

Elaborazione mediante FPGA di immagini ad ampio range dinamico

Niladri Roy
Lattice Semiconductor

Gli odierni FPGA a basso costo e consumi ridotti sono in grado di gestire l'elevato carico di elaborazione delle odierne videocamere

L'elaborazione dei segnali di immagine video (ISP - Image Signal Processing) ha compiuto grandi progressi dai tempi dei segnali analogici. Oggi, l'elaborazione digitale dei segnali rende possibile la manipolazione dei dati di immagine a livello di bit, offrendo un controllo senza precedenti sulla qualità delle immagini. Mentre i DSP hanno goduto di un'ampia diffusione per l'elaborazione dei segnali di immagine video in campo digitale, l'ISP può essere realizzata sfruttando una vasta gamma di dispositivi di elaborazione: DSP, ASIC, ASSP e, in misura sempre maggiore, dagli FPGA.

Esistono diverse ragioni alla base del crescente utilizzo degli FPGA. Due di queste riflettono le più recenti tendenze nelle videocamere di sicurezza, che richiedono un numero decisamente più elevato di dati di immagine da elaborare, mentre la terza è di natura più prettamente economica - la riduzione del numero di componenti richiesti (BOM) e quindi del costo della videocamera.

Sono due le principali tendenze che stanno modificando radicalmente la modalità di realizzazione della videocamera di sicurezza:

1. l'avvento dei sensori megapixel;
2. la necessità di un range dinamico elevato (o ampio) (HDR/WDR - High/Wide Dynamic Range).

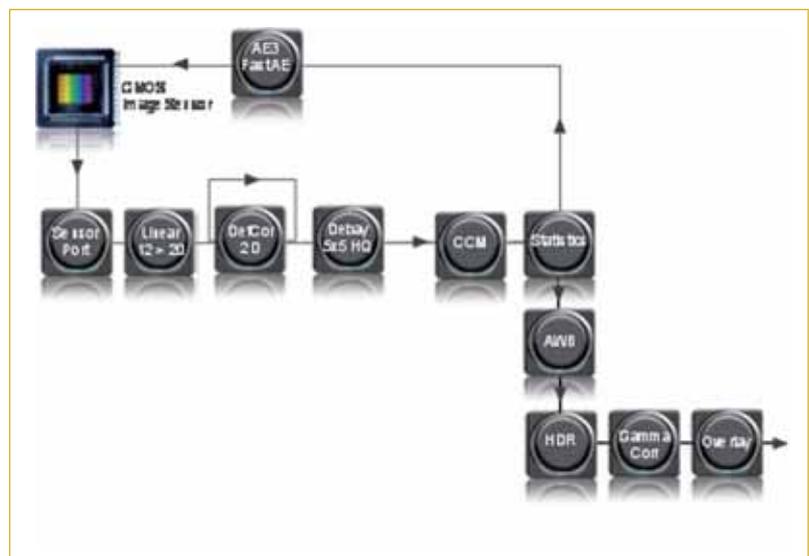


Fig. 1 - Sequenza di elaborazione dei segnali di immagine

Sensori megapixel

Fino a non molto tempo fa nelle videocamere di sicurezza - controllate da un operatore o le cui riprese venivano archiviate per un controllo successivo - era sufficiente un sensore con risoluzione VGA. A causa dell'aumento esponenziale del numero delle videocamere di sicurezza usate nel mondo - con la conseguente impossibilità di controllarle singolarmente con un operatore - l'industria della sicurezza ha iniziato a utilizzare opportuni software per analizzare il video, in tempo reale o in un momento successivo, e scoprire se si sono verificate situazioni anomale nella "regione di interesse". Sono stati svilup-

Tabella 1 - Utilizzo delle risorse per una pipeline ISP nell'FPGA ECP3-35 FPGA di Lattice

Blocco ISP	LUT	EBR 18 kbit	Moltiplicatore 18 X 18
Porta sensore con correzione automatica del livello di nero	1915		
Linearizzazione 12>20	100		
Correzione del Pixel Difettoso & Riduzione del Rumore 2D (2DNR)	1539	12	
De-Bayer 5X5 HQ	1959	12	
Matrice di correzione dei colori (CCM)	1712		18
Unità statistica	2000		
Auto esposizione (AE) veloce	-	-	-
Bilanciamento del bianco & controllo della saturazione automatico (AWB)	2167	2	8
Range dinamico elevato (HDR)	1612	12	16
Correzione della gamma	650	1	2
Sovraimpressione del Logo	336	5	3
Lattice Mico 32	1800	22	2
I2C Master	132		
Livello fisico HDMI di Lattice	276		
TOTALE	16198	66	49
Percentuale di risorse dell'ECP3-35 utilizzate	49,08%	91,67%	76,56%

te. Si prendano ad esempio le foto amatoriali all'aperto scattate con il sole alle spalle delle persone che appaiono in fotografia. Mentre lo sfondo immerso nella luce solare è luminoso e chiaro, i volti delle persone sono molto scuri. Questo succede perché la videocamera (generalmente automatica) regola la sua esposizione in funzione della luce del sole presente nella scena. L'esposizione, tuttavia, è troppo breve per registrare adeguatamente gli oggetti più scuri. Se si impostano

pati algoritmi sofisticati di analisi del video (VA - Video Analytics) per evidenziare i dettagli anomali e distinguerli da quelli ordinari; tuttavia, per essere efficaci, questi algoritmi hanno bisogno di molti più dettagli di quelli che possono essere forniti dalle videocamere con risoluzione VGA.

Le videocamere richiedono una risoluzione superiore affinché l'analisi video sia in grado di discernere il movimento generale in aree ristrette e/o in grandi aree, ad esempio un'area parcheggio può portare al sovraccarico. Una videocamera ha bisogno di circa 30 pixel/pollice per il riconoscimento di una targa, e di circa 150 pixel /pollice per visualizzare attività in modo più dettagliato, come l'identificazione delle transazioni al registratore di cassa. Un megapixel copre i dettagli in un'area di 2,13 metri per 2,13 metri, e occorrerebbero quattro videocamere VGA per eguagliare la potenza di una videocamera megapixel.

Sono stati sviluppati e sono commercialmente disponibili sensori di immagine con risoluzione pari a uno, due, cinque e persino 10 megapixel. Ovviamente, all'aumentare del numero dei pixel, cresce la quantità di dati che devono essere elaborati per sfruttare i vantaggi offerti dalla risoluzione superiore.

HDR - High Dynamic Range: concetti base

L'HDR, noto anche come range dinamico ampio (WDR) misura la qualità della visione del sensore e della funzione ISP sia in aree scure, sia in aree chiaramente illumina-

te. Si prendano ad esempio le foto amatoriali all'aperto scattate con il sole alle spalle delle persone che appaiono in fotografia. Mentre lo sfondo immerso nella luce solare è luminoso e chiaro, i volti delle persone sono molto scuri. Questo succede perché la videocamera (generalmente automatica) regola la sua esposizione in funzione della luce del sole presente nella scena. L'esposizione, tuttavia, è troppo breve per registrare adeguatamente gli oggetti più scuri. Se si impostano

manualmente l'esposizione o l'apertura per lasciare passare più luce, sarà possibile distinguere i dettagli nelle aree scure, ma a discapito dei dettagli nelle aree luminose, che ora sono sovraesposte e completamente schiarite. Questo non è un buon risultato sia per gli operatori umani, sia per il software VA, dato che si perde gran parte dei dettagli nella regione di interesse. I sensori HDR risolvono questo problema in modi creativi, tutti basati sulla cattura di più immagini, ciascuna delle quali con tempi di esposizione diversi; in seguito la sequenza di elaborazione ISP abbina e miscela in modo adeguato queste immagini per preservare e rendere visibili i dettagli sia delle aree luminose sia di quelle scure nella regione di interesse. Ovviamente, più esposizioni della stessa immagine si traducono in una maggiore quantità di dati da elaborare. Ad esempio, quando una videocamera che produce immagini a 60 fotogrammi al secondo funziona con un sensore HDR che acquisisce 3 esposizioni per fotogramma, la sequenza di elaborazioni ISP all'interno della videocamera sta in realtà elaborando l'equivalente di 60 x 3 - o 180 - fotogrammi al secondo.

Gli FPGA e l'aumento del carico di elaborazione

Un sensore megapixel e l'HDR insieme contribuiscono ad aumentare notevolmente il carico di elaborazione della pipeline ISP. I dispositivi DSP, essendo "unità sequenziali" per natura, hanno difficoltà a gestire questo carico enorme di elaborazione dati. Potrebbe essere ancora

I VANTAGGI DELL'USO DI UN FPGA

Di seguito, in sintesi, i principali vantaggi dell'elaborazione del segnale immagine HDR. **Basso costo** - Un FPGA a basso costo da 33.000 LUT gestisce facilmente una sequenza di 1080p60. Il BOM per una videocamera HDR da 1080p60 realizzata con un dispositivo ECP3-35 di Lattice risulta composta da sensore, FPGA e l'associato oscillatore di clock, resistori e condensatori, un regolatore di tensione, un connettore HDMI e l'insieme delle lenti.

Alte prestazioni - La realizzazione qui riportata è caratterizzata da HDR di 120dB, prestazioni da 1080p60, esposizione automatica estremamente veloce e bilanciamento automatico del bianco di qualità molto alta.

Consumi ridotti - Il consumo di potenza statica e dinamica del dispositivo LatticeECP3 è significativamente inferiore rispetto a quello di analoghi FPGA o DSP attualmente disponibili sul mercato.

Supporto DDR3 - L'FPGA supporta l'uso delle memorie DDR3. I produttori che desiderano incorporare una memoria temporanea per il video all'interno dei propri progetti possono trarre vantaggio da questa funzionalità per utilizzare della memoria DDR3 ad alte prestazioni e a basso costo all'interno dei propri progetti di videocamera.

SERDES a basso consumo - Un FPGA a basso consumo con funzionalità SERDES consente ai produttori di realizzare uno strato fisico HDMI direttamente all'interno dell'FPGA, fornendo la funzionalità HDMI senza l'onere di un chip HDMI esterno.

possibile elaborare i dati relativi all'esempio riportato poco sopra - che prevede una pipeline HDR 1080p60 - con un DSP di fascia alta, a fronte però di costi e con consumi di potenza proibitivi e improponibili in termini commerciali. Gli FPGA, grazie al loro parallelismo intrinseco, risultano quindi particolarmente adatti a gestire il crescente carico di elaborazione dei segnali di immagine ad alto intervallo dinamico e ad alta definizione.

L'importanza della programmabilità

Oltre a garantire alte prestazioni, consumi ridotti e bassi costi, gli FPGA sono per definizione programmabili, fatto questo che si traduce in vantaggi significativi rispetto agli ASIC e agli ASSP. Gli ASIC sono estremamente costosi da progettare e da realizzare e, una volta realizzati, non possono essere modificati. Le fotocamere basate su ASSP, dal canto loro, possono essere limitate in termini di funzionalità - limitazioni già intrinseche nei componenti standard utilizzati - e non possono essere modificate. Non a caso, diversi DSP e altri dispositivi ASSP destinati al mercato dell'elaborazione dei segnali di immagine video necessitano di un FPGA che svolga la funzione di "ponte" tra il sensore e il componente standard, allo scopo di gestire le nuove interfacce seriali che i produttori di sensori usano per ottenere dati con risoluzione dell'ordine dei megapixel dai propri sensori. Con una realizzazione basata su FPGA, i produttori di vide-

ocamere possono trarre vantaggio dalla programmabilità per adeguare in tempi brevi i loro progetti alle esigenze dei nuovi sensori e delle nuove tecnologie, oppure modificare altrettanto rapidamente i propri algoritmi ISP.

Implementazione degli algoritmi ISP con HDR in un FPGA

Allo scopo di implementare un algoritmo ISP con HDR in un FPGA è necessario prevedere i blocchi ISP nella sequenza di elaborazione dei segnali di immagine riportati nella figura 1.

I blocchi ISP richiesti sono:

Porta del sensore, con correzione automatica del livello di nero - Questa è richiesta per individuare e configurare i registri del sensore di

immagine e per catturare i dati di immagine.

Correzione del livello di nero - Ciascun canale di colore (color channel) è caratterizzato da un offset dipendente dal tempo. L'elaborazione del colore richiede un comportamento lineare del segnale, per cui ogni segnale deve essere privo di offset. I sensori di immagine CMOS hanno le cosiddette uscite con linee scure per misurare l'offset medio per ciascun canale di colore. La correzione del livello di nero sottrae il rumore specifico del canale di colore e dipendente dalla linea per ottenere un risultato ottimale in termini di livello di nero.

Esposizione automatica - Lo scopo del blocco di esposizione automatica è quello di regolare costantemente l'esposizione per adattarsi in tempo reale alle variazioni delle condizioni di luce.

Linearizzazione - Il sensore HDR Aptina MT9Mo24/34, ad esempio, fornisce 20 bit di informazione per canale di colore. Allo scopo di minimizzare il numero di linee fisiche in uscita dal sensore, Aptina sfrutta uno schema di compressione "intelligente" per comprimere questi dati a 12 bit. La linearizzazione è il processo di decompressione di questi dati da 12 bit al fine di recuperare i 20 bit originari.

Correzione dei pixel difettosi - I pixel spenti ("dead" pixel) o surriscaldati presenti nel sensore a causa dei processi di fabbricazione sono corretti attraverso il blocco di correzione dei pixel per difetto. Questo blocco effettua la correzione dei pixel difettosi mediante interpolazione con

i pixel posti nelle vicinanze nello stesso canale di colore. I metodi tipici di correzione includono la rivelazione dei pixel freddi o caldi usando una stima (media o mediana) sui pixel immediatamente vicini.

Riduzione del Rumore 2-D - A parte i pixel freddi e caldi, i pixel del sensore possono essere caratterizzati da una rumorosità - di tipo casuale - passando da un fotogramma all'altro. Questo significa che essi emettono un'intensità troppo alta o troppo bassa rispetto a quella dei pixel vicini. La correzione del rumore 2D corregge i pixel rumorosi mediante interpolazione con i pixel vicini dello stesso canale di colore, adottando essenzialmente la medesima procedura della correzione dei pixel difettosi.

De-Bayering (interpolazione della matrice di filtri di colore) - Ciascun pixel sul sensore è dotato del cosiddetto filtro Bayer con uno dei tre colori: rosso, verde o blu. Due terzi dei dati sui colori sono di conseguenza assenti da ciascuno di essi e l'immagine risultante è un mosaico dei tre colori. Per ottenere un'immagine a pieni colori, sono stati usati diversi algoritmi di demosaicizzazione (De-mosaic) per interpolare un insieme di valori completi di rosso, verde e blu per ciascun pixel.

Matrice di correzione dei colori (CCM) - I sensori di immagine forniscono spesso una resa non corretta dei colori a causa dei cosiddetti effetti di colore incrociato che sono causati dalla diafonia di segnale fra pixel. Questo effetto porta a immagini di colore scorrette (ad esempio verde con troppo blu). La correzione del colore comporta una moltiplicazione di una matrice complessa di dati relativi ai pixel per ottenere colori chiari.

Bilanciamento automatico del bianco (AWB) - I sensori non si prestano bene a "riconoscere" i colori. L'AWB regola gli altri colori in un'immagine rispetto a un colore bianco ricavabile nell'immagine attraverso un algoritmo conosciuto come "grey world". L'AWB determina il bianco esaminando la frequenza (e quindi la lunghezza d'onda) della luce incidente producendo l'immagine con colori naturali.

Correzione della gamma - I pixel del sensore reagiscono all'intensità della luce incidente in modo lineare.

Allo scopo di fornire i dati sui pixel ai sistemi video più comuni, come ad esempio un tubo CRT con risposta logaritmica potrebbe essere necessaria la conversione a una codifica. La correzione della gamma fornisce questa conversione.

Elaborazione con intervallo dinamico elevato/ampio (HDR/WDR) - Questo è il blocco responsabile della mappatura di 20 pixel di dati del sensore in una rappresentazione

RBG a 8 bit in modo da rendere sia le aree luminose sia quelle scure dell'immagine visibile in un'immagine renderizzata. È richiesta una pipeline interna di ampie dimensioni per assicurare che non venga perso alcun dettaglio nelle aree scure (Fig. 2). L'elaborazione HDR, operando in stretta relazione con un algoritmo rapido di esposizione automatico, è in grado di regolare l'esposizione al variare delle condizioni di luce.

La tabella 1 riporta le risorse FPGA tipiche usate per rea-

lizzare tutti i sopra citati blocchi ISP in un FPGA a basso costo e a basso consumo con 33.000 Look-Up-Table (KLUT).

Oltre ai blocchi ISP già citati, per l'implementazione effettiva sono necessari un'unità statistica che genera istogrammi di immagine usati da blocchi specifici nel sistema, un processore soft Mico32 di Lattice per il controllo dinamico della sequenza, un master I2C per controllare vari segnali, un blocco fisico HDMI per pilotare i segnali HDMI direttamente all'esterno dell'FPGA e persino la sovraimpressione grafica di un logo. Questo dimostra che è possibile integrare un'intera sequenza di elaborazione dei segnali di immagine, oltre a un'uscita HDMI, in un FPGA a basso costo e a basso consumo come il dispositivo Lattice ECP3-35. La pipeline HDR interna, di ampiezza pari a 32 bit, fornisce un range dinamico di 192dB ($20 \log 2^{32}$). In questa implementazione è stato usato un sensore con un intervallo dinamico di 120dB, limitando l'HDR a 120dB - che risulta comunque ancora il più elevato per una realizzazione basta su FPGA. Questa implementazione pratica è in grado di elaborare immagini con risoluzione di 1080p a 60 fotogrammi al secondo fornendo un HDR di 120 dB. ■



Fig. 2 - Immagine elaborata con la tecnica HDR: si noti la mancanza di fenomeni di oscuramento delle aree che si trovano dietro a una potente torcia accesa puntata sulle lenti a una distanza di 25 cm