarmente significativa e chiarisce l'importanza di un'analisi di qualità a supporto delle prestazioni di un algoritmo di riconoscimento, mettendo in luce la diversità di comportamento nel caso di tracce particolarmente rumorose – e ad un basso livello di qualità corrispondente.

L'algoritmo di riconoscimento su tracce reali

Come esaurientemente elencato, tutti i precedenti metodi sono stati validati nello stato dell'arte su tracce non realmente rilevate sulla scena del crimine, ma sinteticamente prodotte in laboratorio. Proprio per ovviare a questo problema, è stato affrontato nell'ambito del dottorato il problema del riconoscimento in casi di impronte reali, che comprensibilmente rispondono quasi sempre alla situazione di qualità più scadente, in quanto immagini affette da rumore complesso ed imputabile a diverse tipologie di situazioni.

La traccia acquisita può essere riprodotta su diverse tipologie di materiali; può essere prodotta da diversi tipi di sostanze; può essere soggetta a diverse tipologie di rumore ed essere parziale; può essere sovrapposta a "*pattern"* geometrici diversi da quello caratteristico della suola.

Proprio sulla tipicità della geometria che caratterizza la suola di calzatura, oltre che su altri semplici assunti riguardanti la natura dei dati in esame, si basa il metodo che viene proposto. Il "*pattern"* di un'impronta, infatti, può essere considerato come una composizione data da una tessitura geometricamente caratterizzata e da una serie di dettagli in riproduzione unica, come ad esempio il logo della marca della calzatura.

Entrambe andrebbero considerati ai fini del riconoscimento, ma il lavoro svolto prende in considerazione solo la parte caratterizzata da tessitura dell'impronta, in modo da poter agevolmente utilizzare un descrittore che si adatti alla geometria della tessitura corrispondente al "*pattern"* da descrivere; in questa sede, cerchiamo di proporre la teoria alla base dell'algoritmo, visibile nelle sue evoluzioni nei diversi lavori pubblicati [49,50],[52, 53, 54].

Il descrittore proposto si basa quindi sulla distanza di Mahalanobis, con mappa calcolata come descritto in [46] e su una metrica di similarità di tipo correlazione. L'operatore scelto, utilizzato precedentemente in algoritmi per il riconoscimento della presenza di persone all'interno di una scena[46], è particolarmente adattabile al caso in esame proprio per le caratteristiche della traccia reale da descrivere.

La descrizione di tale traccia deve infatti preservare le proprietà in termini di distanza relativa fra elementi appartenenti ad una geometria caratteristica, a prescindere da valori di luminanza o di colore, in modo da riconoscere la tipicità di una determinata tessitura.

Per il calcolo della mappa di Mahalanobis, infatti, vanno valutati parametri statistici come media e varianza, confrontate a livello locale in termini di distanza relativa fra aree dell'immagine. Tale

valutazione statistica ha anche l'effetto di ridurre alcuni possibili tipi di rumore presente.

L'area complessiva dell'immagine in esame viene divisa in blocchi di dimensione *pxq* e per ogni possibile coppia ne viene calcolata la distanza, in modo da ottenere come risultato *M* blocchi in direzione orizzontale ed *N* blocchi in direzione verticale per una dimensione complessiva della mappa di distanze che corrisponde ad *(MN)x(MN)*.

Identifichiamo quindi ogni blocco con un indice *i=1,…, MN* ed ai fini del calcolo delle distanze di Mahalanobis *d(i,j)* fra blocco *i* e blocco *j*, per ogni blocco vengono valutate la media m_i e la varianza σ_i^2, per ottenere una mappa risultante i cui elementi sono definiti come segue:

$$d(i,j) = (m_i - m_j)^2/(\sigma_i^2 + \sigma_j^2),$$

$$d(i,i)=0.$$

Il calcolo viene effettuato su blocchi di dimensione pari a 4 × 4 pixel, ottenendo complessivamente 25 × 25 blocchi a partire da immagini di test a livelli di grigio e di dimensione 100 × 100 pixel.

Il descrittore selezionato diviene quindi la densità spettrale di potenza (PSD) della mappa così ottenuta e la misura di similarità diviene la correlazione così definita [60]: per due segnali di dimensione *sxt* ed indicati con $f_1(x,y)$, $f_2(x,y)$ vengono calcolate la media "*mean*" e la deviazione standard "*std*" per definire due corrispondenti segnali normalizzati come:

$$f_{1normalizzato}(x,y)=[\, f_1(x,y)-mean(f_1(x,y))]/std(f_1(x,y)),$$

$$f_{2normalizzato}(x,y)=[f_2(x,y)-mean(f_2(x,y)])/std(f_2(x,y))$$

e coefficiente di correlazione dato da:

$$corr_{1,2} = (\sum_{x=1,..,s} \sum_{y=1,..,t} f_{1normalizzato}(x,y) \cdot f_{2normalizzato}(x,y))/(s \cdot t)$$

Proprio per tenere in considerazione la diversità delle possibili sorgenti di rumore, che possono

riguardare una traccia realmente rilevata sulla scena del crimine, la tessitura di interesse, ovvero il "*pattern"* rappresentativo della traccia in esame, può essere sottoposto a processi di enfasi prima di procedere al calcolo del descrittore scelto.

Differenti tipologie di rumore possono riguardare, ad esempio, le diverse sostanze su cui è stata depositata o dalle quali è costituita la traccia, in quanto immagine reale: così si può pensare di calcolare il descrittore sia direttamente sull'immagine a livelli di grigio sia dopo l'applicazione di un algoritmo di rilevazione dei bordi significativi, proprio per evidenziare ulteriormente la geometria caratteristica, oppure si può ricorrere all'equalizzazione dell'istogramma per aumentare il contrasto, ad esempio per valori di varianza inferiori ad una soglia sperimentalmente determinata.

Se diviene necessaria una fase di enfasi del contrasto, per evidenziare l'informazione di interesse, spesso le operazioni di equalizzazione e di rilevazione dei contorni significativi andranno effettuate in modo combinato, a seconda della varianza calcolata sulla traccia verrà scelto un algoritmo di elaborazione iniziale più o meno sensibile.

Così come, in casi di rumore espresso da valori elevati della varianza sulla traccia, sarà meglio non procedere ad operazioni che ne potrebbero enfatizzare l'entità, come quelle appena illustrate.

La metrica di similarità applicata per confrontare descrittori così costruiti, per l'immagine in ingresso da riconoscere e per l'immagine di riferimento nel database da confrontare, viene scelta come il coefficiente di correlazione calcolato come [60].

Le migliori prestazioni, come vedremo nel seguito, si possono ovviamente ottenere combinando diversi criteri di rilevazione, come ad esempio in [54] dove viene attuato il riconoscimento in due fasi:

1) una prima selezione "grezza", ovvero l'estrazione di un sottoinsieme di P elementi del database, selezionati per similarità mediante l'algoritmo appena proposto per dati rumorosi ed in modo tale da ottenere una sensibile riduzione dei dati da analizzare al passo successivo;

2) una seconda fase più fine basata sull'algoritmo di correlazione modificata di fase (MPOC) come descritto in [65] ed applicato solo sul sottoinsieme estratto al passo precedente, proponendo infine all'operatore forense un insieme estremamente limitato selezionato per similarità mediante questo secondo algoritmo.

Alcune valutazioni sperimentali

Il database di test viene costituito a partire da modelli di calzatura noti, mediante immagini a bassa risoluzione realizzate a partire da quanto disponibile sul sito ENFSI [61]WGM [51].

Due differenti insiemi di test vengono definiti per verificare la differenza di prestazioni dei metodi nello stato dell'arte e del metodo innovativo proposto, come riportato in [53]:

- un insieme prodotto aggiungendo ad immagini di riferimento del database del rumore sintetico (SyntS);

- un insieme di immagini realmente rilevate sulla scena del crimine e realizzate a partire da quanto disponibile sul sito ENFSI WGM (RealS).

Il proposito è proprio quello di considerare la dipendenza delle performance degli algoritmi in analisi a fronte di una variazione della qualità dell'immagine in esame, in modo tale da poter utilizzare quanto di più efficace a disposizione in base all'informazione relativa alla preventiva valutazione di qualità.

Non esiste a questo proposito un database di riferimento standard per la fase di validazione di simili sistemi automatici o semi-automatici e, come visto, i diversi lavori proposti utilizzano tracce sinteticamente prodotte in modo estremamente differente. In questo caso specifico prendiamo in considerazione immagini di test generate sinteticamente con due diverse tipologie di rumore introdotto, come descritto di seguito e rappresentato nelle successive figure:

- 340 impronte vengono ottenute addizionando rumore gaussiano a media nulla ad ogni elemento nel database di riferimento, grazie a funzionalità Matlab, con diversi valori di varianza da 1% a 15%, come indicato in Tabella 4;

- 425 impronte ottenute mediante sfocatura a partire dalle immagini del database di riferimento, sempre utilizzando funzionalità Matlab, con moto a valore angolare nullo e spostamento da 2 a 20 pixel, corrispondenti a lunghezze da 0.15 a 1.5 cm, come indicato in Tabella 4.

I test eseguiti sono pubblicati in [53].

Per quanto riguarda le tracce reali, le immagini di partenza sono considerate in termini di livelli di grigio e sottoposte ad operazioni grezze di ridimensionamento e rotazione, in modo da ottenere una corrispondenza approssimativa con quanto memorizzato nel database di riferimento. Sono inoltre ricavate diverse aree di test in corrispondenza della singola traccia, ottenendo zone di interesse di bassa risoluzione, pari a 100x100, in modo da identificare diverse situazioni che si possono presentare anche sulla singola traccia. Seguendo tale procedura, si ottengono 85 elementi del database di riferimento, a partire da 25 suole di calzatura di modello noto, e 35 casi di test.

Sui diversi set di test a disposizione a questo punto vengono applicati i diversi algoritmi di riconoscimento prescelti.

I risultati vengono confrontati osservando i raggruppamenti di dati proposti come più simili alla traccia da riconoscere ed in riferimento agli elementi noti del database, ripetendo tale procedura per ciascun degli algoritmi di riconoscimento: i primi elementi del database visualizzabili dall'operatore saranno quelli corrispondenti a più elevati valori delle metriche di similarità prescelte nelle diverse situazioni. In particolare, vengono considerati: il primo elemento proposto; i primi cinque; i primi dieci; i primi venti.

Il database di riferimento viene interrogato con l'insieme di test sinteticamente prodotto e con l'insieme di test relativo alle immagini reali per gli algoritmi che implementano: densità spettrale di potenza (PSD) e descritto in [60] ; la correlazione modificata di fase (MPOC) come descritto in [65]; l'algoritmo proposto per tracce reali.

Come emerge dai dati riportati nelle successive Tabelle, vi è molta differenza fra i risultati ottenuti nel caso di rumore sintetico e rumore reale.

La Tabella 4 riporta quanto ottenuto confrontando con gli elementi nel database di riferimento le tracce di test generate in laboratorio ed il metodo MPOC dimostra ottime prestazioni nel caso di rumore sinteticamente prodotto, sia nel caso di rumore gaussiano che di sfocatura da traslazione, mentre gli altri due algoritmi si dimostrano sufficientemente equivalenti in tal senso, con il metodo PSD che fra i due presenta prestazioni leggermente migliori.

I risultati ottenuti per questo primo set di test, con rumore sinteticamente addizionato in ambiente Matlab, mostrano come, nelle diverse condizioni di varianza da rumore gaussiano e di sfocatura, corrispondano i casi di rilevazione corretta, rispettivamente entro primi 1, 5, 10 e 20 elementi del database estratti in termini di similarità: quanto ottenuto viene presentato, in Tabella 4, come

percentuali di successo sui test effettuati.

Al contrario, la Tabella 3 illustra quanto ottenuto per quanto riguarda il caso di rumore reale: il metodo proposto in questo lavoro ottiene risultati migliori rispetto al PSD, contrariamente a quanto rilevato per il caso sintetico, proprio perché pensato per essere robusto rispetto alle diverse tipologie di rumore che possono presentarsi sulla scena del crimine e che non possono essere ricondotte a quanto generato sinteticamente da funzioni di simulazione in ambiente di laboratorio. I risultati ottenuti, per il set di test relativo alle tracce reali e riportati in Tabella 3, si riferiscono in particolare al numero di rilevazioni corrette, rispettivamente entro primi 1, 5, 10 e 20 elementi del database estratti in termini di similarità: quanto ottenuto viene poi tradotto in termini percentuali sui test effettuati.

Si vede inoltre come l'algoritmo MPOC, che in caso di rumore sintetico dimostrava prestazioni sicuramente ottime, peggiori le performance in modo evidente rispetto al caso precedente, ciò a sottolineare ancora una volta l'importanza di una valutazione preventiva riguardo alle prestazioni dei diversi algoritmi di riconoscimento nelle diverse situazioni.

Per ottenere poi le migliori prestazioni, è necessario, come detto, procedere con la combinazione di due procedimenti, selezionando un sottoinsieme del database mediante la procedura più robusta al rumore ed applicando in un secondo passo il riconoscimento di fase solo su tale sottoinsieme, ottenendo il seguente grafico [54]: la prima posizione di riconoscimento viene raggiunta in ben 17 situazioni di test su 35.

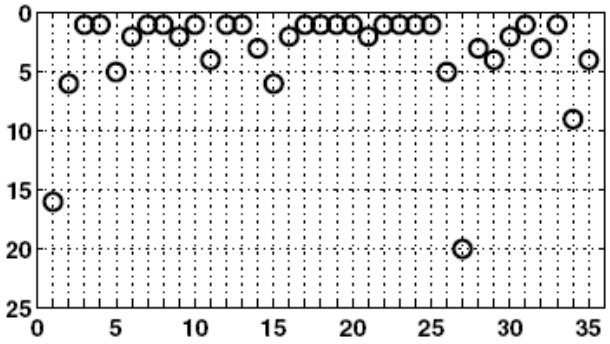

Figura 39: Posizione dell'elemento corrispondente al riconoscimento corretto in caso di tracce reali per un algoritmo combinato

Ciò che è possibile intuire in modo molto chiaro, a partire da questo esempio ma certamente estendibile anche ad altre tipologie di tracce, è che, a seconda delle caratteristiche del dato da riconoscere, va stimata l'efficacia dell'algoritmo di riconoscimento nella situazione in analisi, il che coinvolge in ogni caso una stima di qualità effettuata sull'immagine di partenza.

Questa considerazione è stata semplificata ponendo a confronto tracce sintetiche e tracce reali, caso in cui è facile presagire l'esito di una simile valutazione, ma tale osservazione può essere facilmente estesa a dati di provenienza non nota a priori, unicamente sottoponendo l'immagine ad una valutazione di qualità preventiva.

RealS	Algoritmi											
	PSD				MPOC				Algoritmo proposto			
	1	5	10	20	1	5	10	20	1	5	10	20
#	4	12	15	17	10	14	18	24	5	13	15	22
%	11	34	43	49	29	40	51	69	14	37	43	63

Tabella 3: Numero e percentuali di rilevazione corretta nei primi 1, 5, 10, 20 elementi estratti dai diversi algoritmi nel caso di tracce reali

Figura 40: Elementi di test con rumore gaussiano a varianza crescente

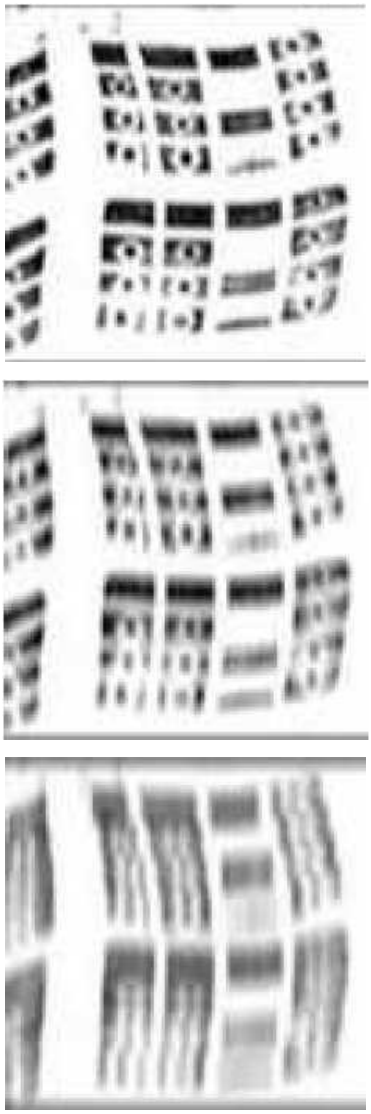

Figura 41: Elementi d test affetti da sfocatura da traslazione

SyntS	Algoritmi											
	PSD				MPOC				Algoritmo Proposto			
	1	5	10	20	1	5	10	20	1	5	10	20
1%	99	100	100	100	100	100	100	100	95	99	100	100
5%	82	89	93	98	100	100	100	100	78	93	99	99
10%	71	75	82	86	100	100	100	100	74	86	92	98
15%	67	72	74	76	100	100	100	100	-	-	-	-
2 pxl	98	100	100	100	100	100	100	100	100	100	100	100
5 pxl	95	100	100	100	100	100	100	100	87	98	99	100
10 pxl	60	69	75	79	100	100	100	100	39	74	79	93
15 pxl	68	82	85	89	96	100	100	100	-	-	-	-
20 pxl	34	42	60	74	100	100	100	100	8	38	52	72

Tabella 4: Percentuali di rilevazione corretta nei primi 1, 5, 10, 20 elementi estratti dai diversi algoritmi selezionati e nelle diverse condizioni di rumore sinteticamente prodotto

VIII. CONCLUSIONI

A conclusione del lavoro svolto, si vuole ulteriormente ribadire la vastità dell'argomento affrontato, ovvero l'analisi di qualità dell'immagine, già presente nello stato dell'arte da lungo tempo ma ancora in fase di studio e di ricerca in svariati settori, proprio per l'ampiezza e la specificità della trattazione necessaria nei diversi ambiti di applicazione.

Per ottenere una panoramica esaustiva in tale campo, sarebbe necessario, anche solamente riferendosi all'ambito delle applicazioni multimediali, considerare la rilevazione di altri artefatti, come ad esempio blocchettatura, già inizialmente menzionata, ed il "*ringing*", per poi ricorrere ad un'integrazione dei parametri ottenuti per ciascuno degli artefatti selezionati ai fini di una valutazione complessiva.

Si vuole inoltre sottolineare come, anche esaminando solamente quanto visto finora, il problema dell'analisi di qualità sia strettamente correlato con altre operazioni particolarmente rilevanti nel campo dell'elaborazione delle immagini, e come in ogni ambito di sviluppo la qualità dell'immagine possa essere considerata come una fase propedeutica ad importanti applicazioni specifiche del settore di interesse.

Abbiamo visto come, in sistemi multimediali, l'analisi di qualità può essere utile per direzionare in modo corretto la fase di miglioramento della qualità percepita; in ambito forense, invece, può divenire una stima dell'efficacia di un algoritmo di riconoscimento automatico, supportando l'operatore forense nella scelta della procedura da applicare per effettuare tale operazione ed eventualmente ricorrere, in fase preliminare, ad opportuni trattamenti di enfasi della traccia di interesse.

Lo studio effettuato vuole essere quindi una base per futuri sviluppi ed applicazioni, sempre se opportunamente adattato alla peculiarità dell'argomento di ricerca e sostenuto da una fase sufficientemente esaustiva di misurazioni e di test soggettivi a sostegno dell'approccio teorico adottato.

Al termine di questo lavoro, un ringraziamento sentito a Philips Consumer Electronics, in particolare nella persona di Jeroen Stessen, per il supporto e l'interesse dimostrato nei confronti degli studi effettuati.

ELENCO ABBREVIAZIONI

H264	High 264 coding
JPEG	Joint Photographic Experts Group
NSS	Natural Scene Statistics
DCT	Discrete Cosine Transform
MPEG	Motion Picture Experts Group
JND	Just Noticeable Difference
HD	High Definition
SD	Standard Definition
FIR	Infinite Response Filer
MOS	Mean Opinion Score
1Pint	1 Pass Intermediate Encoding
CE	Common Encoding
CQ	Constant Quantization Encoding
BR	Blu – Ray Encoding
VQEG	Video Quality Experts Group
MAE	Mean Absolute Error
RMS	Root Mean Square Error
OR	Outlier Ratio
SIFT	Scale Invariant Feature Transform
OSD	Observer Standard Deviation
PSD	Power Spectral Density
FFT	Fast Fourier Transform
MSER	Maximally Stable Extremal Region
MPOC	Modified Phase Only Correlation
ENFSI	European Network of Forensic Science Institutes
WGM	ENFS Working Group on Marks

RIFERIMENTI

[1] J. Xia, Y. Shi, K. Teunissen and Ingrid Heynderickx, “Perceivable artifacts in compressed video and their relation to video quality”, Signal Process.: Image Commun., vol. 24, no 2, pp. 548-556, August 2009.

[2] L. Firestone, K. Cook, N. Talsania, and K. Preston, “Comparison of autofocus methods for automated microscopy,” Cytometry, vol. 12, pp. 195–206, 1991.

[3] Z.M. Parvez Sazzad, Y. Kawayoke, Y. Horita, “No reference image quality assessment for JPEG2000 based on spatial features”, Signal Process: Image Commun, vol. 23, no. 4, pp. 257-268, April 2008.

[4] J. Caviedes and F. Oberti, “A new sharpness metric based on local kurtosis, edge and energy information,” Signal Process.: Image Commun.,vol. 19, no 2, pp. 147–161, 2004.

[5] J. Caviedes and F. Oberti, “An automatic focusing and astigmatism correction system for the sem and ctem,” J. Microscopy, vol. 127,pp. 185–199, 1982.

[6] J. Zhang, S.H. Ong, T. M. Le, “Kurtosis-based no-reference quality assessment of JPEG2000 images”, Signal Proc.: Image Commun., vol. 26, no. 1, pp. 13-23, January 2011.

[7] N. K. Chern, N. P. A. Neow, and M. H. A. Jr, “Practical issues in pixel based autofocusing for machine vision,” IEEE Int. Conf. Robotics and Automation, vol. 3, pp. 2791–2796, 2001.

[8] X. Marichal, W. Ma, and H. J. Zhang, “Blur determination in the compressed domain using dct information,” in Proc. IEEE Int. Conf. Image Processing, vol. 2, pp. 386–390, 1999.

[9] I. van Zyl Marais, W. Herman Steyn, “Robust defocus blur identification in the context of blind image quality assessment”, Signal Process.: Image Commun, vol. 22, no. 10, pp. 833-844, November 2007.

[10] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, “Perceptual blur and ringing metrics: Applications to jpeg2000,” Signal Processing:Image Communication, vol. 19 no.2, pp. 163–172, 2004.

[11] E. P. Ong, W. S. Lin, Z. K. Lu, S. S. Yao, X. K. Yang, and L. F. Jiang, “No-reference quality metric for measuring image blur,” in Proc. IEEE Int. Conf. Image Processing, vol. 2, pp. 469–472, 2003.

[12] R. Ferzli and L. Karam, “A no reference objective image sharpness metric based on the notion of just noticeable blur (jnb),” IEEE Transactions on Image Processing, vol. 18, pp. 717–728, 2009.

[13] F. Dardi, L. Abate, and G. Ramponi, “No-reference measurement of perceptually significant blurriness in video frames,” Signal, Image and Video Processing (Springer), Vol. 5, Issue 3 (2011), Page 271-282

[14] P. Barten, “Physical model for the contrast sensitivity of the human eye,” in Human Vision, Visual Processing, and Digital Display III (M. H. Loew, ed.), vol. 1666 of Proc. SPIE, pp. 57–72, 1992.

[15] P. Perona and J. Malik, “Scale-space and edge detection using anisotropic diffusion,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, pp. 629–639, 1990.

[16] P. Felzenszwalb and D. Huttenlocher, “Efficient graph-based image segmentation,” International Journal of computer Vision, vol. 59, pp. 167–181, 2004.

[17] C. Koch and S. Ullman, “Shifts in selective visual attention: Towards the underlying neural circuitry,” Human Neurobiology, vol. 4, pp. 219–227,1985.

[18] L. Itti, C. Koch, and E. Niebur, “A model of saliency-based visual attention for rapid scene analysis,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 20, pp. 1254–1259, 1998.

[19] I. Bogdanova, A. Bur, and H. Hgli, “Visual attention on the sphere,” IEEE Transactions on Image Processing, vol. 17, pp. 2000–2014, 2008.

[20] “Methodology for the subjective assessment of the quality of television pictures,” ITU-R Rec. BT.500-12, International Telecommunication Union, Geneva, Switzerland, 2009.

[21] J. Xia, Y. Shi, K. Teunissen and Ingrid Heynderickx, “ A Perceptual Blurring Metric for Compressed Video”, 1st International Conference on Information Science and Engineering (ICISE), December 2009.

[22] Z.M. Parvez Sazzad, Y. Kawayoke, Y. Horita, “Spatial Features Based No Reference Image Quality Assessment for JPEG2000”, IEEE International Conference on Image Processing, 2007. ICIP 2007, Sept. 2007.

[23] J. Caviedes and S. Gurbuz, “No reference sharpness metric based on local edge kurtosis,” Image Processing. 2002. Proceedings. 2002 International Conference on, vol. 3, pp. 53-56.

[24] A. K. Moorthy and A.C. Bovik, “A Two-Step Framework for Constructing Blind Image Quality Indices ”, Signal Processing Letters, IEEE, issue 5, May 2010, pp. 513-516.

[25] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, “ A no reference perceptual blur metrics,” Proceedings. International Conference on Image Processing. 2002., vol. 3, pp. 57-60.

[26] R. Ferzli and L. Karam, “Human Visual System Based No-Reference Objective Image Sharpness Metric,” Image Processing, 2006 IEEE International Conference on, Oct. 2006, pp. 2949 – 2952.

[27] Gopalakrishnan V.; Yiqun Hu; Rajan, D., “Random Walks on Graphs for Salient Object Detection in Images”, Image Processing, IEEE Transactions on, Vol. 19, Issue 12, 2010, pp. 3232 – 3242.

[28] Le Meur, O.; Chevet, J.-C., “Relevance of a Feed-Forward Model of Visual Attention for Goal-Oriented and Free-Viewing Tasks,” Image Processing, IEEE Transactions on, Vol. 19 , Issue 11, 2010 , pp. 2801 – 2813

[29] Babu, R.V., Perkis, A. “An hvs-based no reference perceptual quality assessment of jpeg coded images using neural networks,” In: Proc. of IEEE Int. Conference on Image Processing, pp. 433–436 (2005).

[30] Carnec, M., Le Callet, P., Barba, D., “Objective quality assessment of color images based on a generic perceptual reducedreference,” Signal Processing and Image Communication 4, 239–256 (2008).

[31] Cavallaro, A.,Winkler, S.,” Segmentation-driven perceptual quality metric,” In: Proc. of IEEE Int. Conference on Image Processing, pp. 3543–3546 (2004).

[32] Chai, Ngan, K., “Face segmentation using skin-color map in videophone applications,” IEEE Trans. on Circuits and Systems for Video Technology 9, 551–564 (1999)

[33] Farias, M., Mitra, S.,” No-reference video quality metric based on artifact measurements” In: Proc. of the IEEE International Conference on Image Processing, vol. 3, pp. 141–144 (2005)

[34] J.Yang, W.Lu, Waibel, A., “Skin-color modeling and adaptation,” In: Asian Conf. Computer Vision, vol. 2, pp. 687–694 (1998)

[35] Oelbaum, T., Keimel, C., Diepold, K., “Rule-based no-reference video quality evaluation using additionally coded videos,” IEEE Journal of selected topics in signal processing 3, 294–307 (2009)

[36] Ramanarayanan, G., Bala, K., Ferwerda, J., “Perception of complex aggregates,” IEEE Journal of selected topics in signal processing 3, 294–303 (2009)

[37] Shao, L., Zhang, H., de Haan, G., “An overview and performance evaluation of classification-based least squared trained filters,” IEEE Transaction on Image Processing 17, 1772–1782 (2008)

[38] Sheikh, H., Bovik, A., Cormack, L., “No-reference quality assessment using natural scene statistics: Jpeg2000,” IEEE Transactions on Image Processing 14, 1918–1927 (2005)

[39] Wang, Z., Sheikh, R., Bovik, A., “No reference perceptual quality assessment of jpeg compressed images,” In: Proc. of IEEE Int. Conference on Image Processing, pp. 477–480 (2002)

[40] Wu, H., Yuen, R., “A generalized block-edge impairment metric for video coding,” IEEE Signal Processing Letters 4, 317–320 (1997)

[41] Yang, F., Wan, S., Chang, Y., Wu, H.R., “A novel objective noreference metric for digital video quality assessment,” IEEE Signal Processing Letteers 12, 685–688 (2005)

[42] Bravo, M. and Farid, H., “ A scale invariant measure of clutter,” Journal of Vision 23, 1-9 (2008).

[43] F. Dardi, L. Abate, and G. Ramponi: "A set of features for measuring blurriness in video frames", Melecon 2010, 15th IEEE Mediterranean Electrotechnical Conference, Valletta, Malta, 26-28 April 2010.

[44] L. Abate, F. Dardi, and G. Ramponi: "Blurriness estimation in video frames: a study on smooth objects and textures", in Proceeding of the SPIE Electronic Imaging Conference, San Jose (CA) USA, (2010).

[45] F. Dardi, L. Abate, J. Stessen and G. Ramponi, “Causes and visual experience of blurriness in video frames,” sottomesso a Signal Processing: Image Communication, Elsevier.

[46] Utsumi, A. and Tetsutani, N., “Human detection using geometrical pixel value structures,” Fifth IEEE International Conference on Automatic Face and Gesture Recognition, 2002. Proceedings.

[47] R. Gonzalez and R. Woods, “Digital Image Processing”, Pearson Prentice Hall, Upper Saddle River, third edition, 2008.

[48] F. Cervelli, F.Dardi, and S. Carrato: "A translational and rotational invariant descriptor for automatic footwear retrieval of real cases shoe marks", Eusipco 2010, European Signal Processing Conference, Aalborg, Denmark, 23-27 August 2010.

[49] F. Dardi, F. Cervelli, and S. Carrato, "A combined approach for footwear retrieval of crime scene shoe marks", in Proc. ICDP-09, 3rd International Conference on Imaging for Crime Detection and Prevention, (London (UK)), Dec. 2009. ISBN: 978-1-84919- 207-1, paper No. P09.

[50] F. Cervelli, F. Dardi, and S. Carrato, "A texture recognition system of real shoe marks taken from crime scenes", in Proc. 2009 IEEE International Conference on Image Processing, (Cairo (Egypt)), pp. 29052908, Nov. 2009.

[51] ENFSI Working Group on Marks website. www.intermin.fi/intermin/hankkeet/wgm/home.nsf/.

[52] F. Dardi, F. Cervelli, and S. Carrato, "A texture based shoe retrieval system for shoe marks of real crime scenes", in proc. ICIAP09, (Vietri sul Mare (Salerno, Italy)), pp. 384-393, Sept. 2009.

[53] F. Cervelli, F. Dardi, and S. Carrato, "Comparison of footwear retrieval systems for synthetic and real shoe marks", in Proc. ISPA09, 6th International Symposium on Image and Signal Processing and Analysis, (Salzburg (Austria)), pp. 684689, Sept. 2009.

[54] F. Dardi, F. Cervelli, and S. Carrato, "A full automatic footwear retrieval system for shoe marks from real crime scenes", in Proc. ISPA09, 6th International Symposium on Image and Signal Processing and Analysis, (Salzburg (Austria)), pp. 668672, Sept. 2009.

[55] G. Algarni and M. Amiane, “A novel technique for automatic shoeprint image retrieval,” Forensic Sci. Int., 181:10–14, 2008.

[56] W. Ashley, “What shoe was that? The use of computerized image database to assist in identification,” Forensic Sci. Int., 82:7–20, 1996.

[57] A. Bouridane, A. Alexander, M. Nibouche, and D. Crookes, “Application of fractals to the detection and classification of shoeprints,” In Proc. Int. Conf. Image Processing, volume 1, pages 474–477, 2000.

[58] P. Brodatz, “Textures: a photographic album for artists designers,” Dover, New York, 1966.

[59] N. Sawyer, “SHOE-FIT a computerized shoe print database,” In Proc. Eur. Convention Secur. Detect., pages 86–89, 1995.

[60] P. De Chazal, J. Flynn, and R. Reilly, “Automated processing of shoeprint image based on the Fourier transform for use in forensic science,” IEEE Trans. Pattern Analysis Machine Intelligence, 27:341–350, 2005.

[61] European Network of Forensic Science Institutes. Website: www.enfsi.eu.

[62] Z. Geradts and J. Keijzer, “The image-database REBEZO for shoeprint with developments for automatic classification of shoe outsole designs” Forensic Sci. Int., 82:21–31, 1996.

[63] A. Girod, “Computerized classification of the shoeprints of burglars’ shoes,” Forensic Sci. Int., 82:59–65, 1996.

[64] A. Girod, “Shoeprints: coherent exploitation and management,” In European Meeting for Shoeprint Toolmark Examiners, The Netherlands, 1997.

[65] M. Gueham, A. Bouridane, and D. Crookes, “Automatic recognition of partial shoeprints based on phase-only correlation,” In Proc. Int. Conf. Image Processing, volume 4, pages 441–444, 2007.

[66] Zhou Wang nd Alan C. Bovik, “Reduced – and No – Reference Image Quality Assessment”, IEEE Signal Processing Magazine, vol. 28, n. 6, 2011.

[67] M. Pavlou and N. Allinson, “Automated encoding of footwear patterns for fast indexing,” Image Vision Computing, 27:402–409, 2009.

[68] Ulrich Engelke, Hagen Kaprykowsky, Hans-Jürgen Zepernick and Patrick Ndjiki-Nya, “Visual Attention in Quality Assessment”, IEEE Signal Processing Magazine, vol. 28, n. 6, 2011.