DISS. ETH NO. 24319

TWO-STREAM VISION SENSORS

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES of ETH ZURICH
(Dr. sc. ETH Zurich)

presented by

Chenghan Li
M. sc. ETH Zurich
born on 05 December 1986
citizen of China

accepted on the recommendation of

Prof. Tobias Delbruck
Prof. Kevan Martin
Prof. Bernabe Linares-Barranco
PD. Dr. Shih-Chii Liu

24 May 2017

Abstract

This thesis explores the why and how of two-stream vision sensors. The physics of visual transduction poses fundamental limits on a vision system – a vision system dedicated to feature perception cannot perform well in motion perception and vice versa. Studies of mammalian vision systems have taught us that perception of both feature and motion demands a two-stream vision sensor: one stream with high spatial resolution, high signal-to-noise ratio and color sensitivity to perceive features, the other with high temporal resolution and low latency to perceive motion.

Through surveying prior works, the state-of-the-art frame-based active-pixel sensor (APS) technology was found well suited for feature perception because of its compact size, high signal-to-noise ratio and color sensitivity through the use of a color filter array (CFA). On the other hand, the neuromorphic event-based dynamic vision sensor (DVS) technology was found well suited for motion perception because of its high temporal resolution, low latency and sparse output. Combining the APS and DVS technology, the dynamic and active-pixel vision sensor (DAVIS) developed in 2013 produced concurrent intensity frames and temporal contrast events, however, both of which were monochromatic and had the same spatial resolution. Following the exploratory step made by the DAVIS, this thesis presents the design and silicon results of the dynamic and color active-pixel vision sensor (CDAVIS), the first two-stream vision sensor that produces concurrent color VGA frames and monochromatic QVGA events.

The CDAVIS employed a heterogeneous pixel array consisting both state-of-the-art APS pixels and modified DAVIS pixels. The compact size of the additional APS pixels allowed the CDAVIS to achieve 7.1 times more pixels of the prior DAVIS for the frame output with only doubled pixel array area. Color sensitivity was achieved through the addition of CFA. The CDAVIS chip was fabricated in Towerjazz 0.18um CMOS image sensor technology. The frame output of the CDAVIS was characterized based on European Machine Vision Association (EMVA) Standard 1228. To standardize the event output characterization method, this paper proposes a set of event output measurement and analysis methods. The event output of the CDAVIS was characterized with the proposed methods. For better benchmarking, a prior DAVIS chip was fully characterized with the same methods. The results showed the frame output of the CDAVIS had better linearity, absolute sensitivity threshold, dynamic range, and consistency in SNR than the prior DAVIS. Both sensors had comparable event output performance. However, the CDAVIS had a lower quantum efficiency (QE) at 3.3% (G pixel at 550nm), due to more complex in-pixel circuits.

To demonstrate the advantages of the concurrent color VGA frames and fast sparse events, this thesis elaborates on one proof-of-concept application example, towards the vision system of RoboCup Small-Sized League (SSL) soccer using the CDAVIS. A Java-based feature detection and tracking algorithm called the Red-Dot-Tracker was built to perform a simple task of tracking red dots on green background using the CDAVIS. The CDAVIS exhibited the capability to support real-time color-based object tracking with about 15ms shorter worst-case latency than using a 60fps (frames per second) camera. When applying the Red-Dot-Tracker on a real-time playback of the CDAVIS recording, the event output produced less than 1/10 of data traffic and

required 5~10 times less processing power than the frame output in the same Java environment.

Lastly, this thesis presents preliminary results from two pieces of ongoing development efforts on two-stream vision sensors. Firstly, aiming to reduce spatial redundancy in the event output of a two-stream vision sensor, this thesis proposes a modified DVS design implementing center-surround receptive field topology. The center-surround DVS consists of a novel compact photoreceptor design with two antagonistic outputs, a diffuser network, and a sum-differencing amplifier. Based on simulation studies, this thesis discusses the feasibility of the proposed center-surround DVS design and outlines the next steps to take this concept further.

Secondly, to address the low QE issue that plagued neuromorphic vision sensors with complex in-pixel circuits, this thesis presents preliminary silicon results of the first two-stream vision sensor implemented in BSI technology. Through side-by-side comparison of the backside illuminated (BSI) and front-side illuminated (FSI) versions of an identical DAVIS design, it was found that BSI quadrupled QE of the DAVIS pixel. However, the BSI version had more inter-pixel crosstalk and was more vulnerable to parasitic light. This thesis discusses further works needed to quantitatively understand the effects of BSI on two-stream sensors.

Abstract

Questo lavoro di tesi esplora il perché e il come dei sensori di visione a doppia uscita. La fisica della trasduzione visiva pone limiti fondamentali nei sistemi di visione - un sistema di visione dedicato alla percezione di caratteristiche visive non percepisce il movimento in maniera accurata e vice versa. Lo studio dei sistemi di visione dei mammiferi ci ha insegnato che la percezione di caratteristiche visive e del moto richiede sensori a due flussi di informazione: un flusso ad alta risoluzione spaziale, con alto rapporto tra segnale e rumore e con sensitività ai colori per la percezione di caratteristiche visive, il secondo flusso richiede alta risoluzione temporale e bassa latenza di risposta per la percezione del movimento.

Attraverso l'analisi dei lavori precedenti, lo stato dell'arte dei sensori basati su fotogrammi e su pixel attivi (APS) si e' rivelato ideale per l'estrazione di caratteristiche visive e risulta dimensione compatta, ha un alto rapporto tra il segnale e il rumore ed e' sensibile ai colori grazie all'utilizzo di matrici a filtri colorati (CFA). D'altro canto, il sensore visivo dinamico neuromorfo basato ad eventi (DVS) si e' dimostrato adatto per la percezione del movimento in quanto ha una risoluzione temporale elevata, bassa latenza di risposta e emette informazione sparsa. La combinazione delle tecnologie APS e DVS, ovvero del pixel dinamico e attivo, nel sensore DAVIS che fu sviluppato nel 2013 e produce nello stesso momento sia fotogrammi che eventi che rappresentano il contrasto temporale. Ad ogni modo, entrambe le uscite erano monocromatiche e avevano la stessa risoluzione spaziale. Seguendo i passi esplorativi fatti con il sensore DAVIS, questa tesi presenta il design e i risultati di fabbricazione (su silicio) del sensore dinamico con pixel attivi a colori CDAVIS: il primo sensore con due uscite visive che produce concorrentemente fotogrammi a risoluzione VGA e eventi monocromatici a risoluzione QVGA.

Il sensore CDAVIS utilizza una matrice eterogenea di pixel che consiste nella realizzazione di pixel APS che sono lo stato dell'arte, e di pixel DAVIS modificati. La dimensione compatta dei circuiti APS addizionali ha permesso al sensore CDAVIS di ottenere 7.1 volte piu' pixel del suo predecessore DAVIS per l'uscita a fotogrammi a discapito di solo due volte l'area utilizzata. La sensibilita' ai colori e' stata ottenuta grazie all'utilizzo dei filtri a colori CFA. Il dispositivo CDAVIS e' stato fabbricato in un processo CMOS per sensori visivi in tecnologia Towerjazz 0.18 um. L'uscita a fotogrammi del sensore CDAVIS e' stato caratterizzato in accordo con gli standard 1228 dettati dall' "European Machine Vision Association" (EMVA). Per standardizzare i metodi di caratterizzazione dell'uscita ad eventi, questo lavoro di tesi propone un insieme di misure che di metodi di analisi dell'uscita del sensore ad eventi.

Per una migliore analisi comparativa, un chip precedente DAVIS e' stato caratterizzato con gli stessi metodi. I risultati hanno dimostrato che l'uscita a fotogrammi del sensore CDAVIS vanta di una maggiore linearita', di una soglia di sensibilita' assoluta, un piu' ampio intervallo dimanico, e un rapporto segnale rumore consistente con il suo predecessore DAVIS. Entrambi i sensori hanno comparabili prestazioni per quanto riguardo l'uscita ad eventi. Ad ogni modo, il sensore CDAVIS si e' dimostrato avere una minor efficienza quantistica (QE) a 3.3% (G pixel a 550nm), dovuta a un circuito piu' complicato del pixel stesso.

Per dimostrare i vantaggi dell'uscita concorrente a colori VGA e degli eventi sparsi, questa tesi elabora una verifica teorica di un esempio applicato, nel contesto di visione nel gioco del calcio praticato dai robot nella RoboCup Small-Sized League (SSL) utilizzando una camera CDAVIS. Un algoritmo sviluppato in Java per l'estrazione di caratteristiche visive e per il tracciamento chiamato "Red-Dot-Tracker" e' stato costruito per performare un semplice compito di tracciamento di punti rossi su sfondo verde utilizzando la CDAVIS. La camera CDAVIS ha esposto la capacità' di supportare in tempo reale e basandosi sui colori tracciamento di oggetti con una risposta di circa 15ms minore del caso peggiore di risposta utilizzando 60fps. Quando l'algoritmo "Red-Dot-Tracker" e' stato applicato alla riproduzione in tempo reale delle registrazioni del CDAVIS, gli eventi di uscita hanno prodotto meno di 1/10 di traffico dati e hanno richiesto 5~10 volte meno potenza di calcolo rispetto all'uscita del frame nello stesso ambiente Java.

Infine questa tesi presenta i risultati preliminari di due lavori in corso su sensori a due uscite visive. Il primo lavoro ha lo scopo di ridurre la ridondanza spaziale negli eventi di uscita del sensore, questa tesi propone un DVS con design modificato che implementa una topologia di campo ricettivo "center-surround'. This DVS consiste nel design di un nuovo fotorecettore compatto con due uscite antagoniste, una rete diffusore, e un amplificatore. Basandosi su studi in simulazione, questa questa tesi discute la fattibilità' del design del "center-surround' DVS descritto e descrive i prossimi passi per sviluppare questo concetto. Il secondo lavoro ha lo scopo di indirizzare il problema di basso QE che affligge i sensori di visione neuromorfi con complessi circuiti in-pixel, questa tesi presenta risultati preliminari in silico del primo sensore di visione a due canali implementato in tecnologia BSI. Attraverso confronti delle versioni de retro-illuminato (BSI) e del fronte-illuminato (FSI) di in identico DAVIS design, e' stato osservato che il BSI quadruplica il QE del pixel del DAVIS. Comunque, la versione BSI aveva piu' crosstalk a era piu' vulnerabile alla luce parassita. Questa tesi discute lavori futuri necessari per capire quantitativamente gli effetti del BSI sul sensore a due canali visivi.