

Video scene analysis for a configurable hardware accelerator dedicated to Smart-camera

Imen Charfi, Wajdi Elhamzi,
Julien Dubois, Mohamed Atri, Johel Mitéran

Le2i lab, Burgundy University (France)
EμE lab, Monastir University (Tunisia)

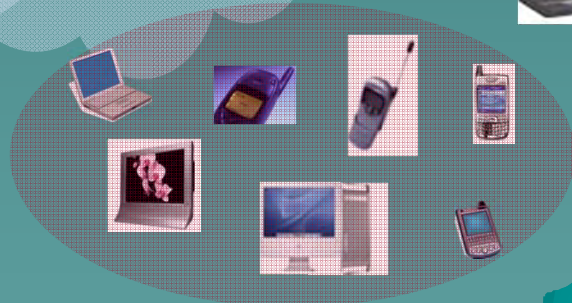
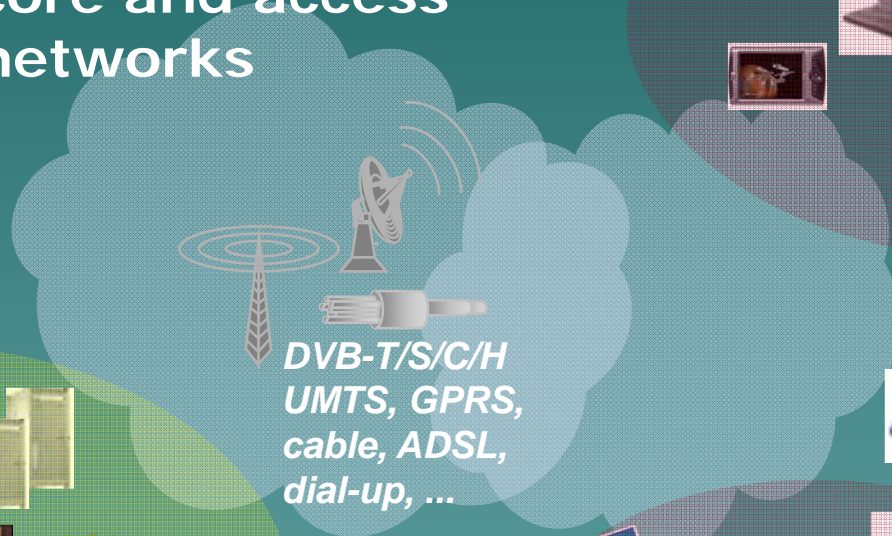
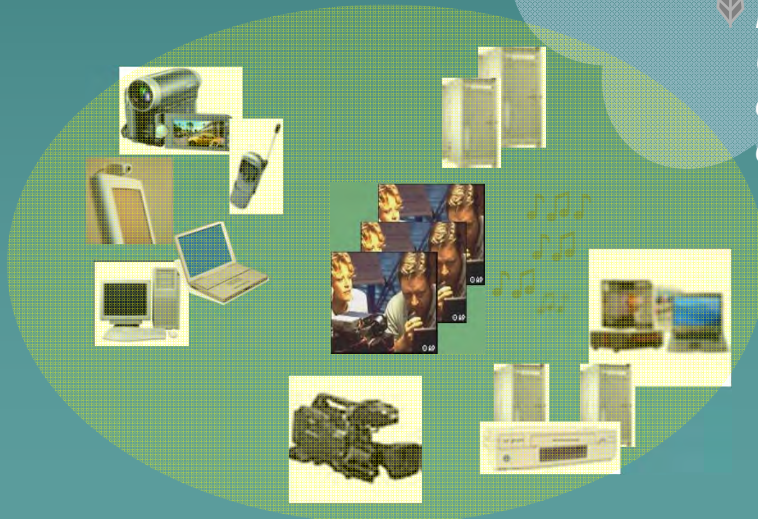


Video coding diversity

Diversity in client devices

Different types of core and access networks

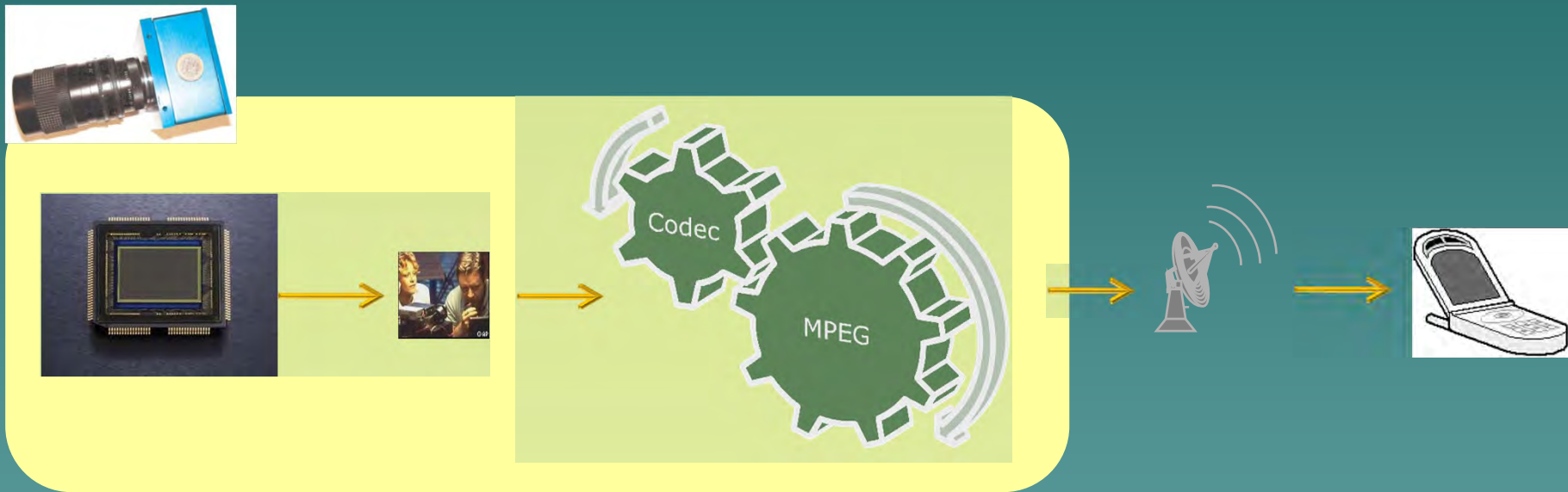
Diversity of content formats /standards



How to handle with such diversity ?

How to provide a video codec able to support such diversity?

Video codec embedded in a Smart camera



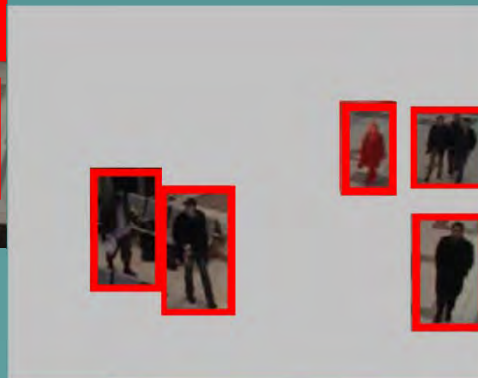
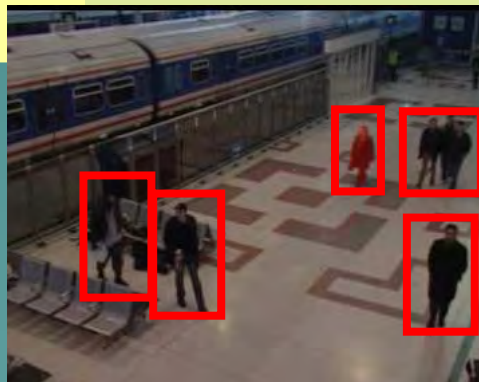
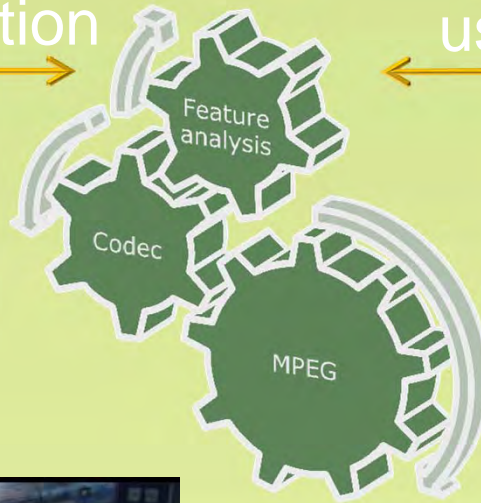
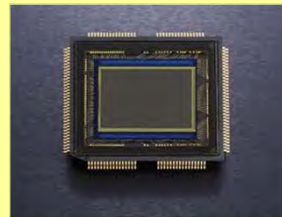
A significant contribution to handle with the diversity !
More challenging : flexibility to deal with environment constraints

Dynamic adaptation



Scene analysis :
event detection

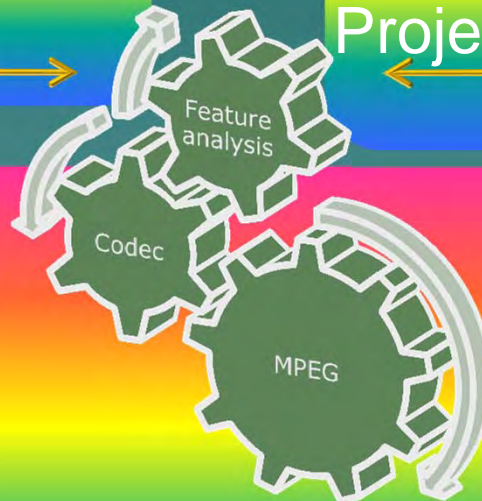
QoS analysis (network,
user configuration)



Current investigations

Scene analysis :
PhD on automatic Fall
detection

QoS analysis (network, user
configuration)
Project with industrial partner

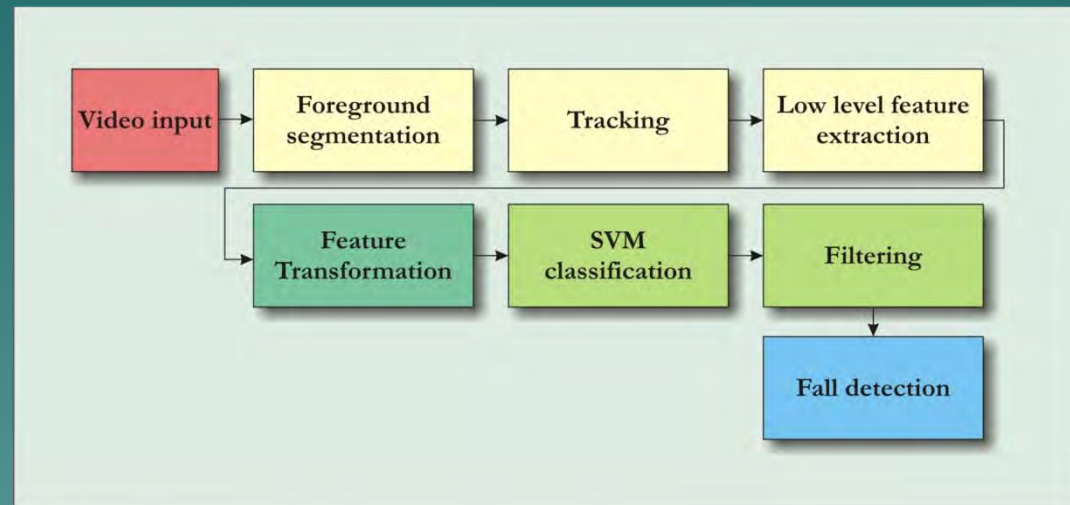


(Re)Configurable Codec
PhD on motion estimation



Automatic Fall detection

Imen Charfi's PhD



1 vector of 3584 features / frame



Data base available (60 videos, extension at 130 videos done)

Detection Results

SVM : Classification error rate 3%

	Final level error				
	Specificity	Accuracy	Precision	Recall	Global error
Fd	0.99984	0.99823	0.94444	0.62963	0.00177
Sd	0.99866	0.99466	0.66667	0.4	0.00534
FFT	0.99951	0.99577	0.84211	0.41026	0.00423
W	0.99951	0.99675	0.88235	0.4848	0.00325
Fd + Sd	0.99968	0.99806	0.89474	0.62963	0.00194
Fd + Sd + FFT	0.99984	0.99823	0.94444	0.62963	0.00177
Fd + W	0.99967	0.99886	0.9	0.78261	0.00114
Fd + FFT + W	0.99984	0.99984	0.94737	1	0.00016
FFT + W	0.99968	0.9955	0.94118	0.37209	0.00404
FFT + Sd	0.99984	0.99791	0.94444	0.58621	0.00209
Fd + Sd + W	1	0.99839	1	0.62963	0.00161
Fd+ Sd + FFT + W	0.99984	0.99823	0.94444	0.62963	0.00177

error < 0.02 %
per frame

After final filtering and Bounding box Manually annotated

	Final level error				
	Specificity	Accuracy	Precision	Recall	Global error
Fd + W	0.99967	0.99935	0.89474	0.89474	0.00065
Fd + FFT + W	0.99984	0.99951	0.94737	0.9	0.00049
Fd+ Sd + FFT + W	0.99984	0.99822	0.94444	0.62963	0.00178

error < 0.05 %
per frame



After final filtering and Bounding box automatically annotated

Detection Results

- ✓ Two preliminary evaluations show similar performances achieved :
 - on the extended video-data set (130 videos)
 - using boosting instead of SVM (from 10 to 100 time faster)
- ✓ The regularity and the complexity of boosting method enables a FPGA hardware implementation to be investigated.
Technical lock : the real-time processing of all features

Johel MITERAN, Jiri MATAS, Elbey BOURENNANE, Michel PAINDAVOINE, Julien DUBOIS
"Automatic Hardware Implementation Tool for a Discrete Adaboost-Based Decision Algorithm", EURASIP Journal on Applied Signal Processing, Hindawi, 2005 (7), pp. 1035-1046, 2005

Fethi SMACH, Johel MITERAN, Mohamed ATRI, Julien DUBOIS, Mohamed ABID, Jean Paul GAUTHIER, *"An FPGA-based accelerator for Fourier Descriptors computing for color object recognition using SVM"*, Journal of Real-Time Image Processing (JRTIP), Springer, vol.2, pp. 249-258, 2007.

Khalil Khattab, Julien Dubois and Johel Miteran, *"Cascade Boosting Based Object Detection from High Level Description to Hardware Implementation"*, EURASIP Journal of Embedded Systems, Special Issue "Design and Architectures for Signal Image Processing", Hindawi, 12 pages, 2009.

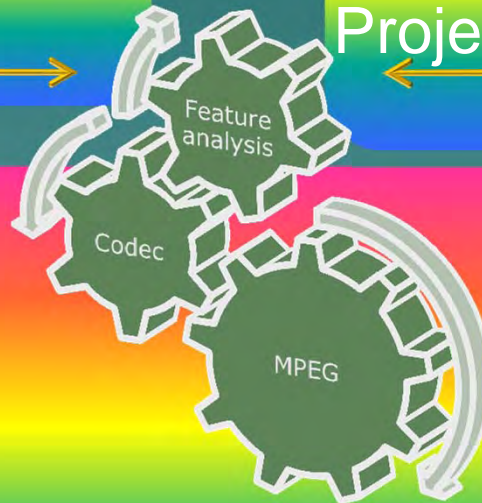


Current investigations

How to use the detection results to adjust the video coding performances ?

Scene analysis :
PhD on automatic Fall
detection

QoS analysis (network, user
configuration)
Project with industrial partner



(Re)Configurable Codec
PhD on **motion estimation**

More than 60 % of the compression

More than 60 % of the processing time

ME principle

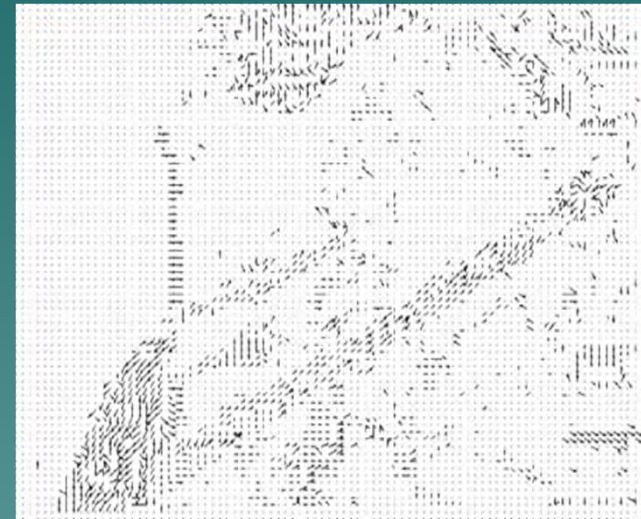
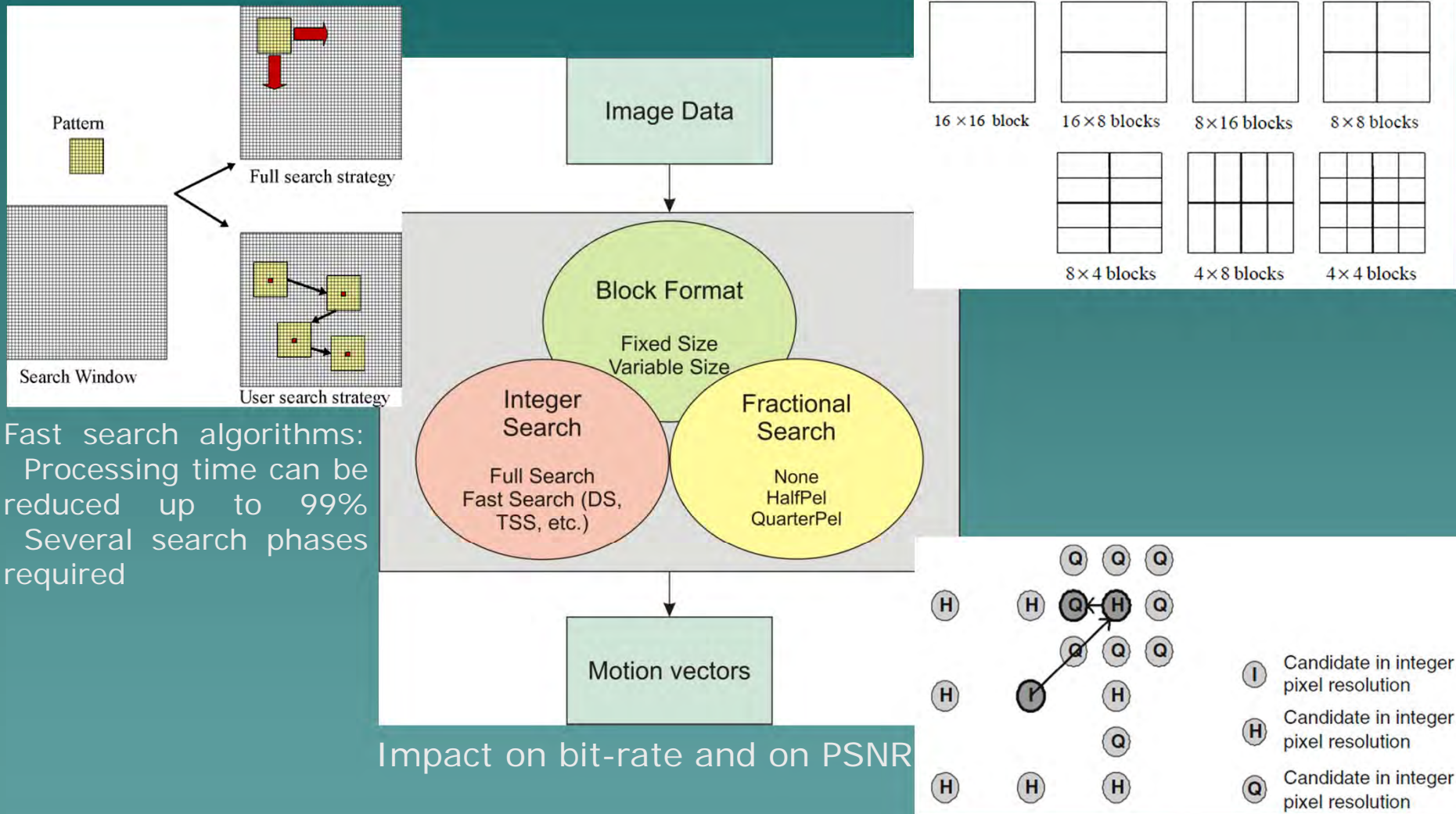


Image split in 16x16 macro-block
Each macro-block => one motion vector

How ME be flexible ?



Fast search algorithms:
Processing time can be reduced up to 99%
Several search phases required

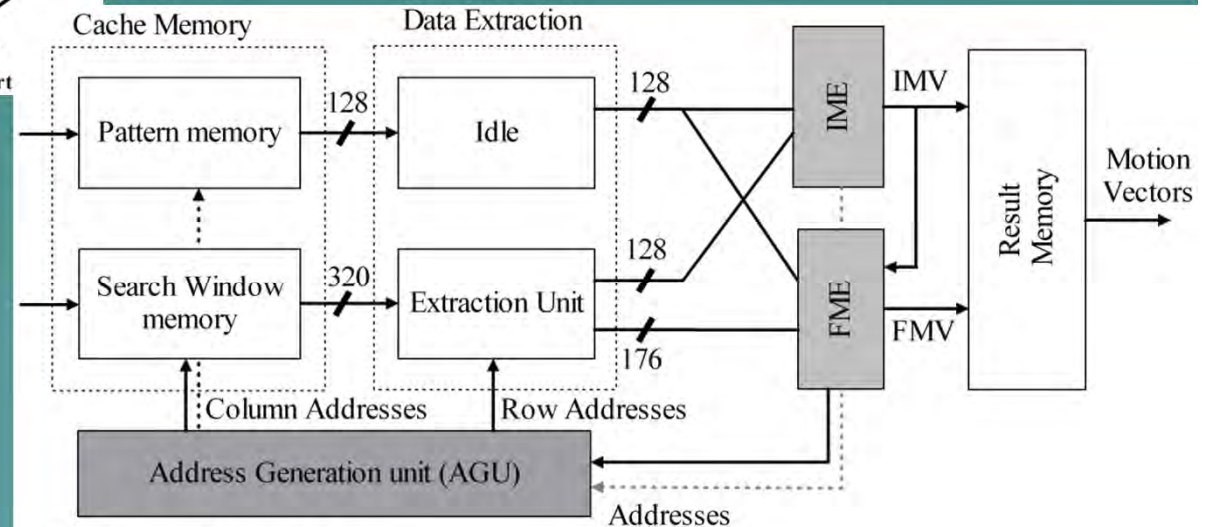
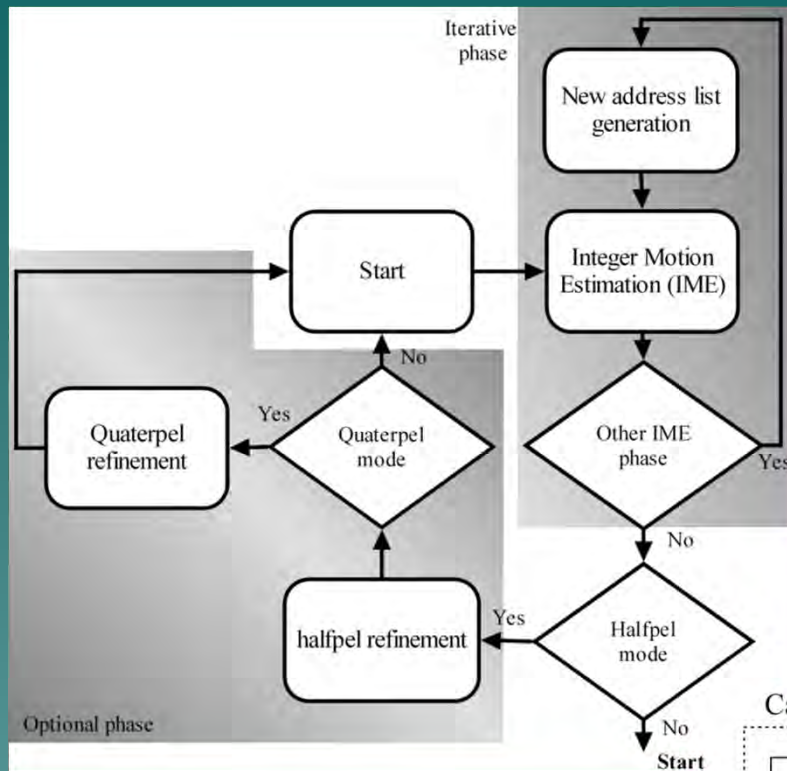
Impact on bit-rate and on PSNR

+ 4 dB (also 45 % of overall processing time)



Architecture Overview

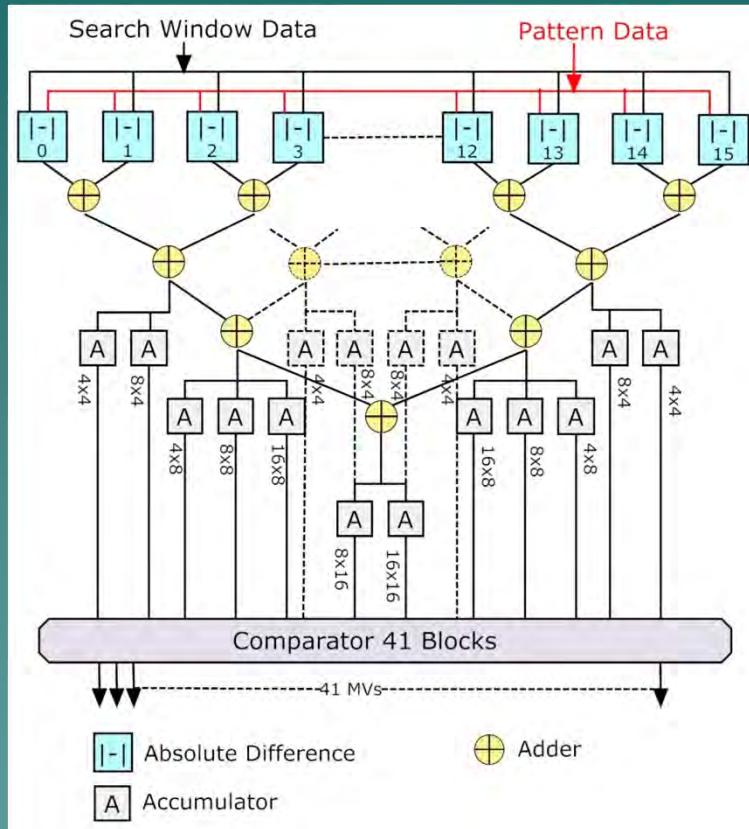
Wajdi Elhamzi's PhD



IMV : Integer Motion Vectors
FMV: Fractional Motion Vectors

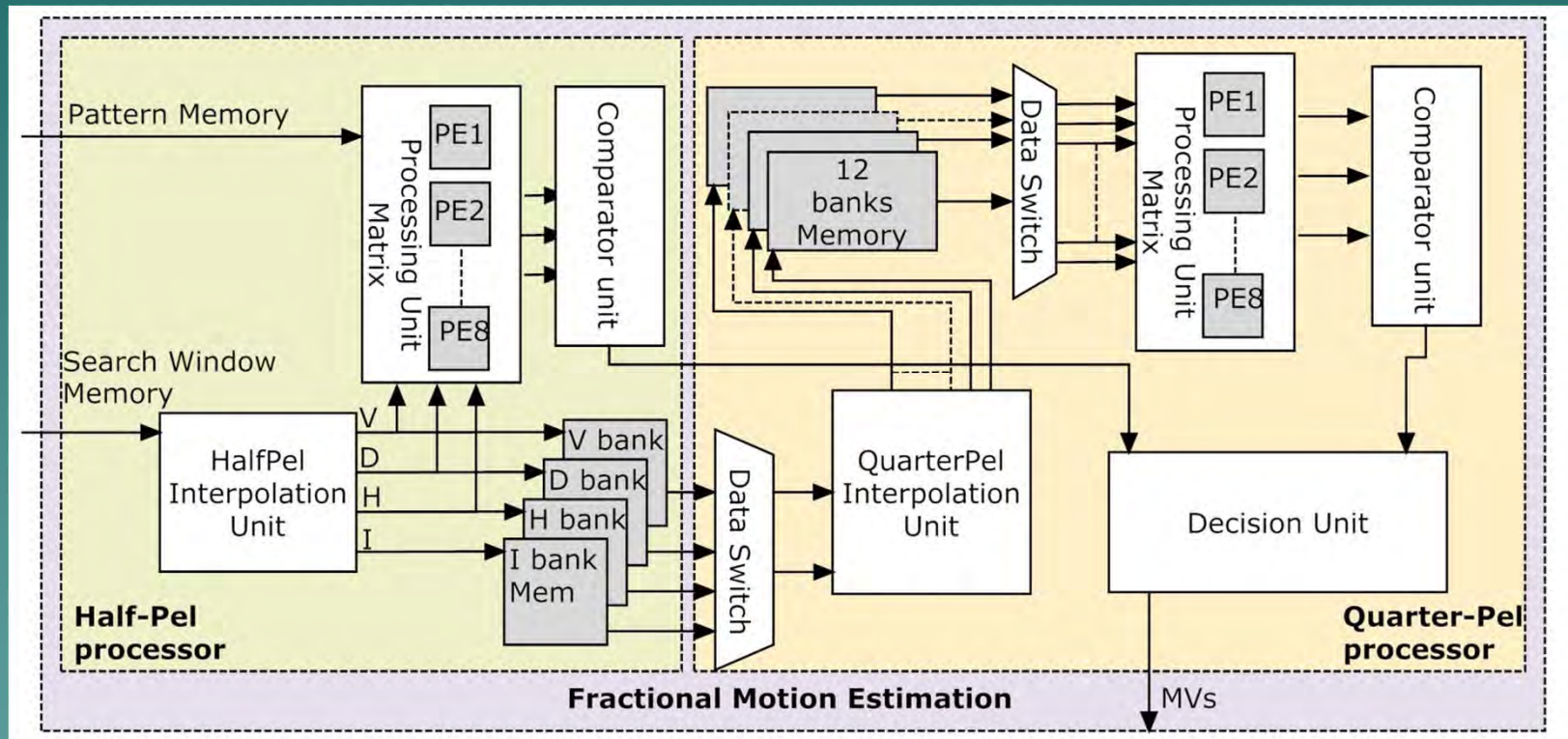


Integer ME architecture



Ref.	Year	searching range	Latency	Tech (μm)	Freq (MHz)	Max fps using FS	search strategies
[3]	2004	16x16	4096	TSMC 0.13	294	720x576@45fps	FS
[15]	2006	16x16	4096	TSMC 0.18	266	352x288@25fps	FS
[18]	2009	32x32	26624	TSMC 0.18	316	352x288@30fps	FS
Ours	2012	16x16	4096	Virtex 6vlx240 tff784-3	438	720x576@67fps	FS / DS

Halfpel & Quaterpel Co-processors



Halfpel & Quaterpel Co-processor

	Chen's [2]	Yang's [13]	Ruiz's[21]	Our's
Year	2004	2006	2010	2012
Cycles	1664	790	870	1062
Nbr of PEs	9	18	8	8
Interpolation Unit width	10	22	10	22
Tech (μm)	UMC 0.18	TSMC 0.18	UMC0.18	Virtex-6 vlx240tff784-3
Freq (MHz)	100	285	290	253
Throughput (KMB/s)	49	250	NA	232

610 cycles!!!!

1080 HD video streams at frame rate of 29 fps !!!!



Overview of the proposed implementation

Motion Estimation (Device :6v1x240tff784-3)					
Logic utilization	Used (IME/FME)		Available	Utilization (IME/FME)	
Number of slice register	1168	11944	301440	>1%	3%
Number of slice LUTs	1281	17426	150720	>1%	11%
Number of BRAMs/FIFOs	1	32	416	>1%	8%
Maximum frequency	Frequency IME : 438MHz / Frequency FME : 253MHz				

In IME : 752x576 video streams at 67 fps (in FS mode)

In FME: 1080 HD (1920x1088) video streams at frame rate of 29 fps (around 232K Macroblocks/s)

Future works :

- ✓ Improve FME (preliminary results very promising)
- ✓ Dynamic reconfiguration of the motion estimator
- ✓ Investigation on other configurable parts of the codec (DCT, quantification...)
- ✓ Design of a hardware video codec with the configurable motion estimation

Conclusion

Our contributions for a smart camera design with adaptive video coding:

- ✓ Configurable low-cost motion estimator based on FPGA component with competitive performances
- ✓ Fall detection algorithm defined and hardware implementation currently investigated



Questions ?

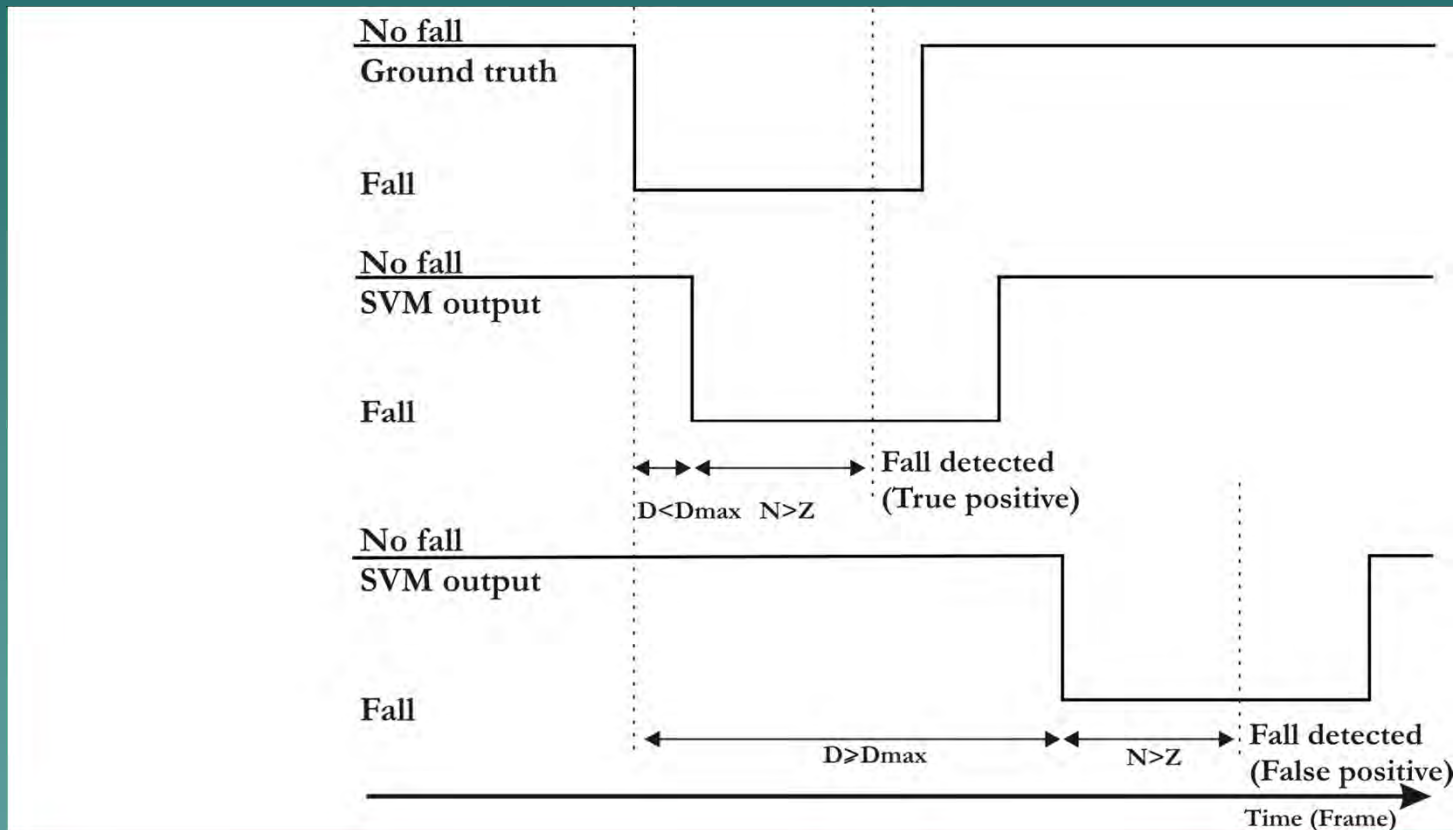


Event detection

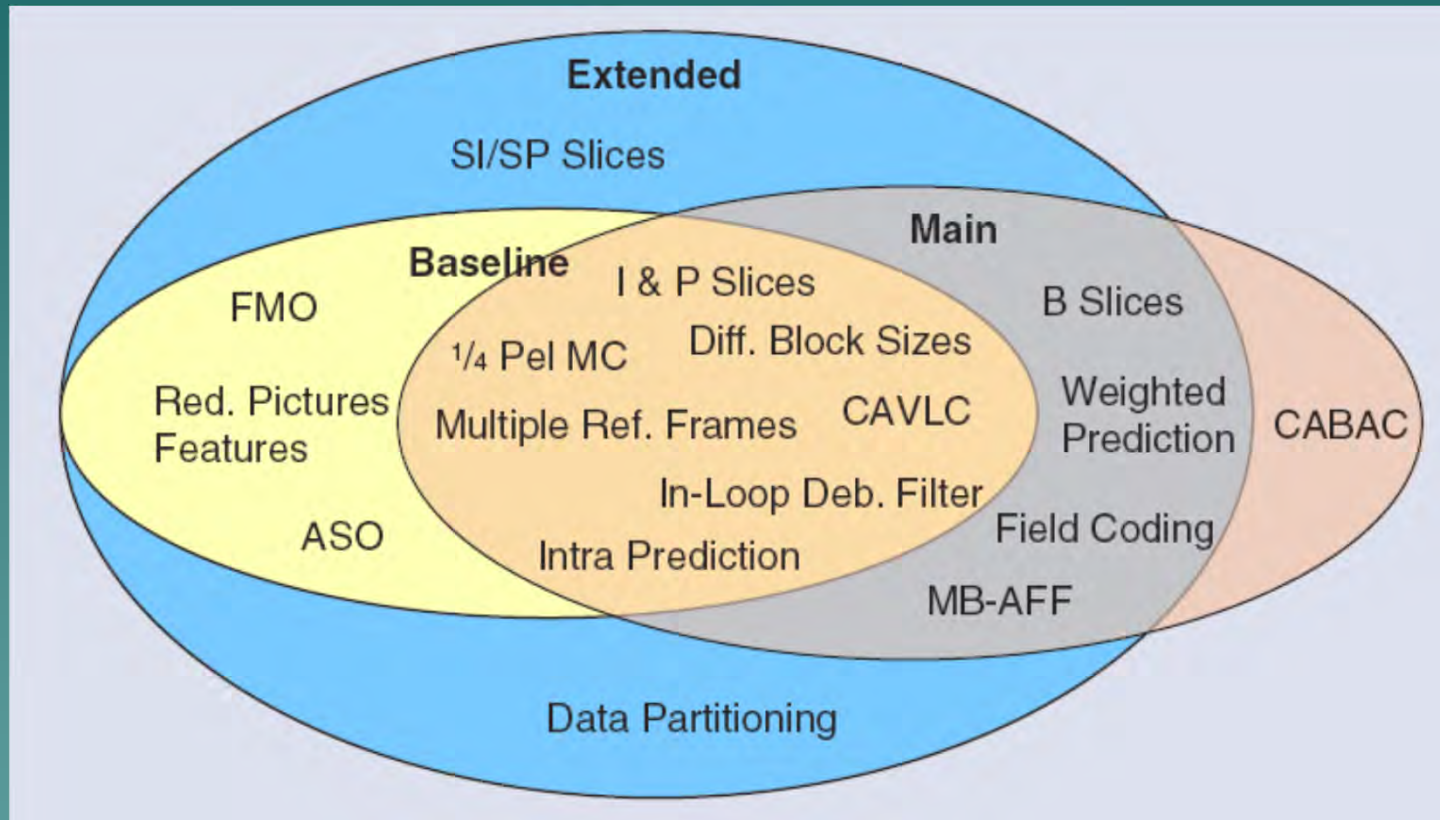
Imen Charfi's Thesis



Temporal Filtering of classification results



Commun parts in H.264



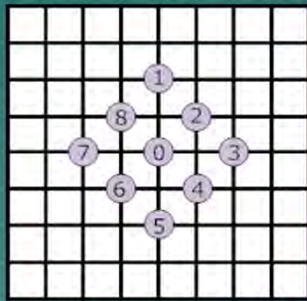
Motion estimator architecture which support any search strategy!

Must support H.264 features!

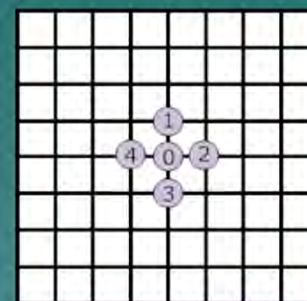
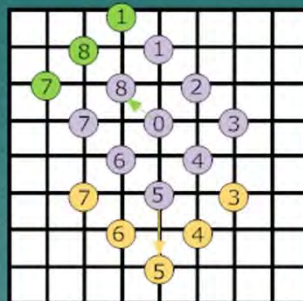
Fast search strategy ?

H.264 features supported

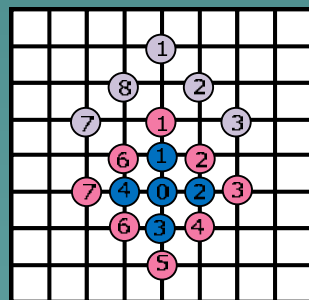
+ Different search strategies supported



LDSP



SDSP

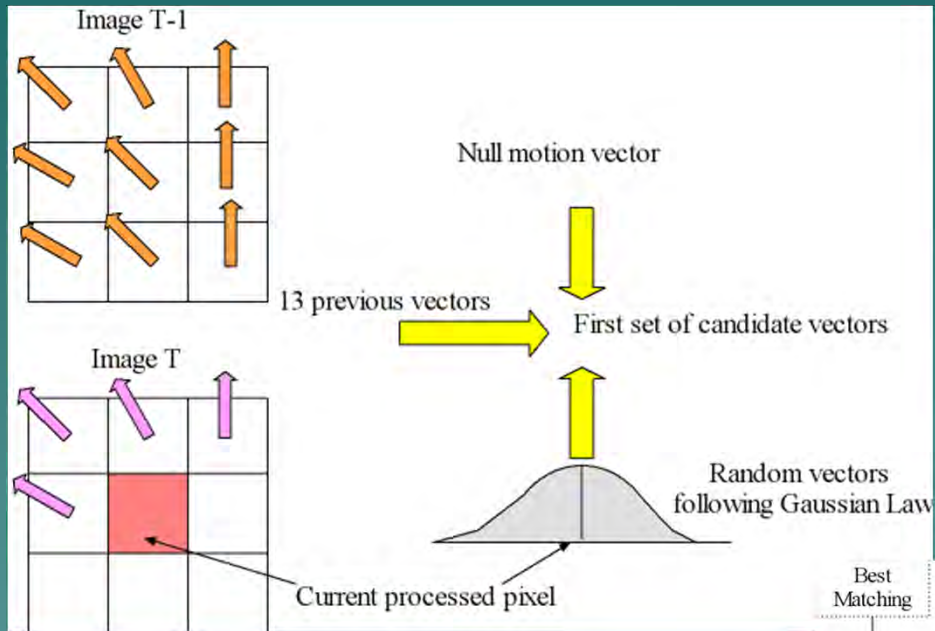


Example

Others reduced search:
Three Steps....

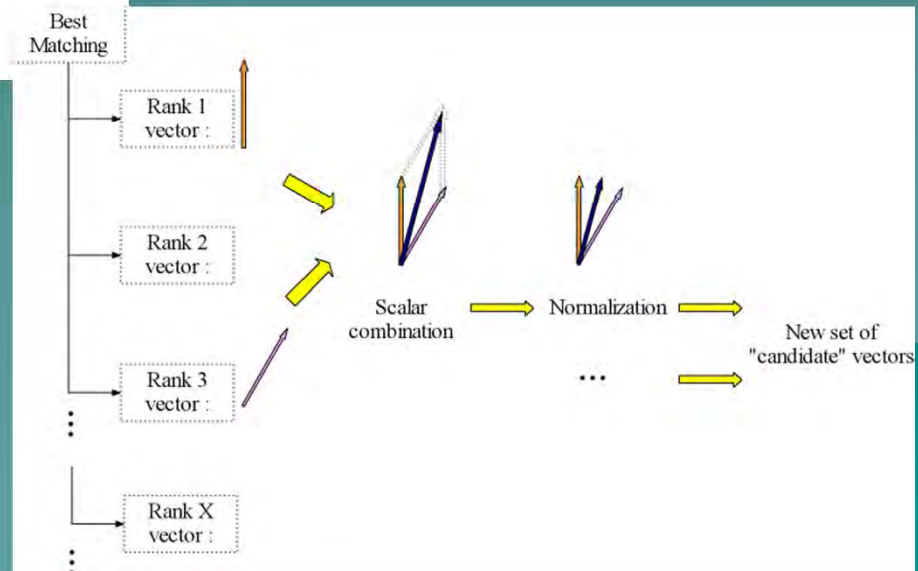
Random search

One solution : prediction + random search

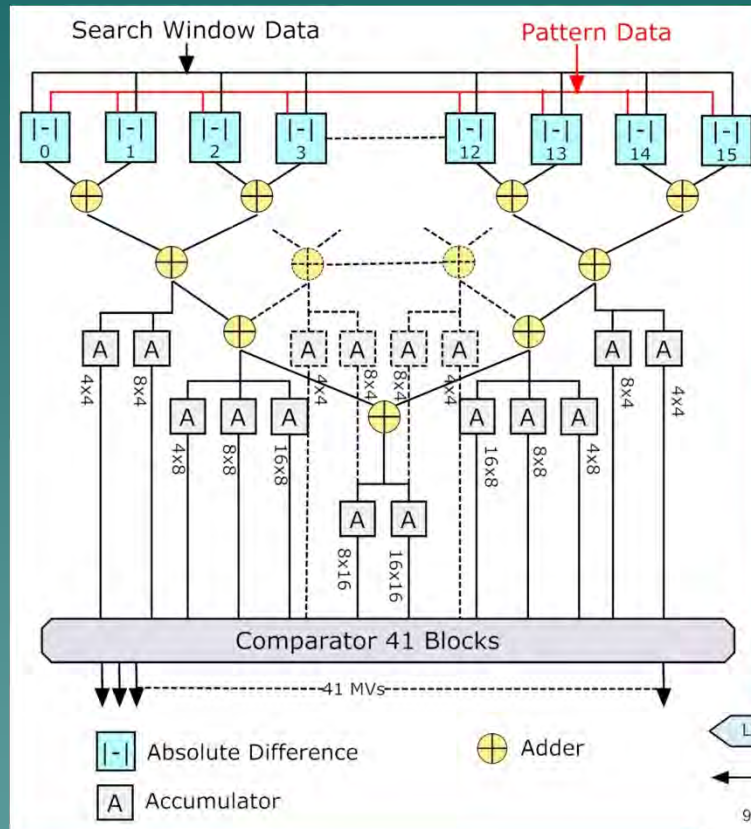


Several phases requested !

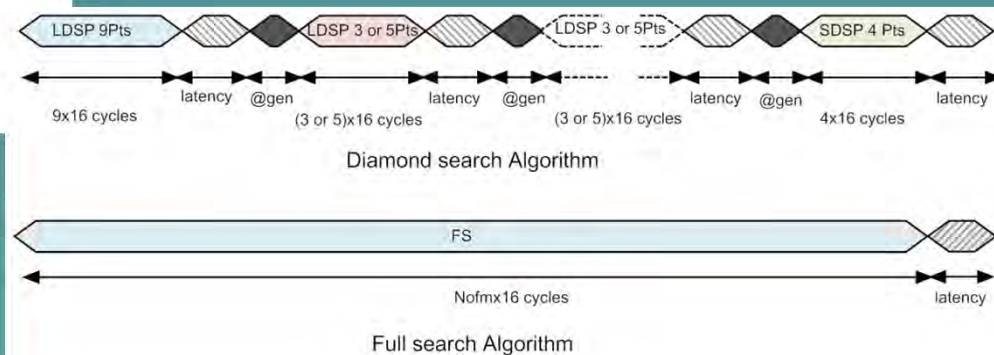
More complex nevertheless dynamic modifications can be considered



Integer ME architecture



Ref.	Year	searching range	Latency	Tech (μm)	Freq (MHz)	Max fps using FS	search strategies
[3]	2004	16x16	4096	TSMC 0.13	294	720x576@45fps	FS
[15]	2006	16x16	4096	TSMC 0.18	266	352x288@25fps	FS
[18]	2009	32x32	26624	TSMC 0.18	316	352x288@30fps	FS
Ours	2012	16x16	4096	Virtex 6vlx240 tff784-3	438	720x576@67fps	FS / DS

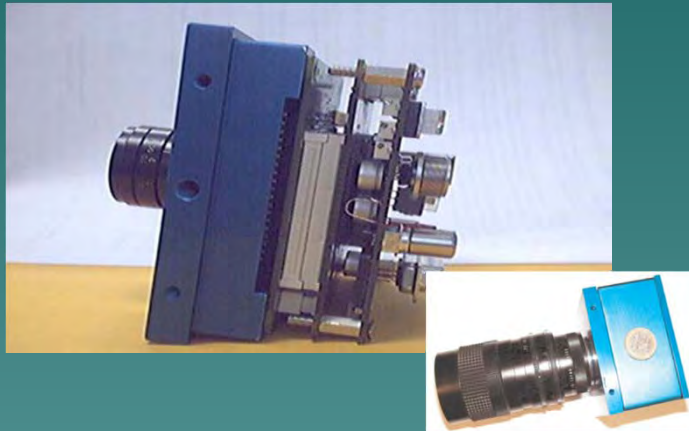


R. Mosqueron

J. Dubois

M. Paindavoine

Smart Camera



CMOS sensor :

500 images/s,

1280 * 1024 pixels sur 10 bits

Bandwidth required 6.55Gb/s.

USB2 Bandwidth :

Peak 340Mb/s upto 480Mb/s

Average 200Mb/s

Compression ration 30 :

$6.55 / .20 \approx 30$

Others pre-processing implented



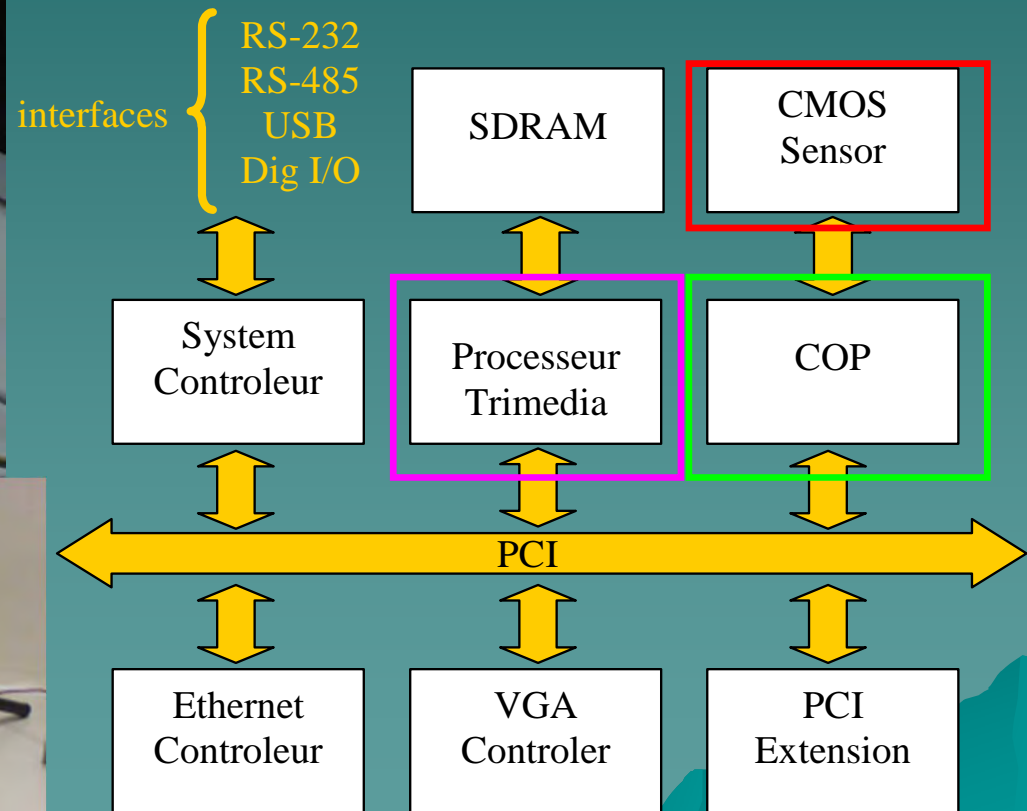
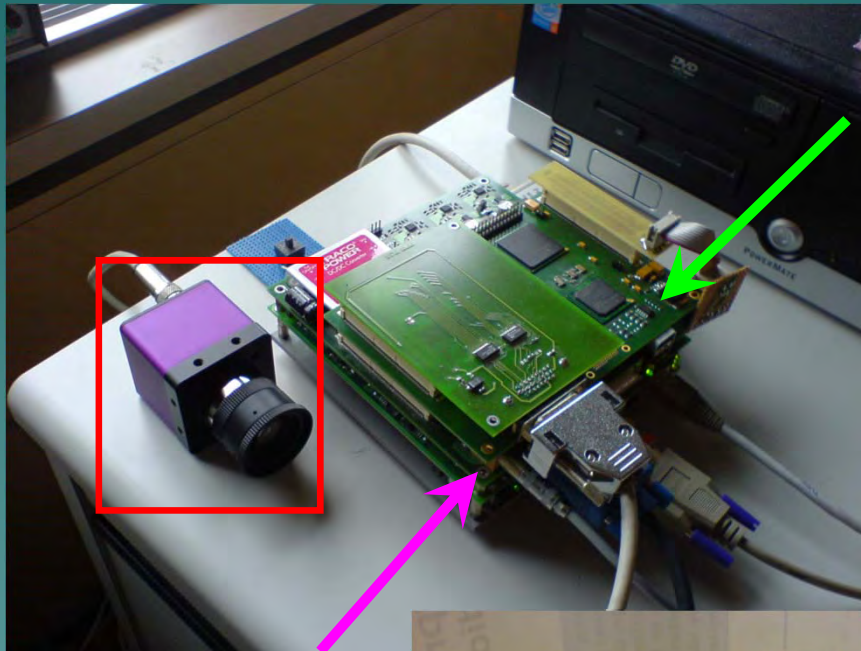
Image originale



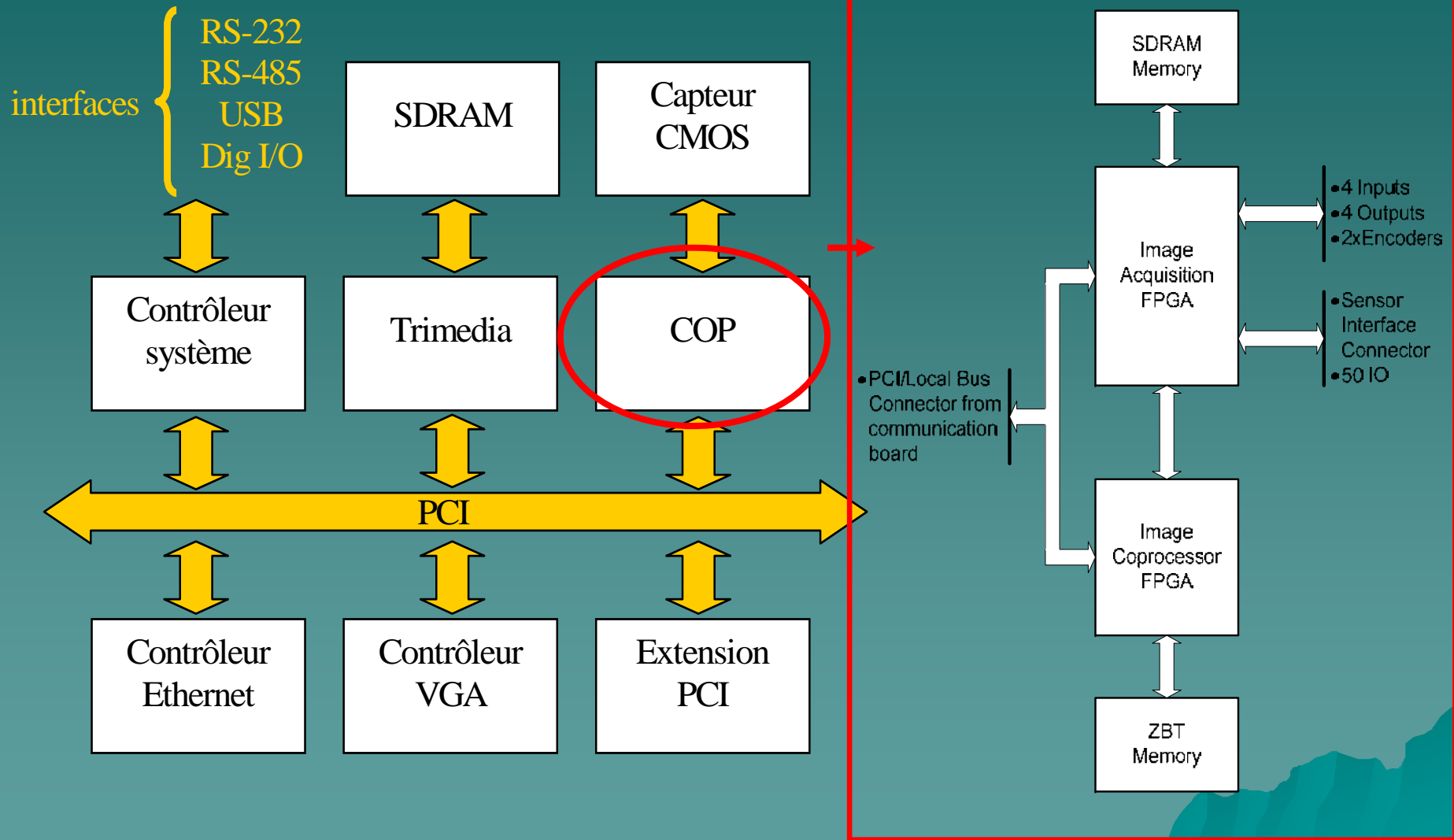
Image obtenue avec
le codage en ondelettes
PSNR=31dB
compression=10



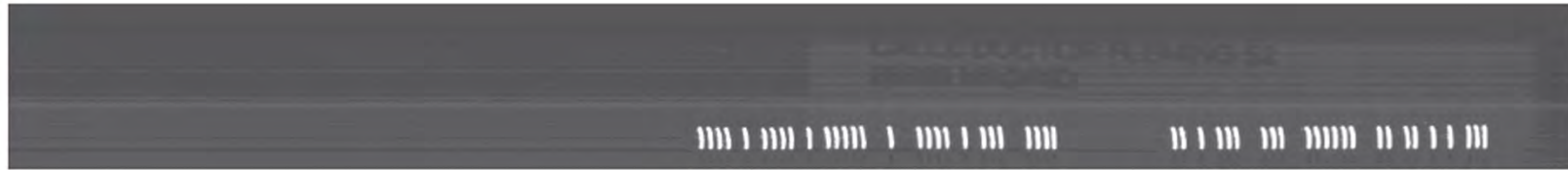
Smart camera with heterogeneous architecture



Heterogeneous smart camera prototype



Application : postal sorting



(a) Original image



(b) Low pass filtering



(c) Dilatation



(d) Sub-sampling



(e) Blobbing

Application : postal sorting



(a) Zoom of the sub-sampling image



(b) Zoom of the blobbing image



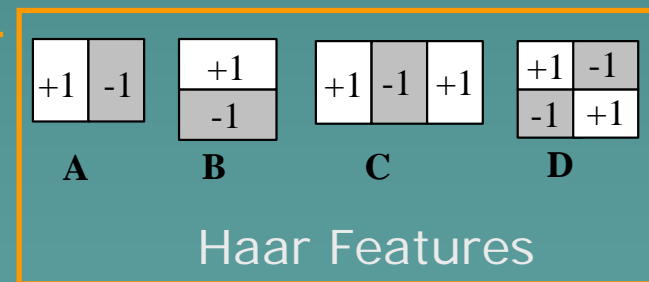
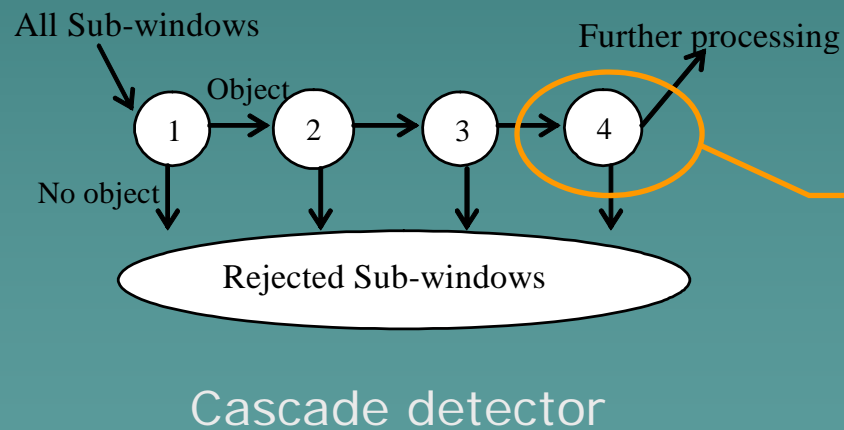
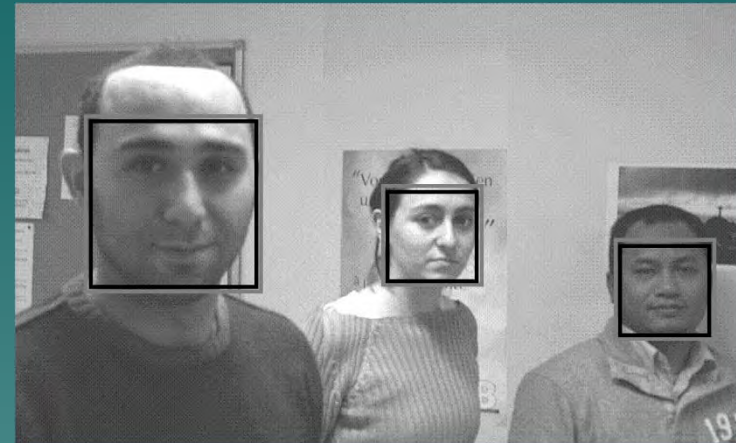
J. Dubois J. Mitéran

W. Elhamzi I. Charfi

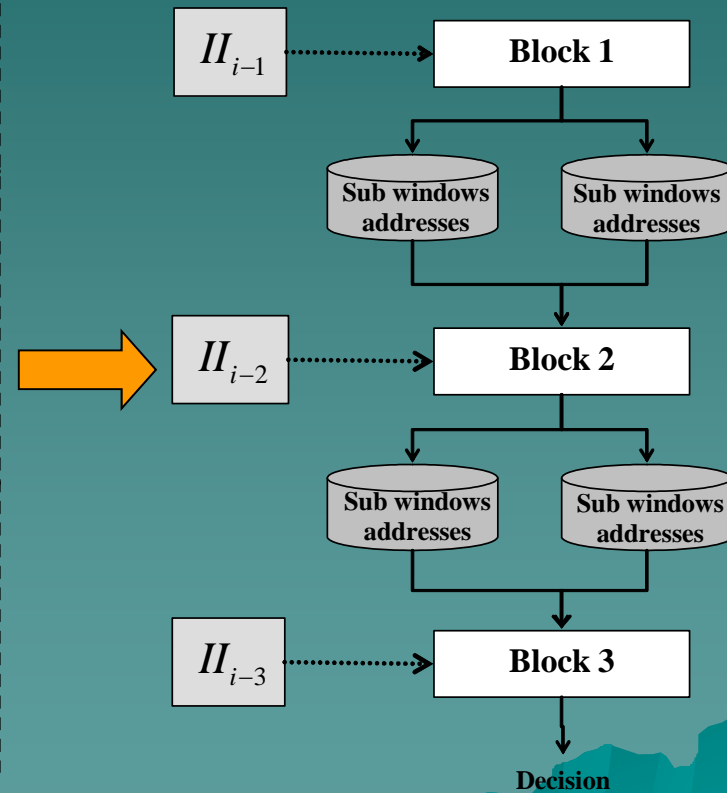
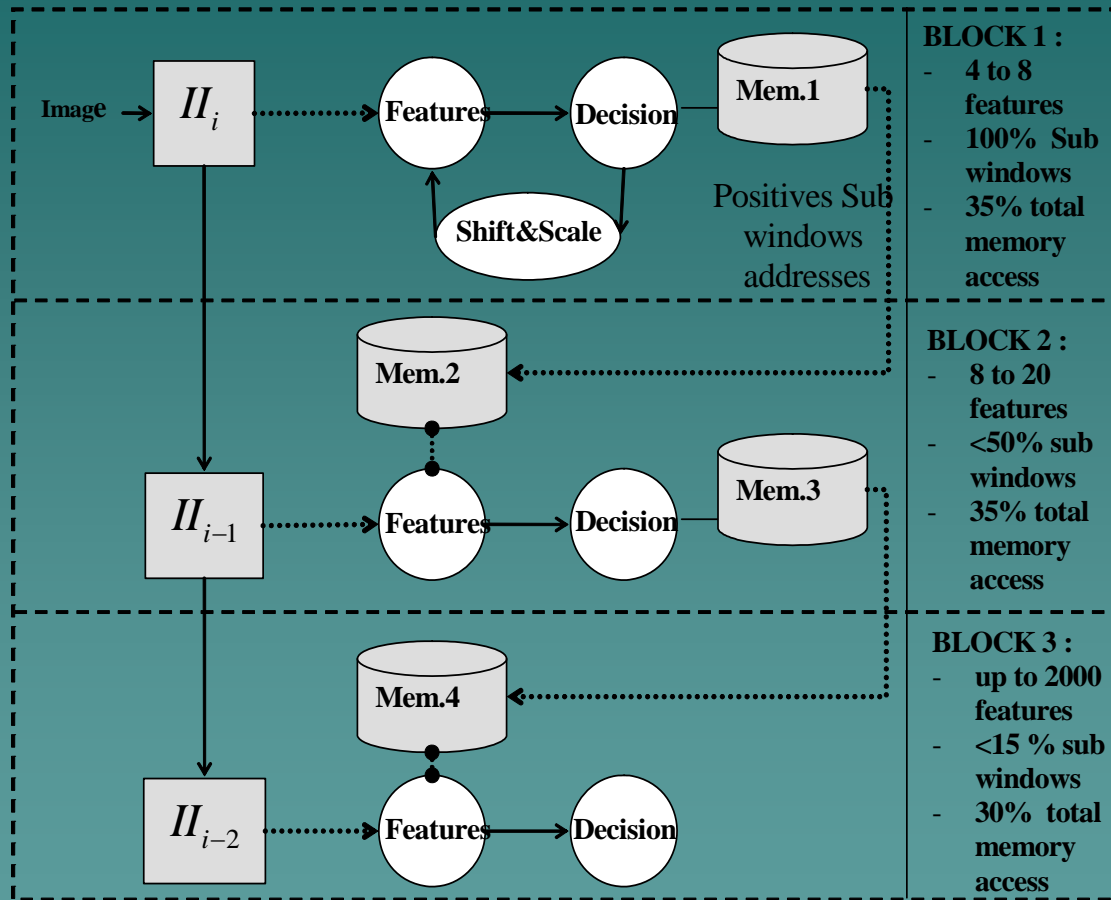
K. Khattab

SystemC Modelisation

Face Detection and localisation
based Viola-Jones method



Architecture model



Implementation

Tool : SystemCrafter

