The Demand for Many Cores: Tera-scale Usage Models

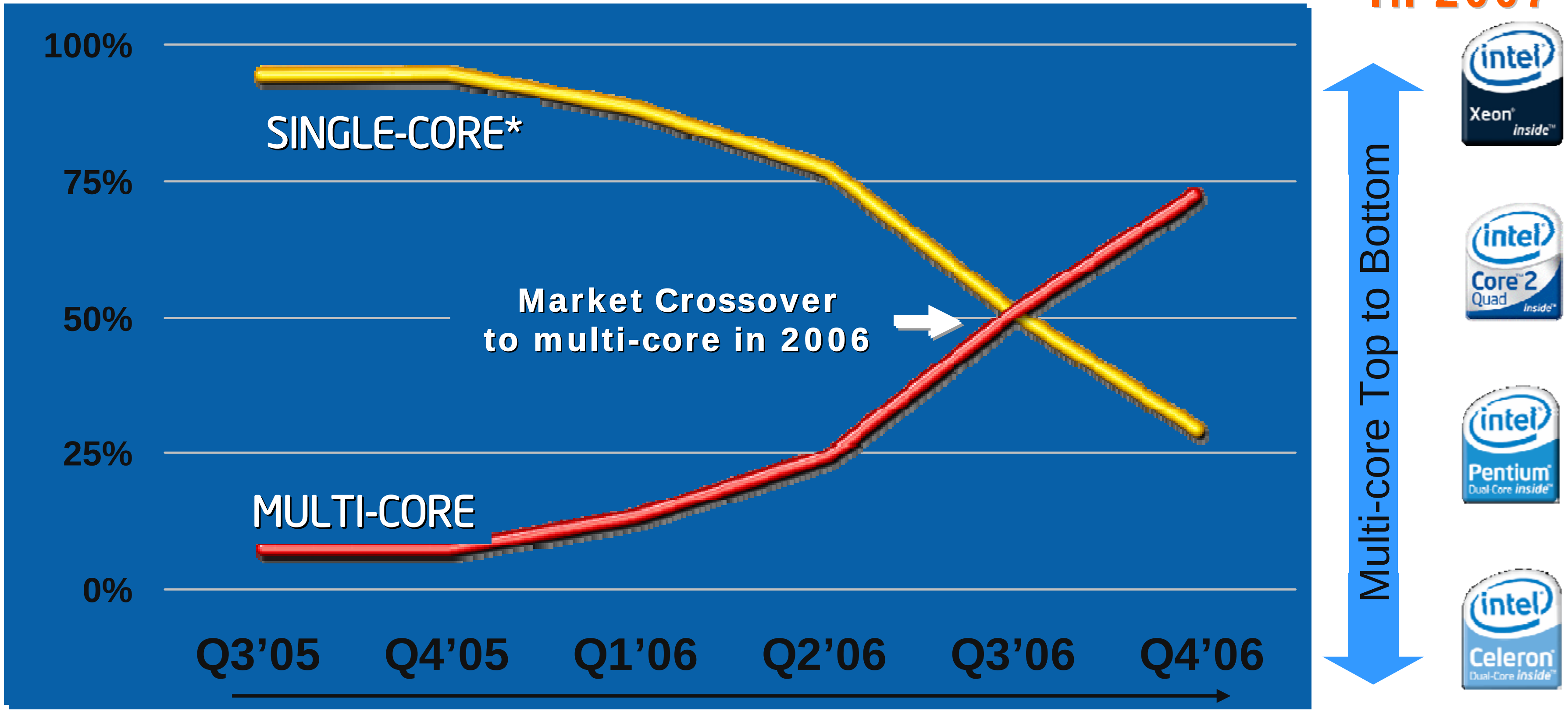
Dr. Yimin Zhang
Research Manager
Intel China Research Center

Agenda

- Introduction to Tera-scale
- Tera-scale Usage Models
- Enabling future applications
- Deeper look – Visual Media Research in China

Multi-core is now mainstream

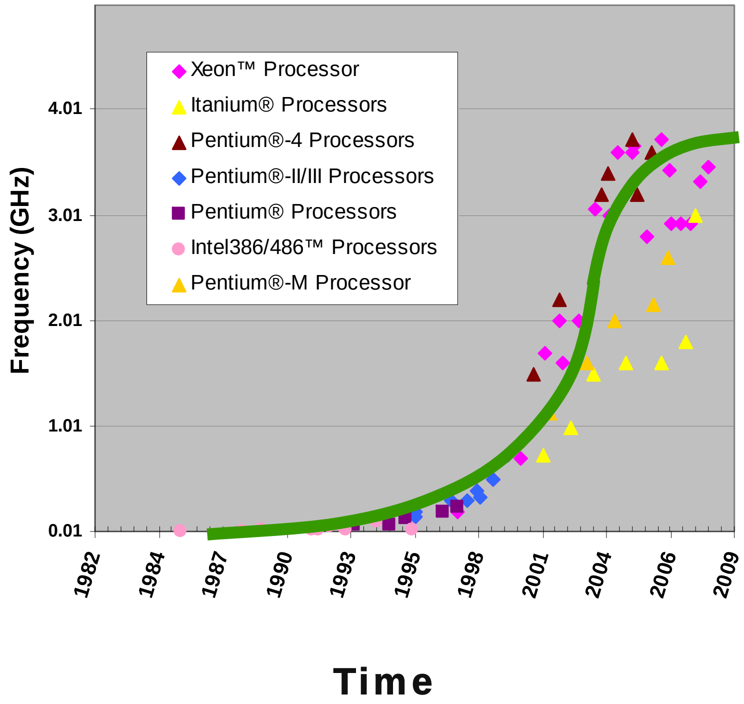
Over 6 MU Quad-core In 2007

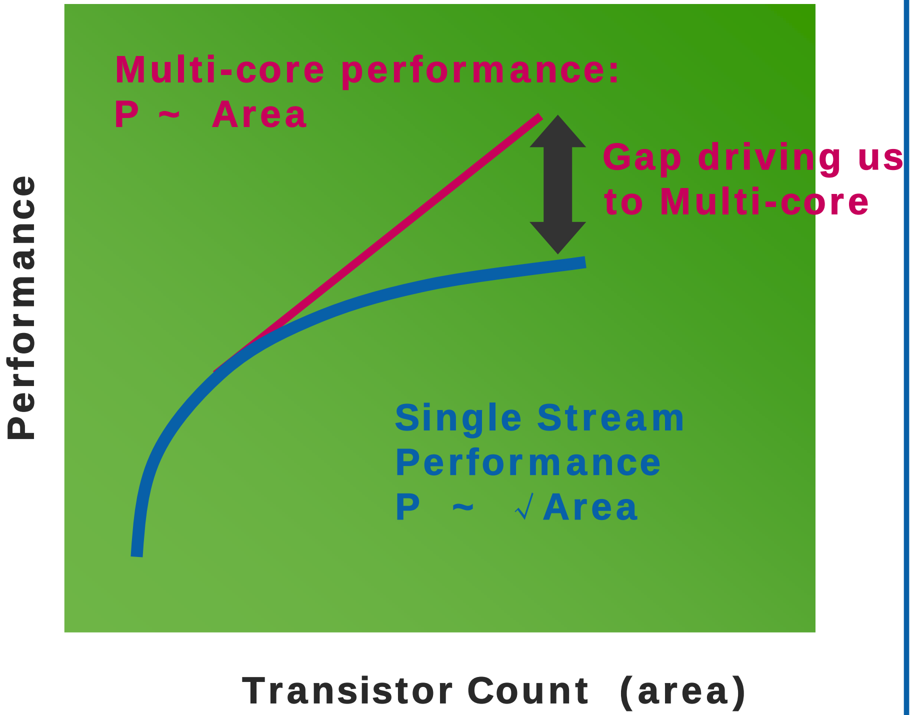


Market Crossover to multi-core in 2006

Question: How do we continue adding cores, and why?

Transition from more Clock to more Cores

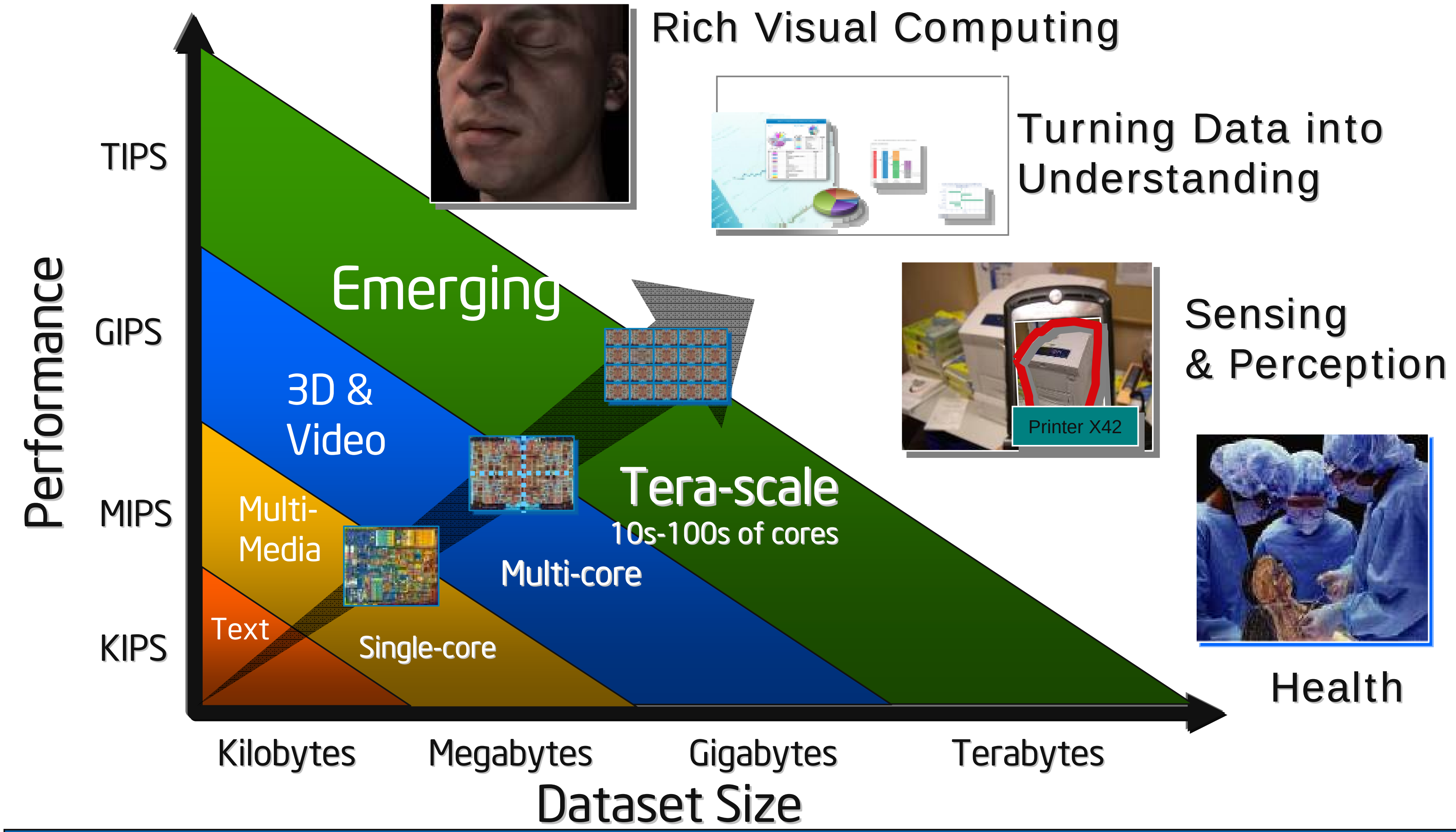




- Frequency limited by leakage and power ($P = CV^2f$)
- Transistor counts continue to increase with Moore’s Law
- Multi-core adds performance for less power

Intel Tera-scale Research

Scaling multi-core to bring tera-scale performance the mainstream

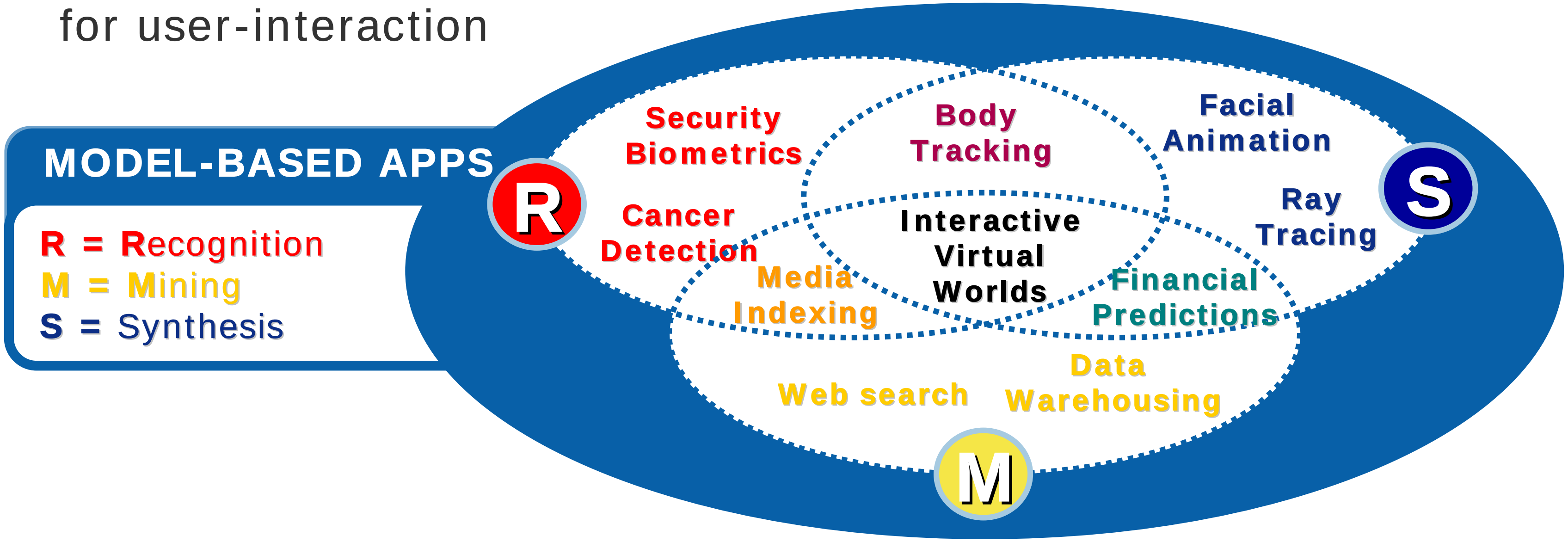


Agenda

- Introduction to Tera-scale
- Tera-scale Usage Models
- Enabling future applications
- Deeper look – Visual Media Research in China

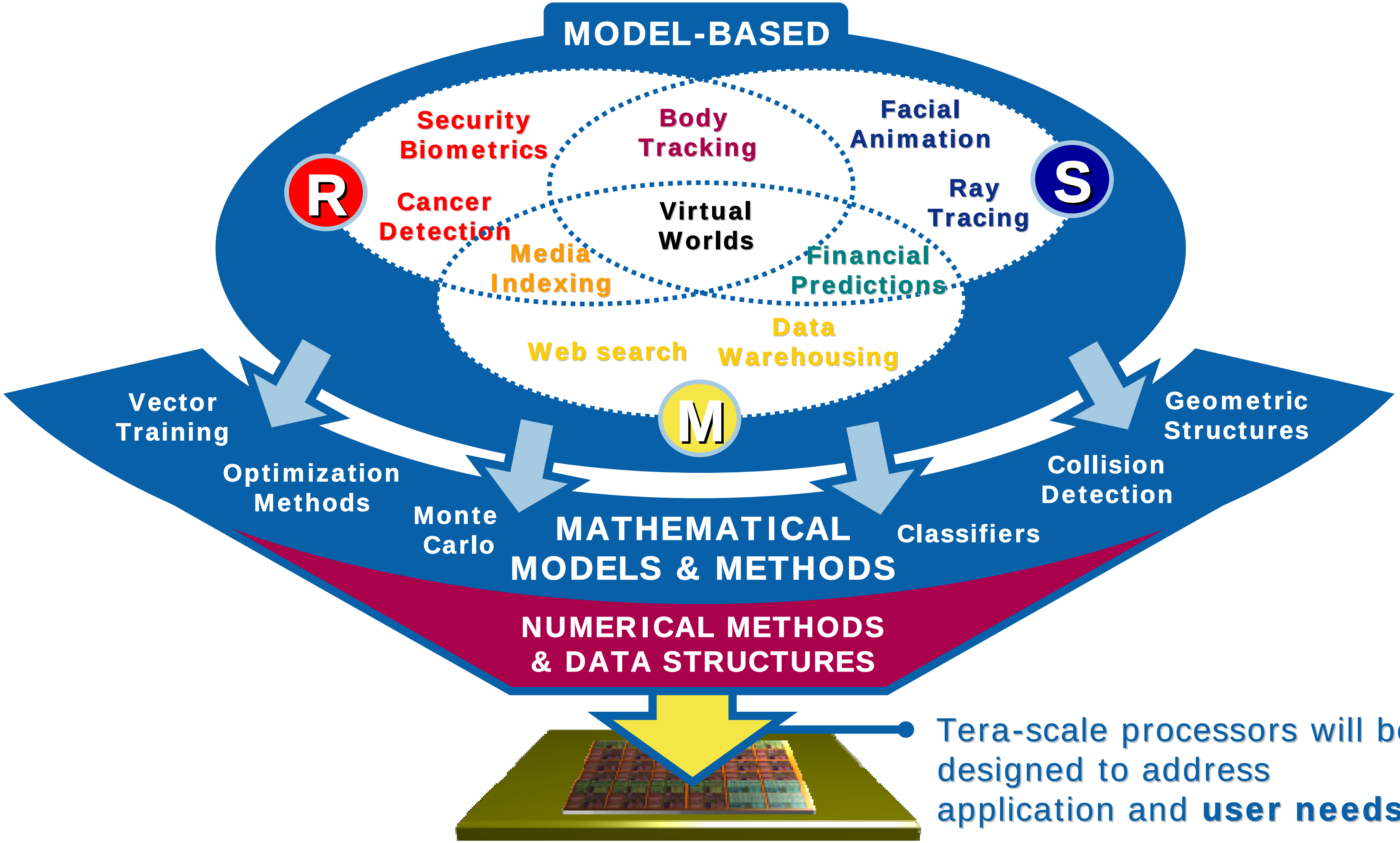
Emerging Performance-Driven Apps

- Based on modeling or simulating the real world
- Make technology more immersive and human-like
- Algorithms are highly parallel in nature
- **Real-time** results essential for user-interaction



In 2004, these observations led us to explore tera-scale

A Top-down Approach



Tera-scale processors will be designed to address application and **user needs**

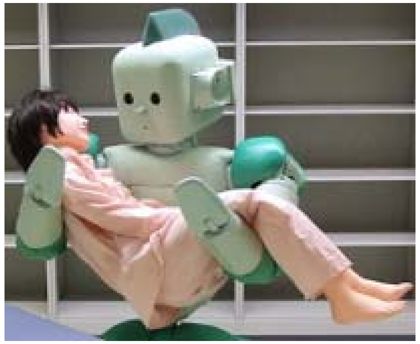
Example: Visual Computing (part 1)

Computational Perception: Expanding human capabilities through improved machine understanding of our world

Location-based Services

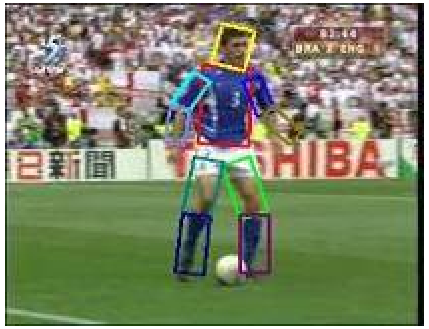
Medical Imaging

Smart cars



Personal Robotics

New Interfaces

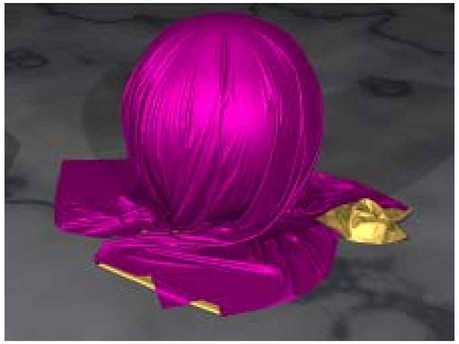


Video Mining

Face tracking

Example: Visual Computing (part 2)

Physics-based realism: graphics that can look, act and react like real-world objects.



Virtual Materials

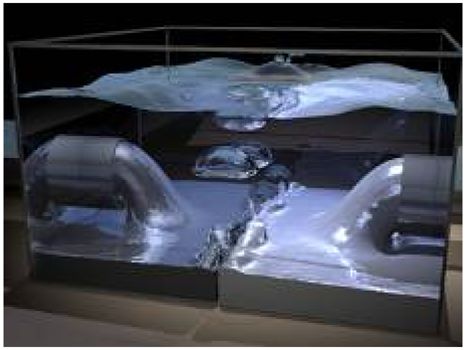
www.3dbrigade.com

Ray Tracing

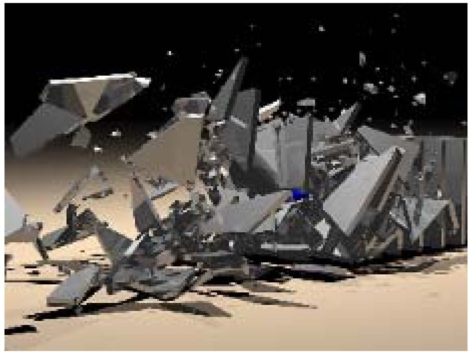
Modeling the physics of light. Used in movies today for more photorealistic graphics.

Sophisticated collisions

Expressive Faces



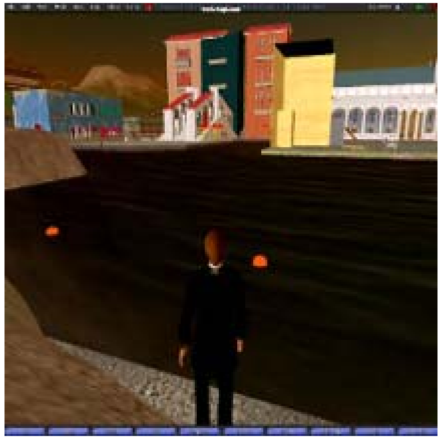
Fluids physics



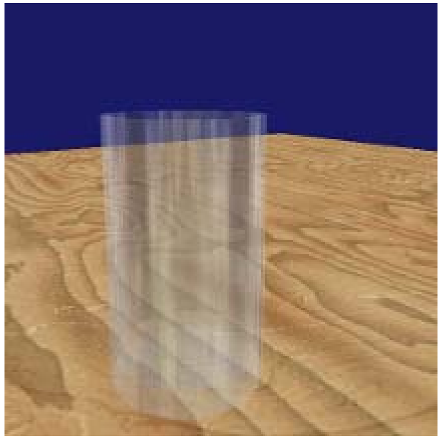
Breakable objects

More Compute → Better Experience

- Perceptual accuracy & graphical realism scale with compute
- Algorithms ideal for array of general purpose CPUs
- Shared algorithms… i.e. ray-tracing for lighting aids physics collision detection and path selection AI



Today: Second Life Fluid
1X



Tomorrow: Particle Fluid
10X Compute

Future: Natural Looking Fluid
1000X Compute

Now → **Effects Physics** → 10+ years

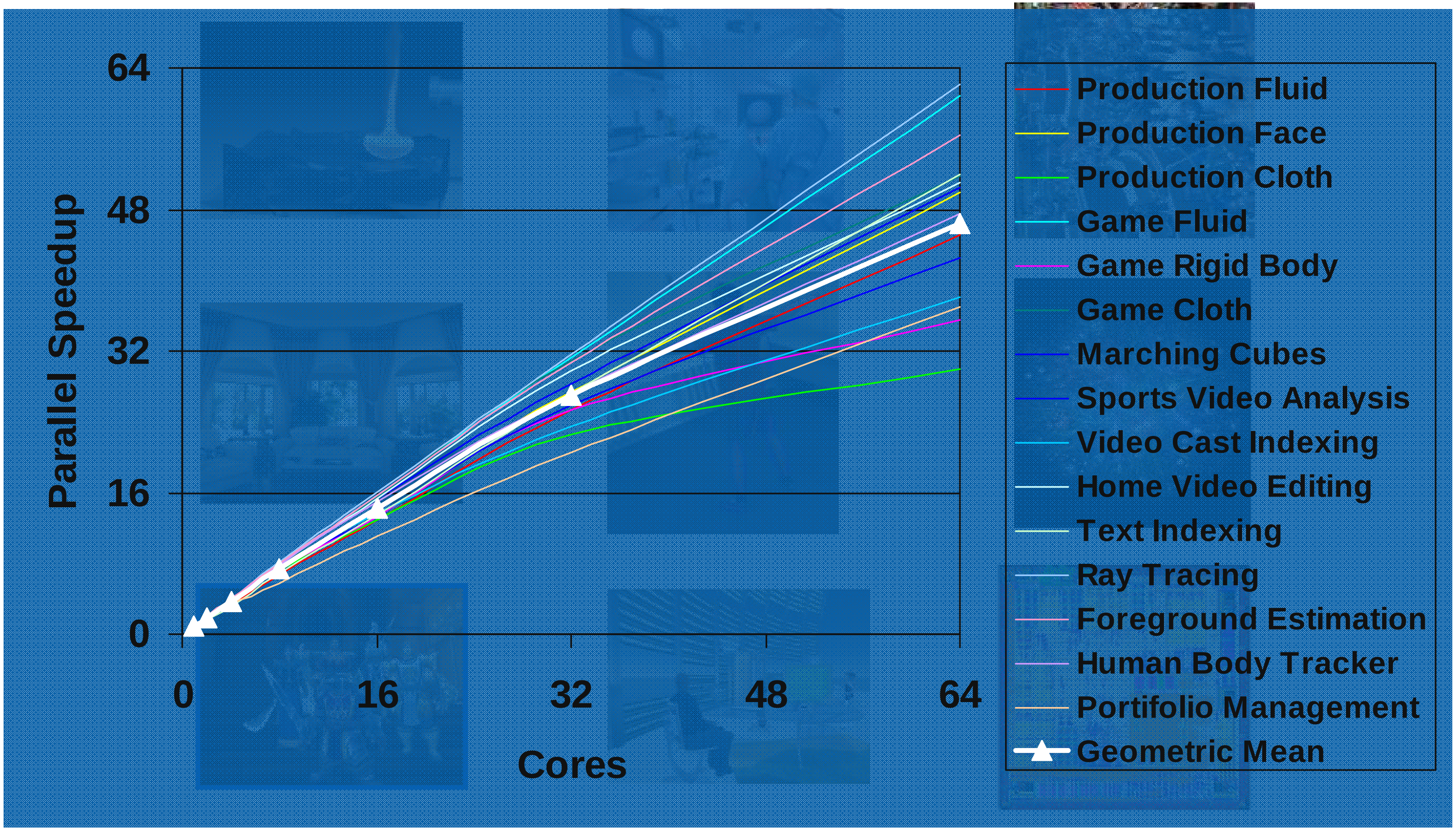
Progress in Real-Time Ray Tracing

- Approaching real-time functionality for ray-traced graphics
- Possible due to increased parallel computing and software innovations

IDF 2004	Early 2007	Fall 2007
50 Intel® Xeon™ Processors	Yorkfield (45nm Quadcore)	Dual-X5365 (total: 8 cores)
4 Frames per Second 640x480	~90 frames per Second 768x768	~90 frames per Second **1280x720**

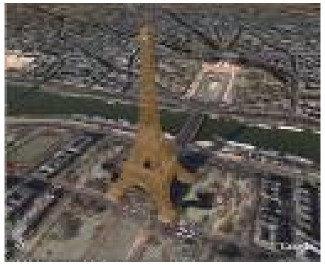
See demo in the technology showcase (BU026)

Research shows Applications Scale Well



Graphics Rendering – Physical Simulation -- Vision – Data Mining -- Analytics

Virtual Worlds: "Connected" Visual Computing



Eiffel Tower in Google Earth

Users Create

World of Warcraft Avatar

Virtual Teamroom

Users Collaborate & Play

Scenario Play

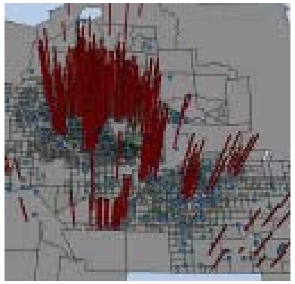
Users Explore and Learn

Machinima Interactive Movies

Qwaq Treefort Virtual Room

Users Enhance the Actual World

Visualizing Real World Information -- Dust storm in Morocco

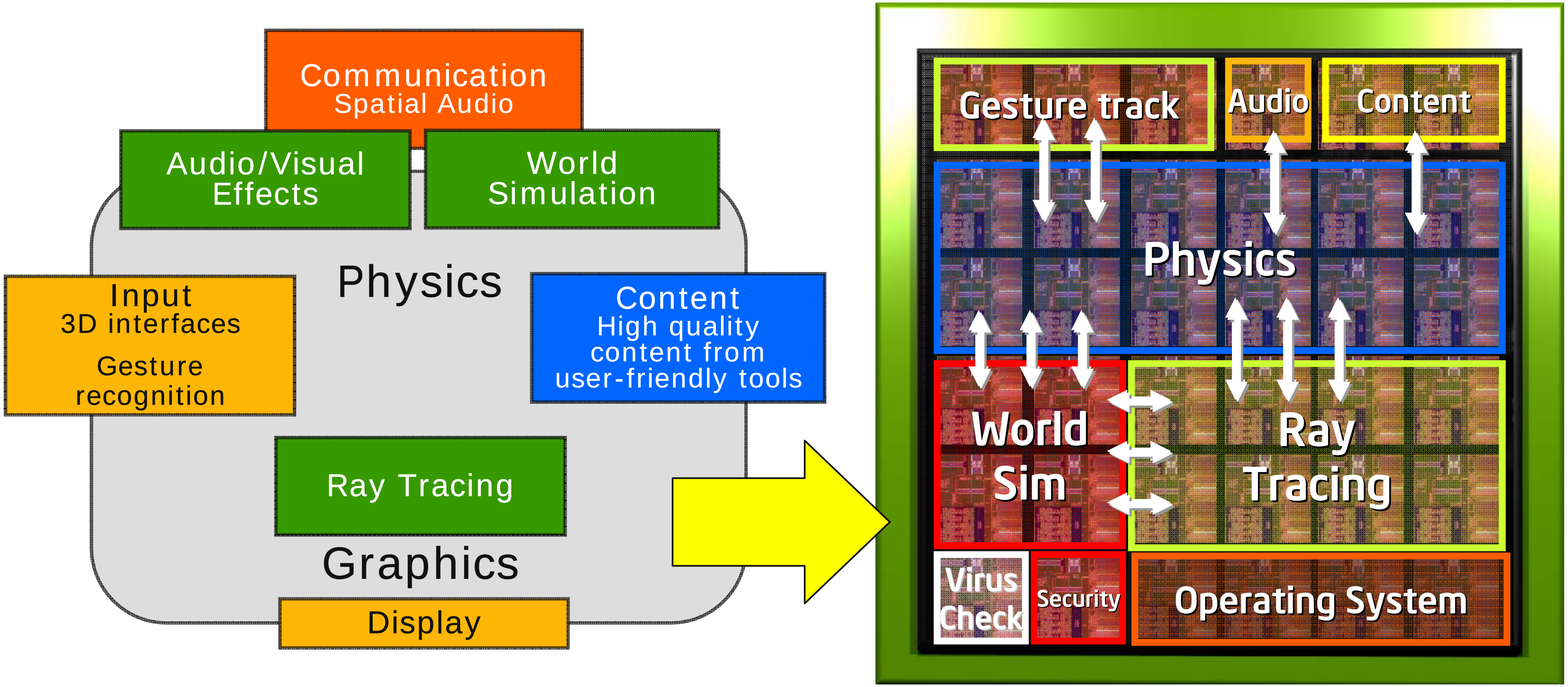


West Nile Virus Visualization

CVC apps will transform the Internet from 2D to 3D

…but require LOTS of compute horsepower

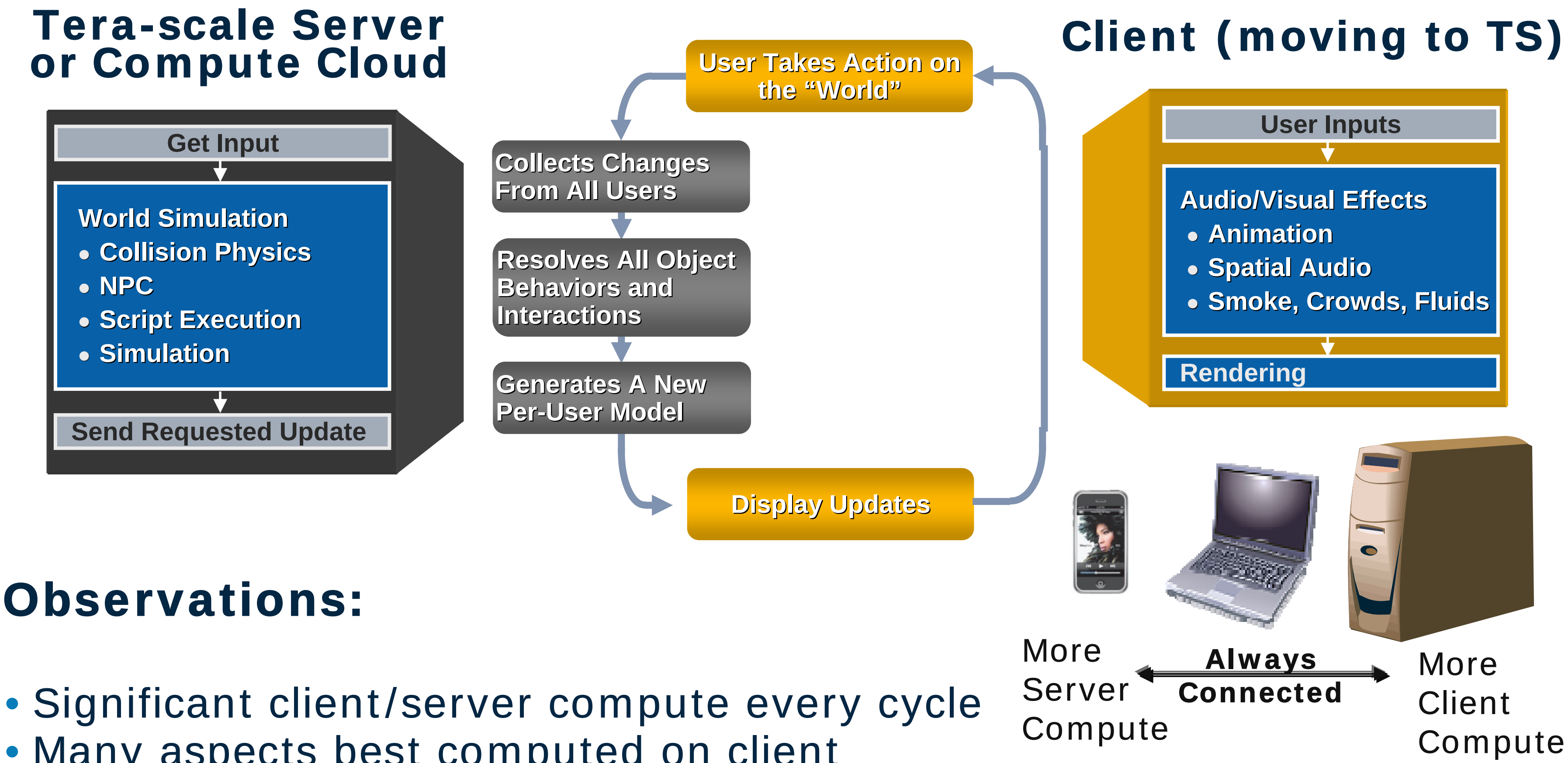
Tera-scale will enable CVC apps



All elements processed on a common platform in parallel.

Tera-scale chips could provide this.

Processing Will Span Client / Server



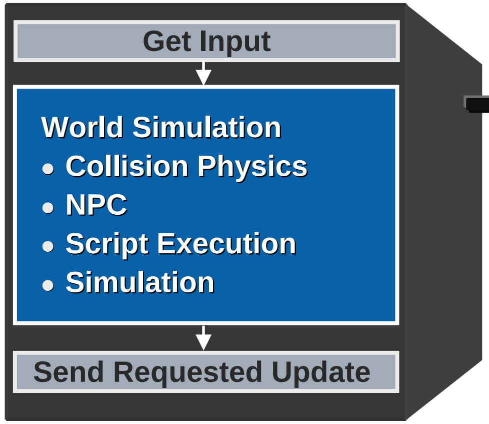
Observations:

- Significant client/server compute every cycle
- Many aspects best computed on client
- Extensive use of MIPS, FLOPS, threads
- Partitioning depends on client capability, connectivity

Mobile CVC includes Augmented Reality

Visual computing + mobility + sensors will mix the virtual and real

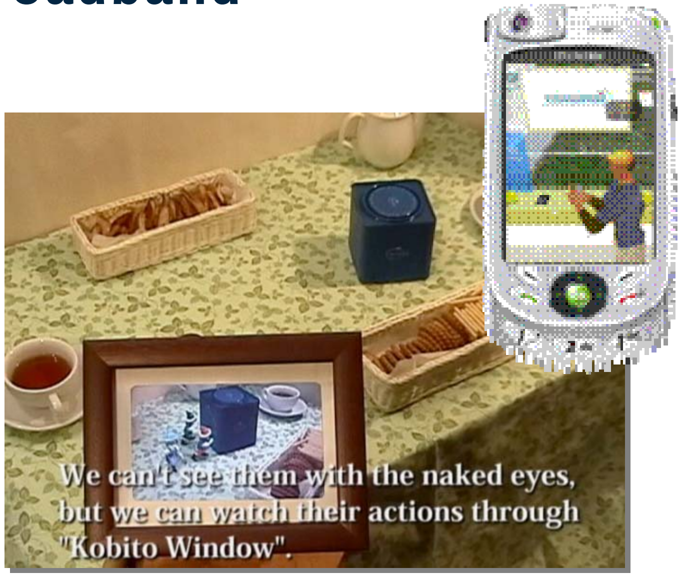
Tera-scale Computer



Mobile Broadband

Mobile Internet Device with Camera, Sensors

- Mobile broadband will connect future, more capable MIDs to tera-scale compute resources
- This will allow us to augment our view of reality


We can't see them with the naked eyes,
but we can watch their actions through
"Kobito Window".

Agenda

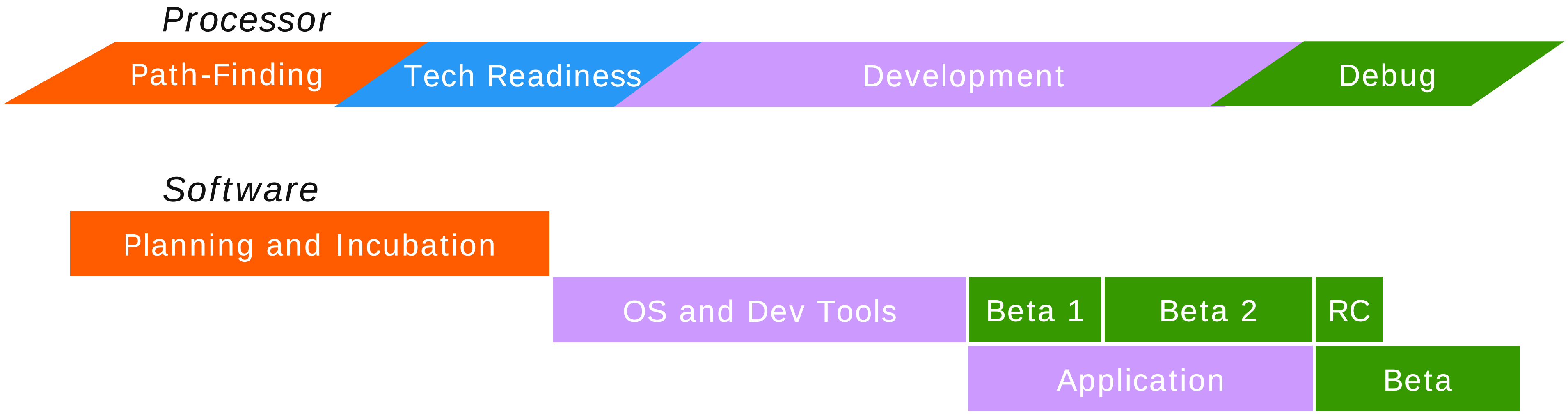
- Introduction to Tera-scale
- Tera-scale Usage Models
- Enabling future applications
- Deeper look – Visual Media Research in China

HW/SW interdependence

- **Scaling multi-core will be challenging**
- **Parallel programming is a <u>major</u> shift for mainstream software**

Hardware must be built to fit needs of SW

Software is developed to exploit existing HW



Processor

Path-Finding

Tech Readiness

Development

Debug

Software

Planning and Incubation

OS and Dev Tools

Beta 1

Beta 2

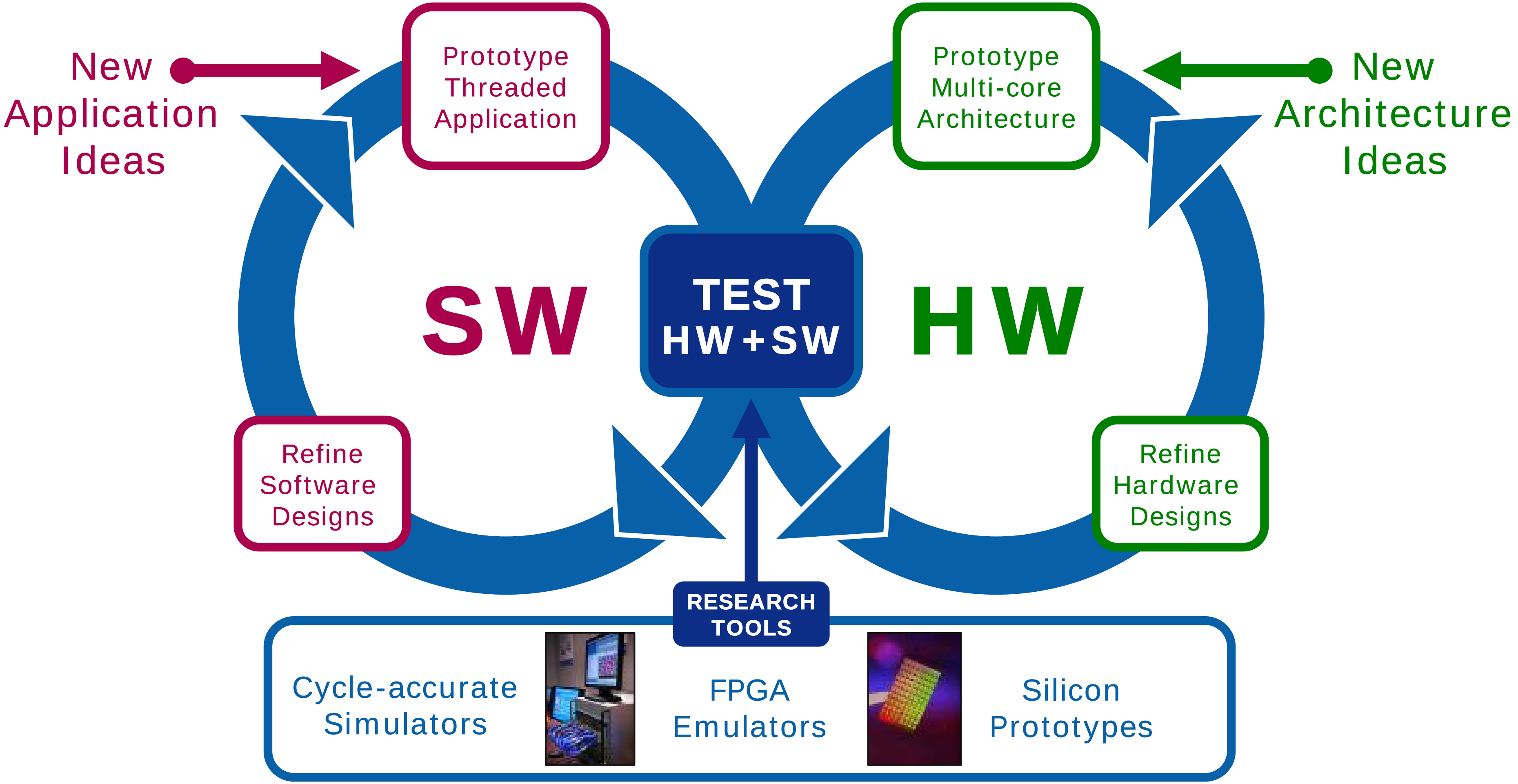
RC

Application

Beta

Both take ~5 yrs how do we avoid a ~10yr transition?

Joint Hardware & Software R&D



New Application Ideas

Prototype Threaded Application

Prototype Multi-core Architecture

New Architecture Ideas

SW

TEST HW+SW

HW

Refine Software Designs

Refine Hardware Designs

RESEARCH TOOLS

Cycle-accurate Simulators

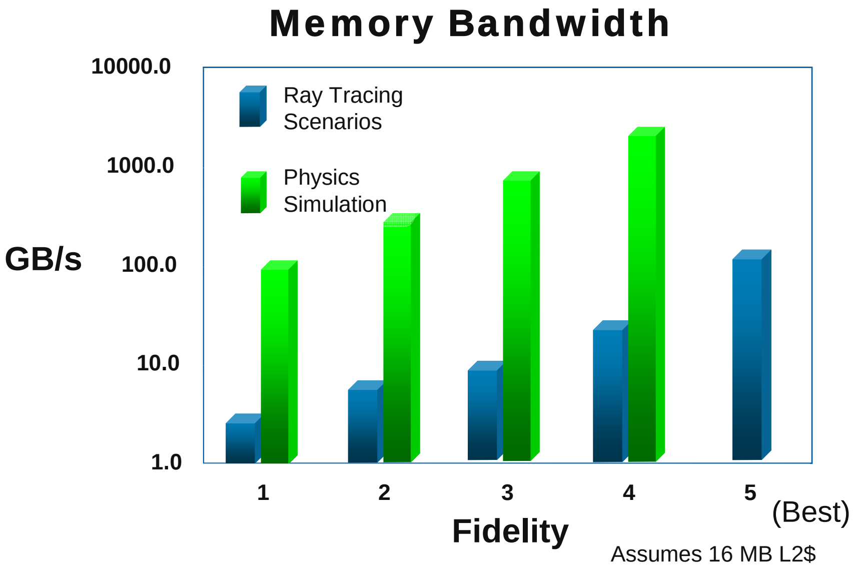
FPGA Emulators

Silicon Prototypes

HW/SW co-development & emulation critical

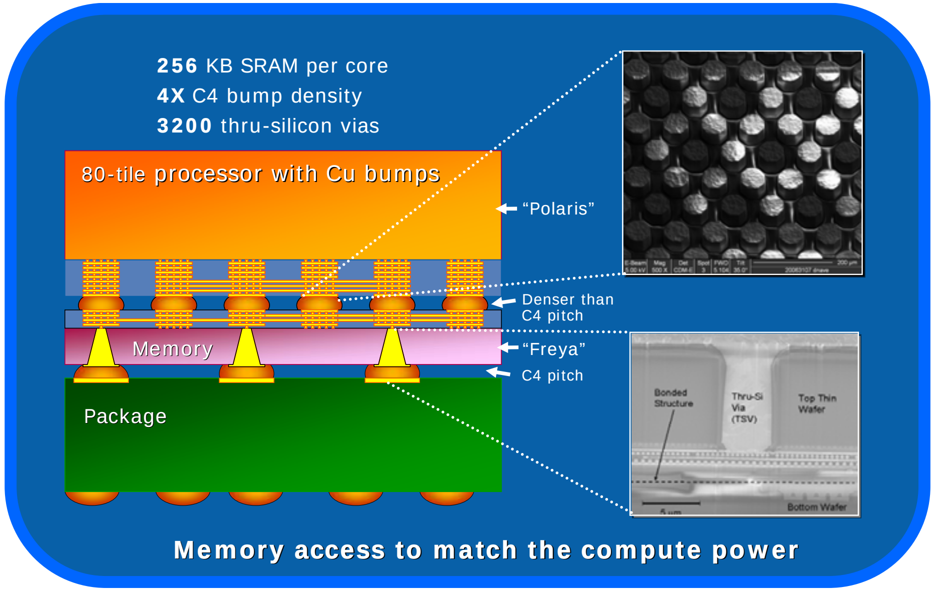
Example: Memory Bandwidth

App Research Findings



Assumes 16 MB L2$

3D Memory Stacking R&D



256 KB SRAM per core
4X C4 bump density
3200 thru-silicon vias

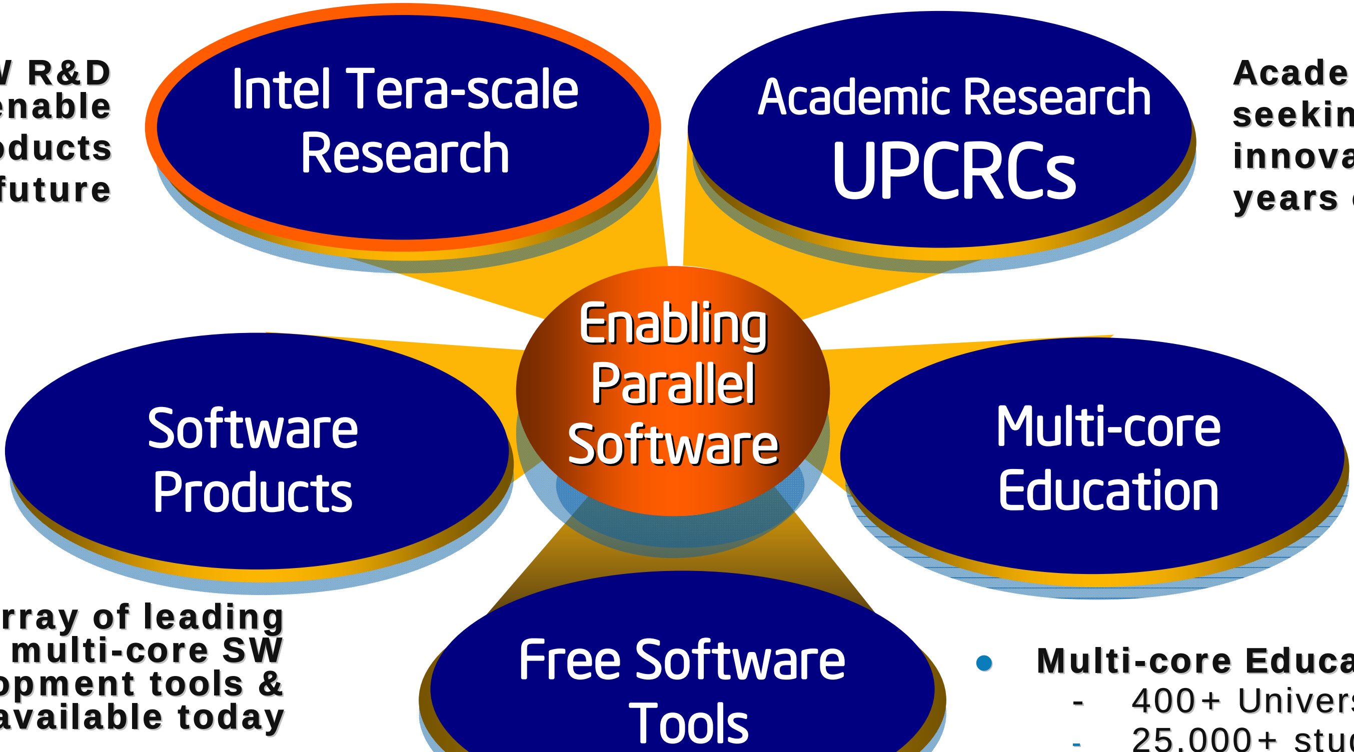
Memory access to match the compute power

- Visual Computing application research shows a tremendous increase in memory bandwidth requirements
- In parallel, we are developing new memory options

Many efforts to enable many cores

Joint HW/SW R&D program to enable Intel products 3-7+ in future

Academic research seeking disruptive innovations 7-10+ years out



Enabling Parallel Software

Intel Tera-scale Research

Academic Research UPCRCs

Software Products

Multi-core Education

Free Software Tools

Wide array of leading multi-core SW development tools & info available today

- TBB Open Sourced
- STM-Enabled Compiler on Whatif.intel.com
- Parallel Benchmarks at Princeton's PARSEC site

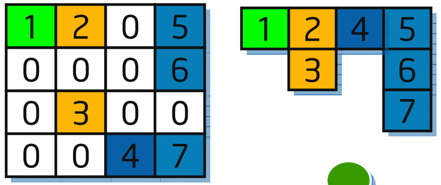
- Multi-core Education Program
  - 400+ Universities
  - 25,000+ students
  - 2008 Goal: Double this
- Intel® Academic Community
- Threading for Multi-core SW community
- Multi-core books

Must work closely with customers, and industry and academic partners

Tera-scale Programming Research

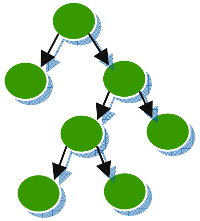
Transactional Memory
Lock-free Parallel Memory Management

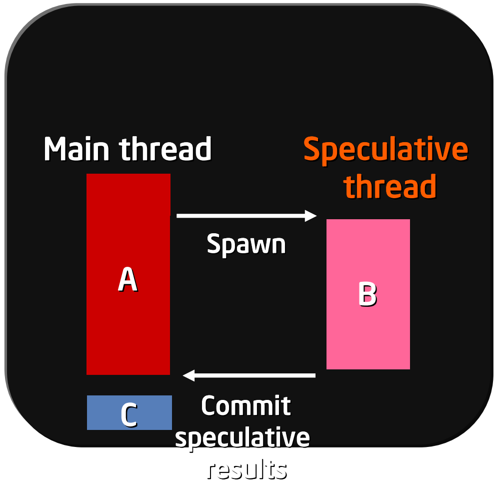
Ct: C/C++ for Throughput Computing
Making it easier to program of a wide array of high-throughput applications.



1	2	0	5
0	0	0	6
0	3	0	0
0	0	4	7

1	2	4	5
	3		6
			7





Speculative Multi-threading
Threading serial code segments at the hardware level

1/3 of Tera-scale research is software enabling

PARSEC

['pär-"sek] A unit of measure

Princeton Application Repository for Shared memory Computers

blackscholes*	Standard Financial analytics benchmark.
bodytrack*	Build body model from video input.
facesim*	physical modeling, face animation.
fluidanimate*	smoothed particle hydrodynamics
freqmine*	frequent item set mining
Swaptions*	financial Monte Carlo code
ferret	Server for image similarity search
dedup	Enterprise Storage
streamcluster	streaming clustering of multidimensional data
vips	Image processing system
x264	H.264 (MPEG-4) video encoding
canneal	VLSI placement program using simulated annealing

Open Source Parallel App Benchmarks

From Kai Li's and J. P. Singh's groups at Princeton

*** Benchmarks provided by Intel**

http://parsec.cs.princeton.edu/

March 2008: New "Universal Parallel Computing Research Centers"

- $20 million committed by Intel and Microsoft
- Two University centers: Berkeley and Illinois

Berkeley
University of California

Professor David Patterson
UCB UPCRC Director

ILLINOIS
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

Prof. Wen-Mei Hwu

Prof. Marc Snir

UIUC UPCRC Co-Directors

Catalyze breakthrough research to help make parallel computing mainstream in 7-10+ years

Agenda

- Introduction to Tera-scale
- Tera-scale Usage Models
- Enabling future applications
- Deeper look – Visual Media Research in China

Market Trends

#1 Explosion of digital content

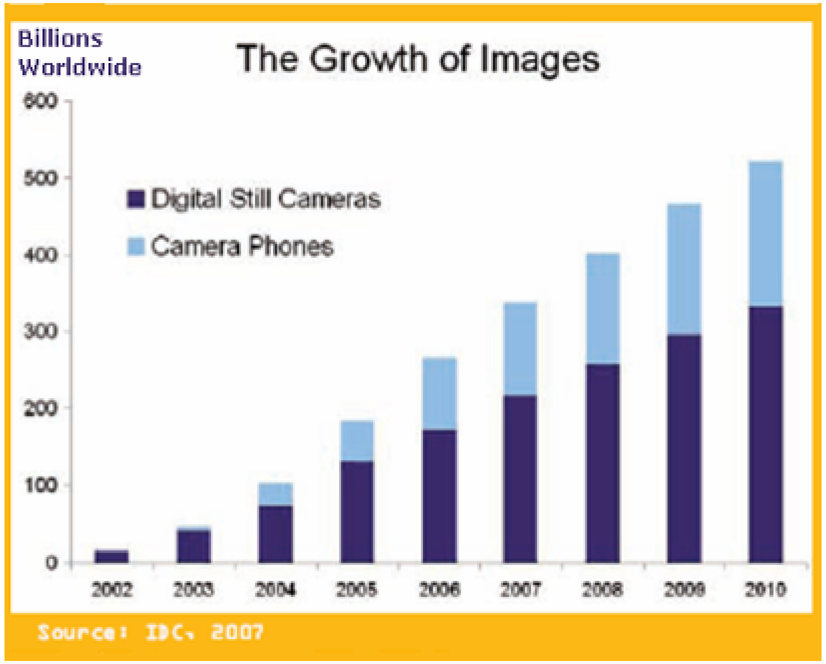
- Over 400M of digital photos shot daily (IDC)
- 11% of U.S PCs have >10K photos (Tabblo)

#2 Consumers want smart filters based on their preferences

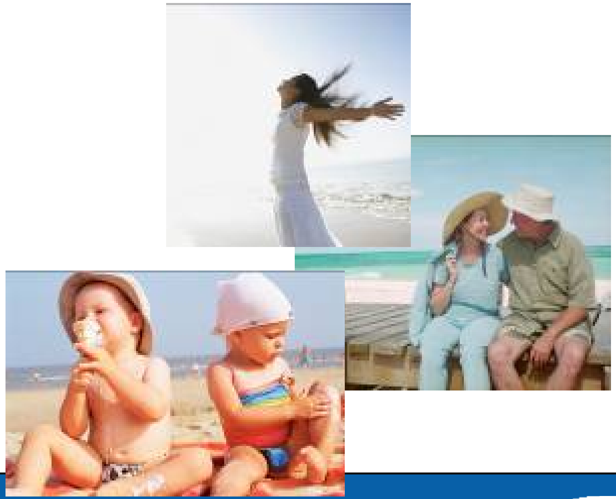
- Locating photos is cumbersome today
- Text search uses file & folder names
- Time & hassle to rename & tag image files

#3 Image recognition is slow

- > 20 seconds / photo / iteration (Pentium M)

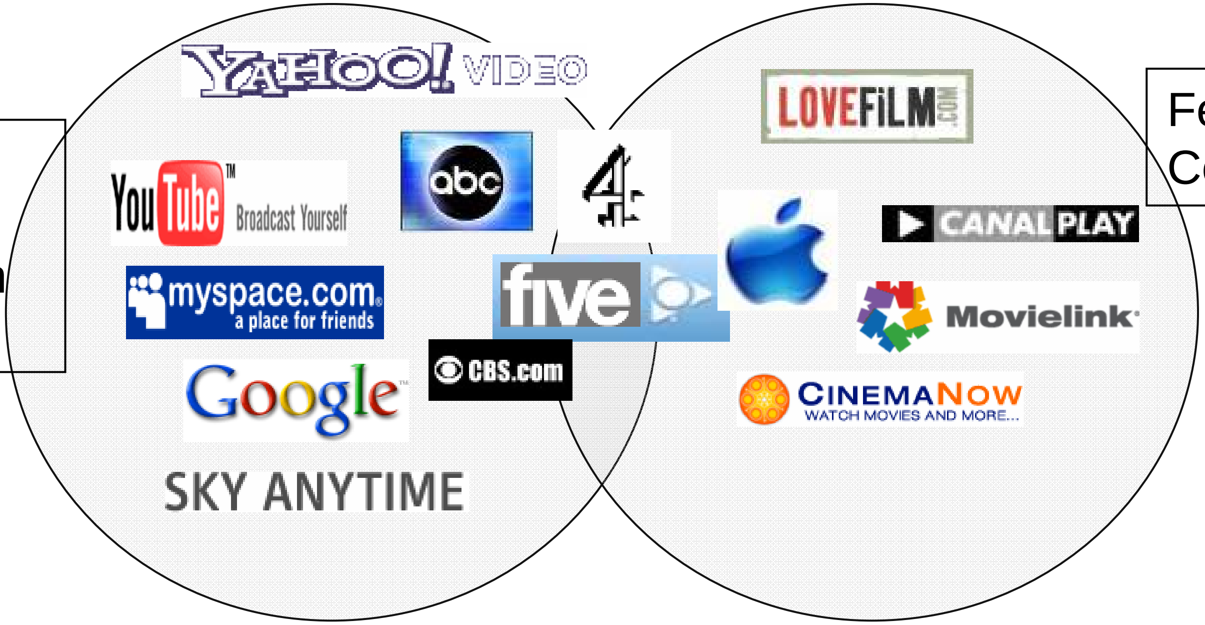


Source: IDC, 2007



Online Video Services

Generally Free Content with Ads


YAHOO! VIDEO
YouTube Broadcast Yourself
abc
4
myspace.com a place for friends
five
CBS.com
Google
SKY ANYTIME
LOVEFiLM.com
CANAL PLAY
Movielink
CINEMANOW WATCH MOVIES AND MORE...

Fee-Based Content

Professional and User-Generated Multimedia content growing rapidly

Key technology for media mining

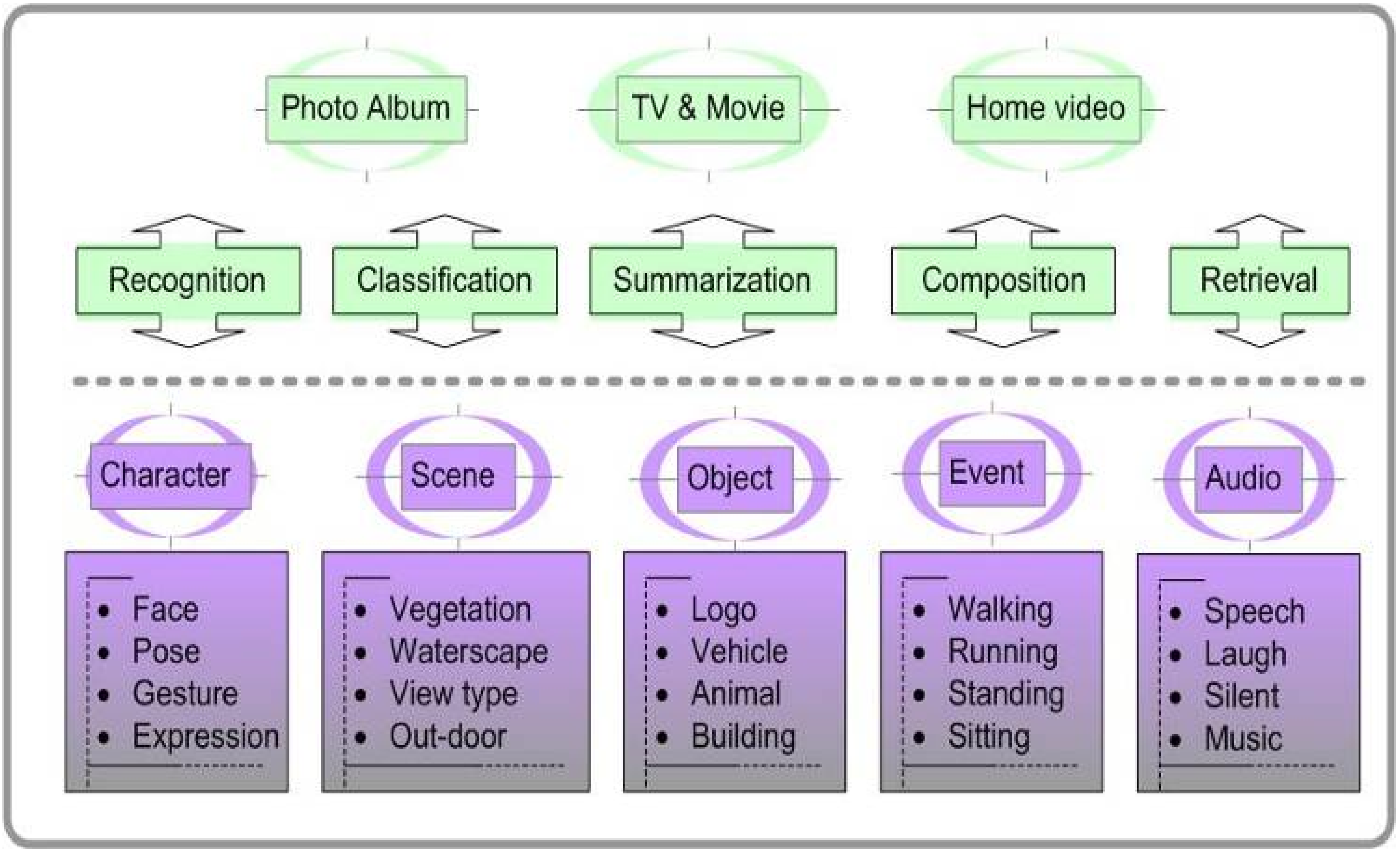


Photo Album

TV & Movie

Home video

Recognition

Classification

Summarization

Composition

Retrieval

Character

- Face
- Pose
- Gesture
- Expression

Scene

- Vegetation
- Waterscape
- View type
- Out-door

Object

- Logo
- Vehicle
- Animal
- Building

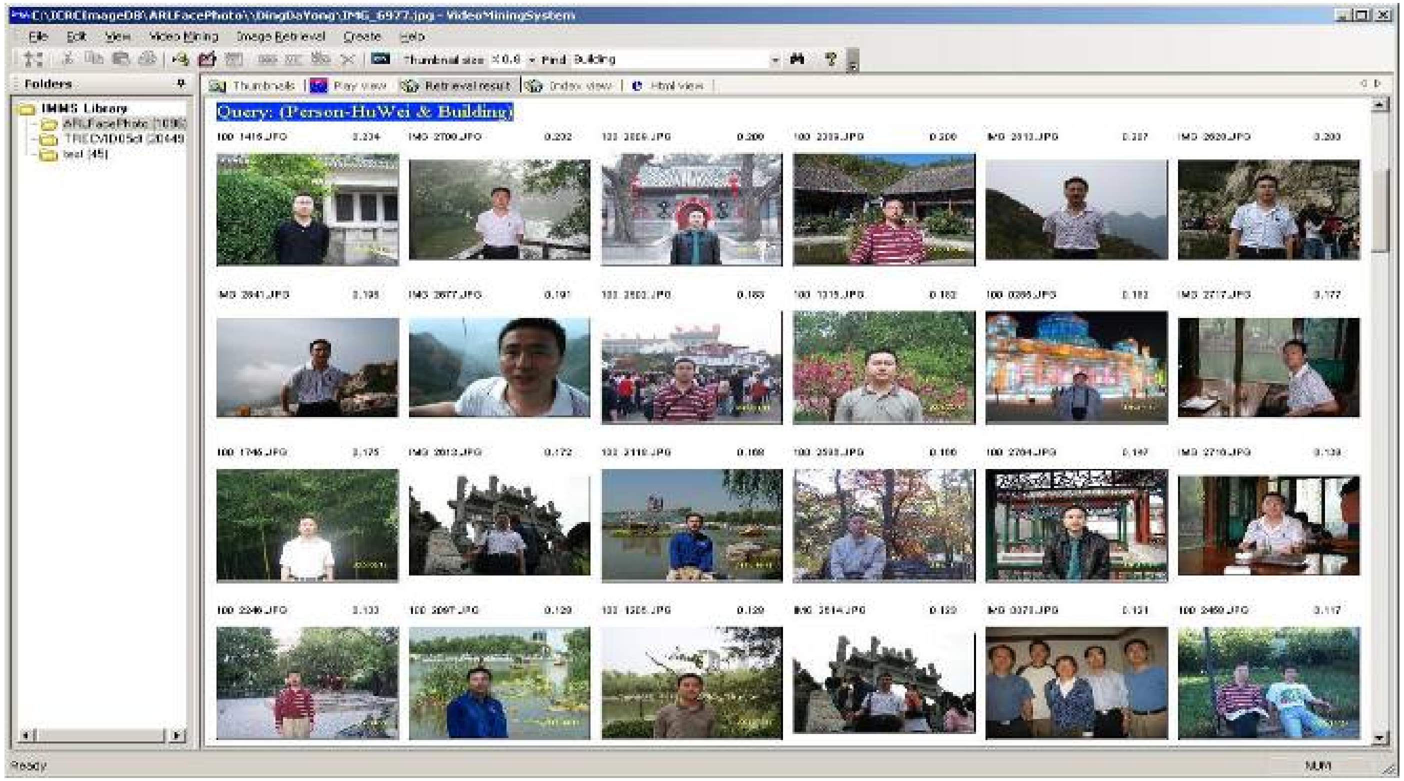
Event

- Walking
- Running
- Standing
- Sitting

Audio

- Speech
- Laugh
- Silent
- Music

Person recognition: Face recognition in photo

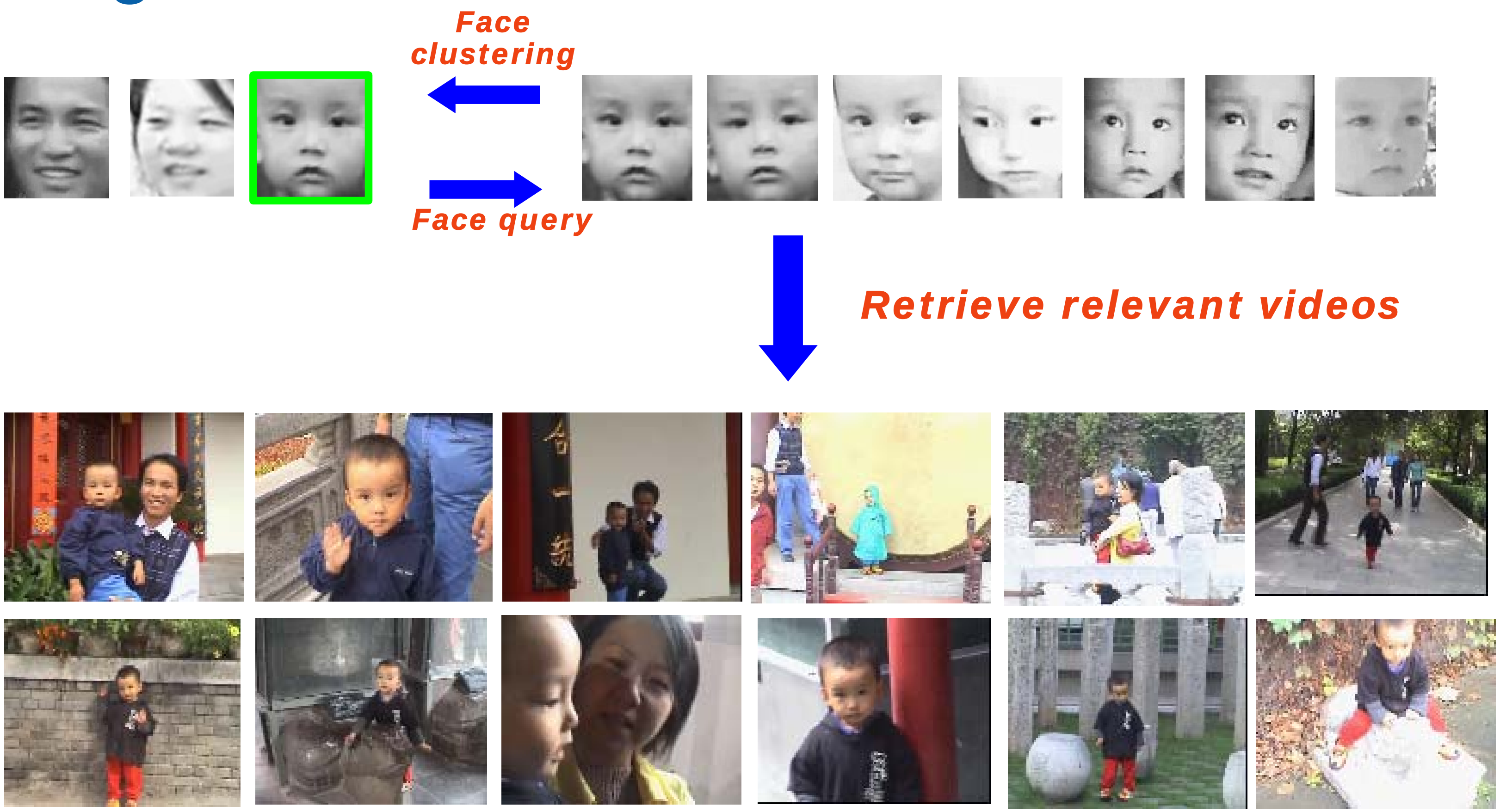


Query: (Person-HuWei & Building)

- *Highly accurate multi-view face detection + FIGHT feature + LDA*
- *90% accuracy in 1000 personal photos from 24 people*

Best accuracy reported

Person recognition: Cast Indexing / face recognition in video



Face clustering

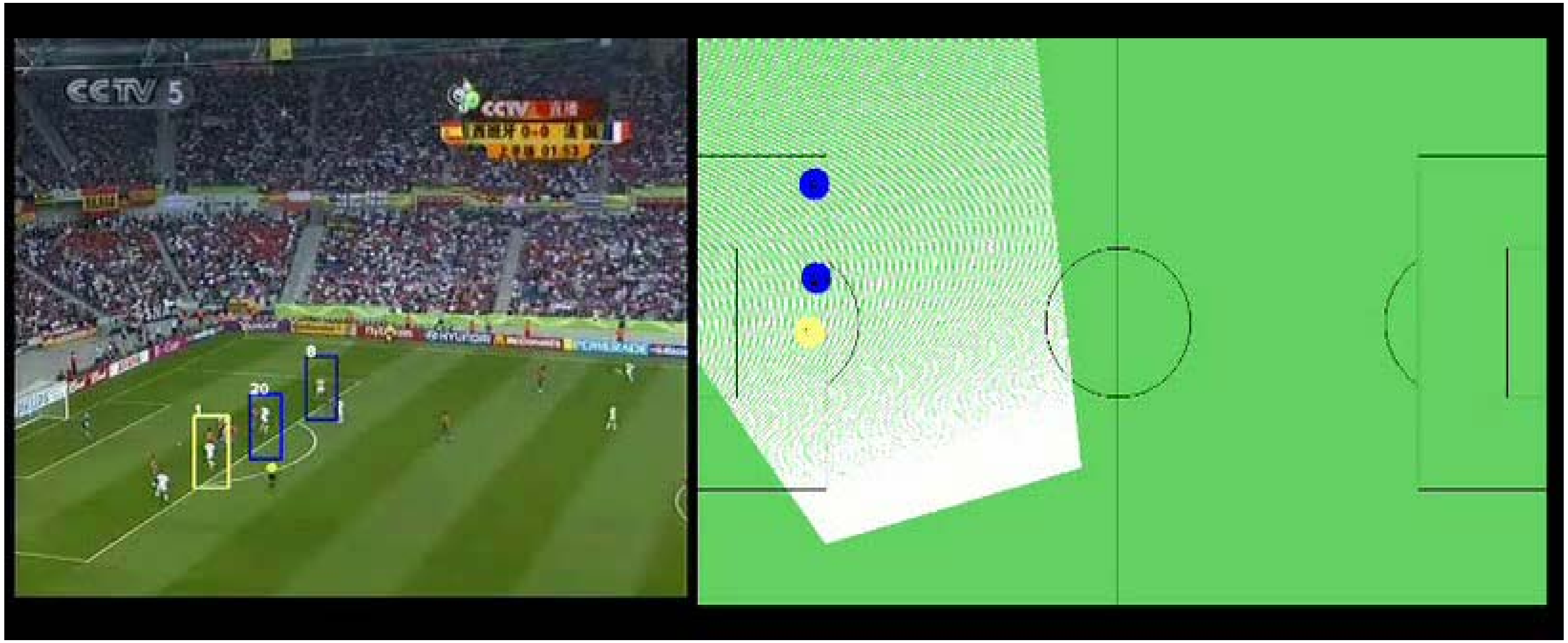
Face query

Retrieve relevant videos

- *Multi-modality (face / speaker), Hybrid (supervised + unsupervised)*
- *Promising results: 90% in news video, 60-80% in movie and home video*

Person recognition: Human detection / tracking

- Human detection/tracking
  - High detection accuracy: precision 92.38%, recall 88.82%
  - Tracking: 75%, some ability to handle merge and occlusion



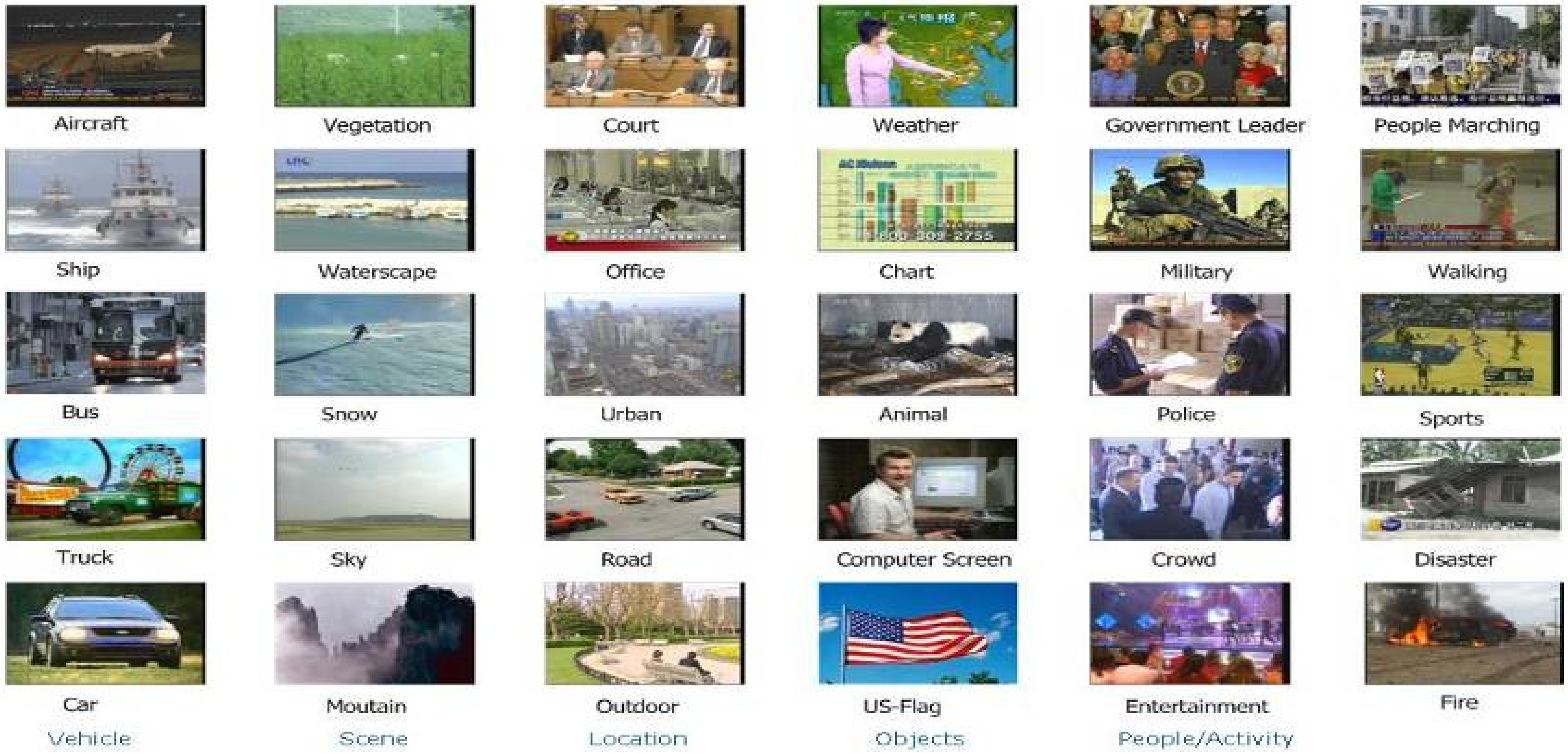
1.81 fps

8 objects, obj-size: 15x55, 100 particles/obj

Generic Concept Detection in TRECVID

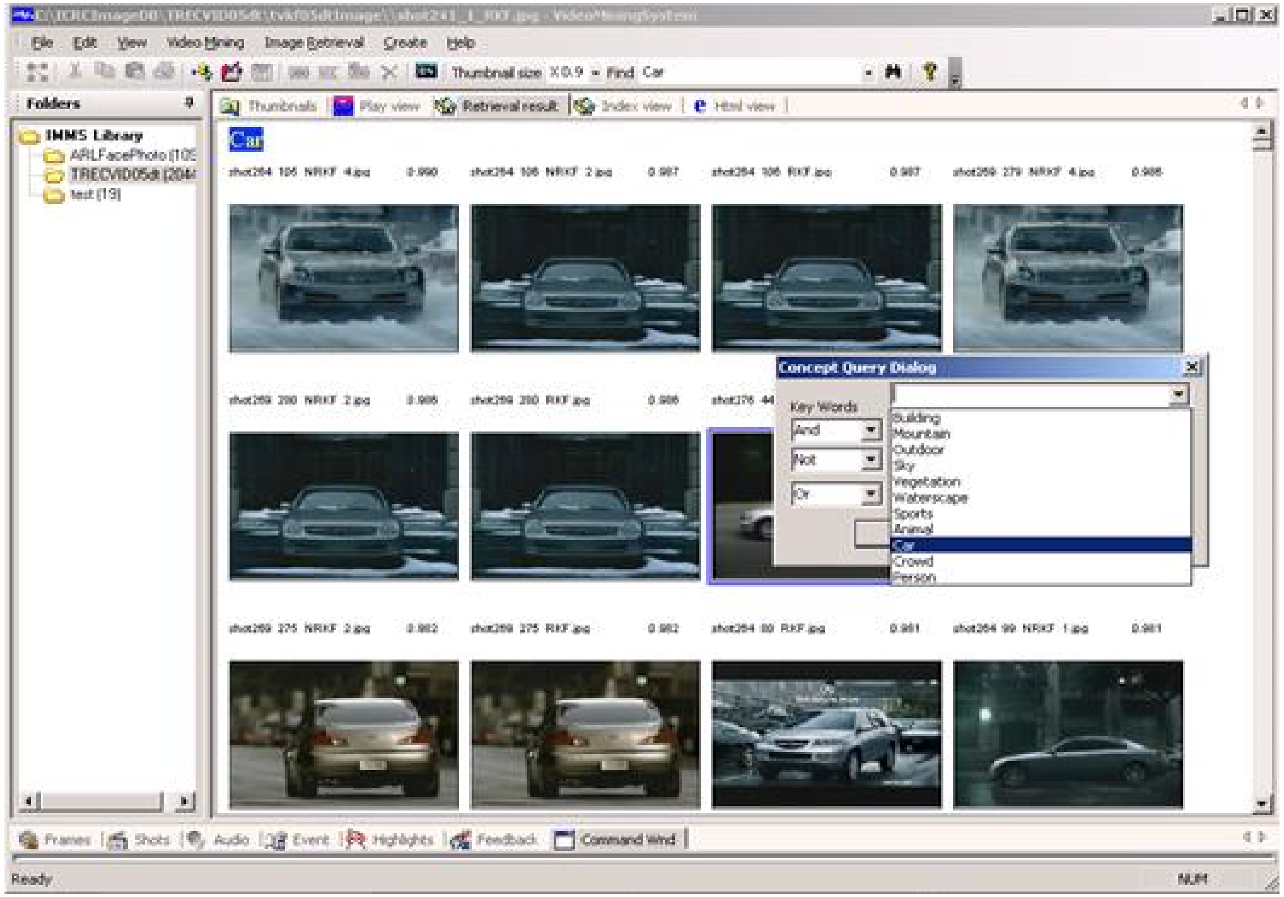
TRECVID

- Yearly international workshop sponsored by NIST for evaluation of research in content-based retrieval of digital video.
- High level feature extraction/concept detection (joint with Tsinghua)



Video Mining

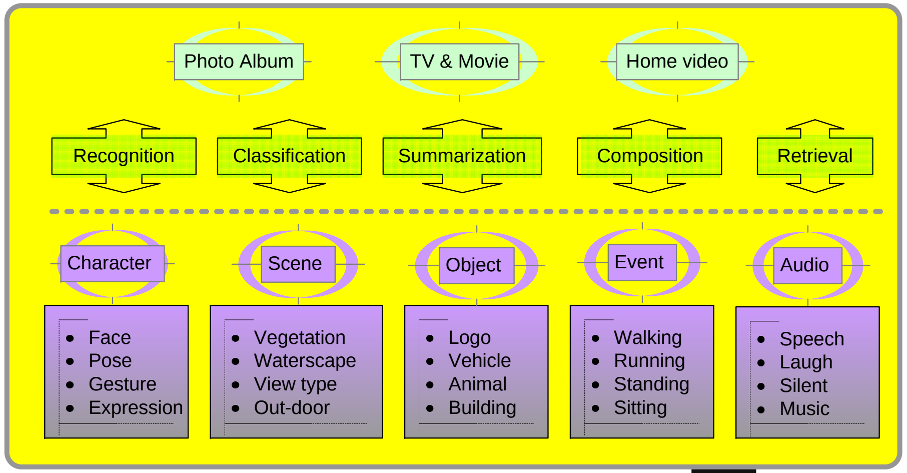
Concept Detection Results



Concept	Precision
Animal	0.6054
Building	0.5002
Car	0.626
Crowd	0.7068
Dog	0.2778
Food	0.6656
Mountain	0.5498
Outdoor	0.9464
TV-Screen	0.6066
Sky	0.7934
Sports	0.7916
Vegetation	0.4424
Walk/Running	0.4522
Waterscape	0.6075
Person	0.9745

- ***Fully automatic***
- ***State-of-the art accuracy***

Future media applications



Organized by person/face

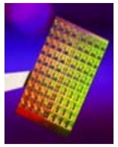
Photo/video Organizer

Search by person/place/event/ example image

Media Search

Script-based authoring

Coarse to fine customization

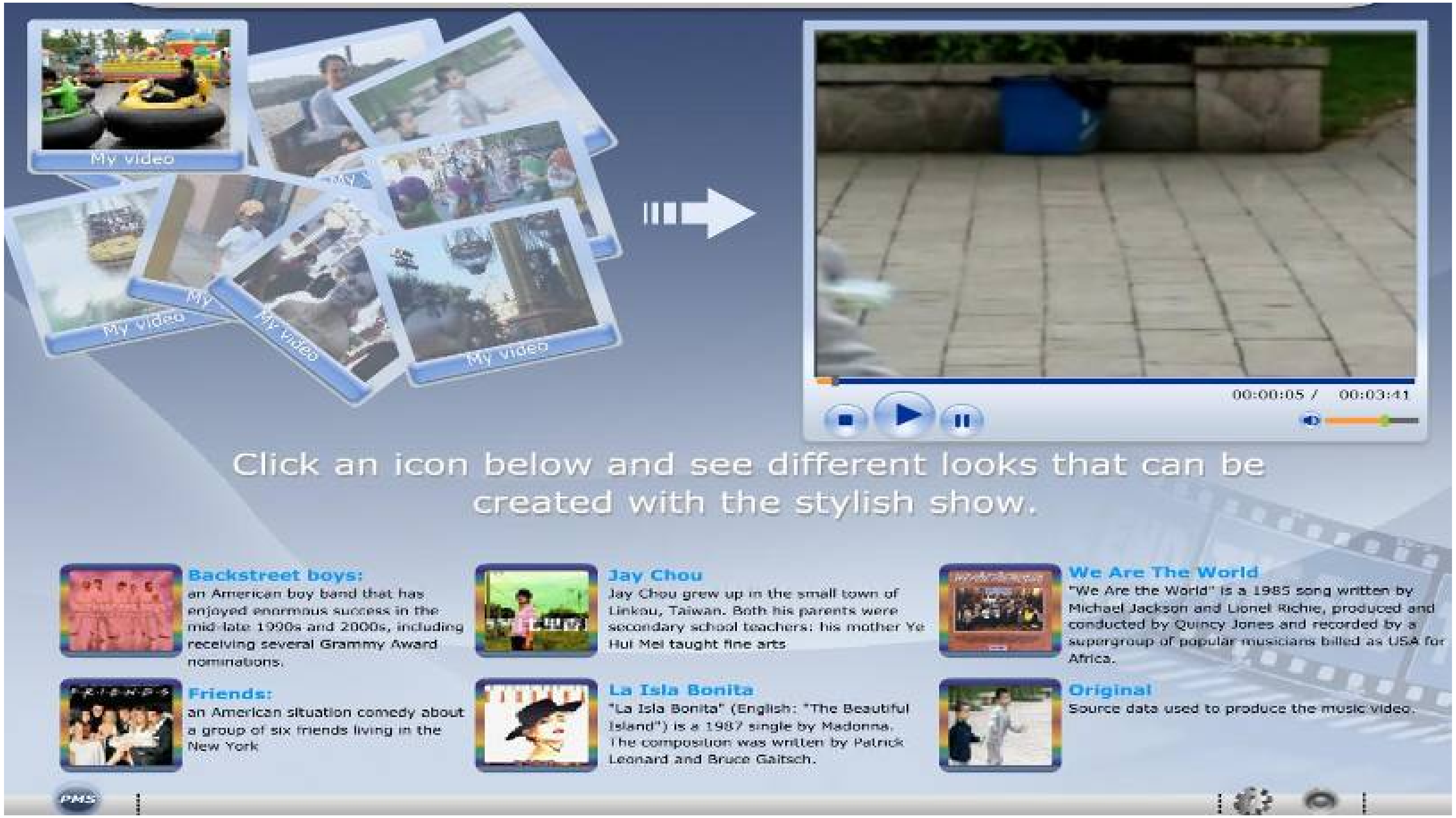


Multi-core / Many-core platform

Video Editing

Video Editing

Sample usage model: Automatic personalized music video generation



Click an icon below and see different looks that can be created with the stylish show.

Backstreet boys:
an American boy band that has enjoyed enormous success in the mid-late 1990s and 2000s, including receiving several Grammy Award nominations.

Jay Chou
Jay Chou grew up in the small town of Linkou, Taiwan. Both his parents were secondary school teachers: his mother Ye Hui Mei taught fine arts

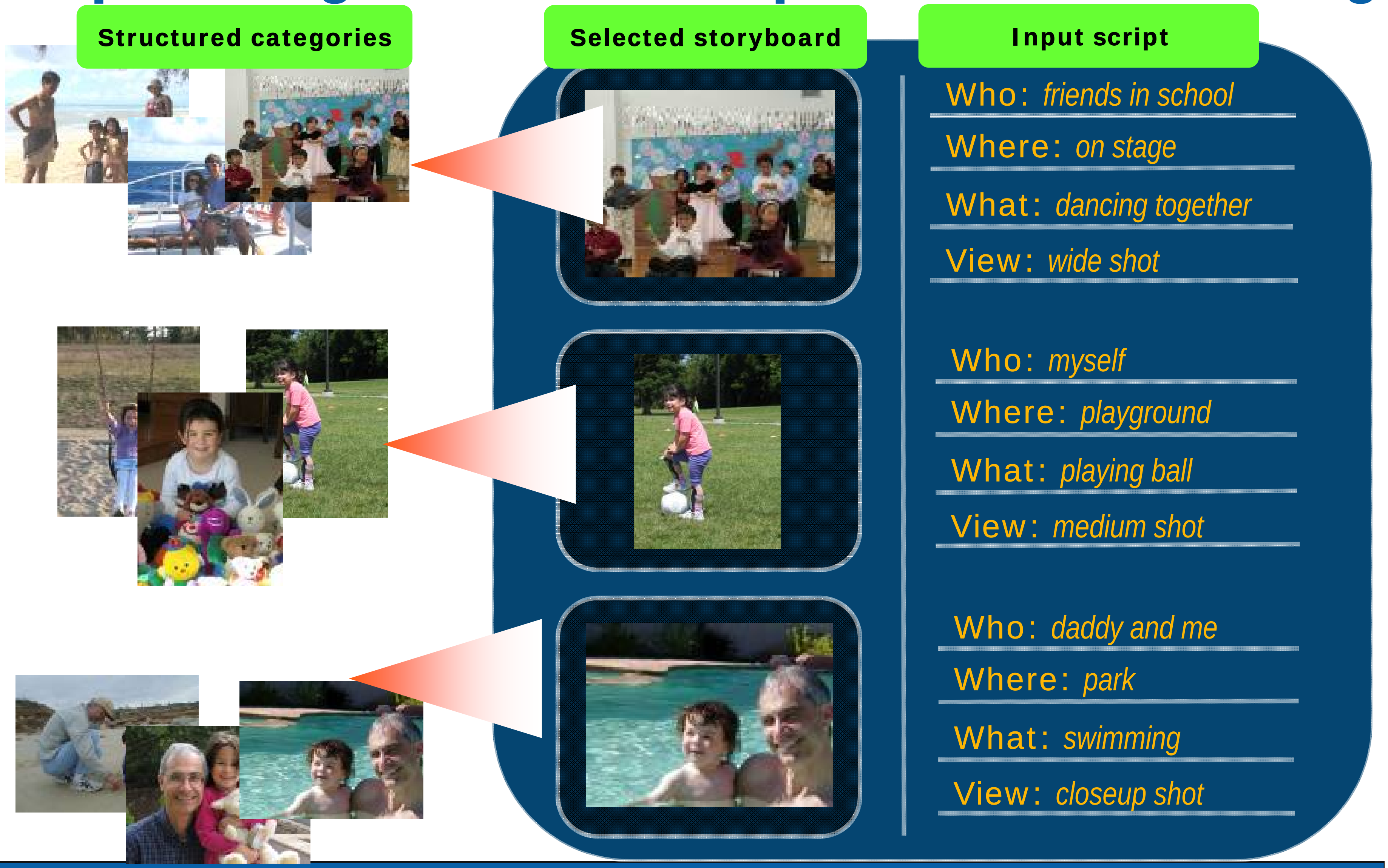
We Are The World
"We Are the World" is a 1985 song written by Michael Jackson and Lionel Richie, produced and conducted by Quincy Jones and recorded by a supergroup of popular musicians billed as USA for Africa.

Friends:
an American situation comedy about a group of six friends living in the New York

La Isla Bonita
"La Isla Bonita" (English: "The Beautiful Island") is a 1987 single by Madonna. The composition was written by Patrick Leonard and Bruce Gaitsch.

Original
Source data used to produce the music video.

Sample usage model: Script Based Authoring



Structured categories

Selected storyboard

Input script

Who: *friends in school*

Where: *on stage*

What: *dancing together*

View: *wide shot*

Who: *myself*

Where: *playground*

What: *playing ball*

View: *medium shot*

Who: *daddy and me*

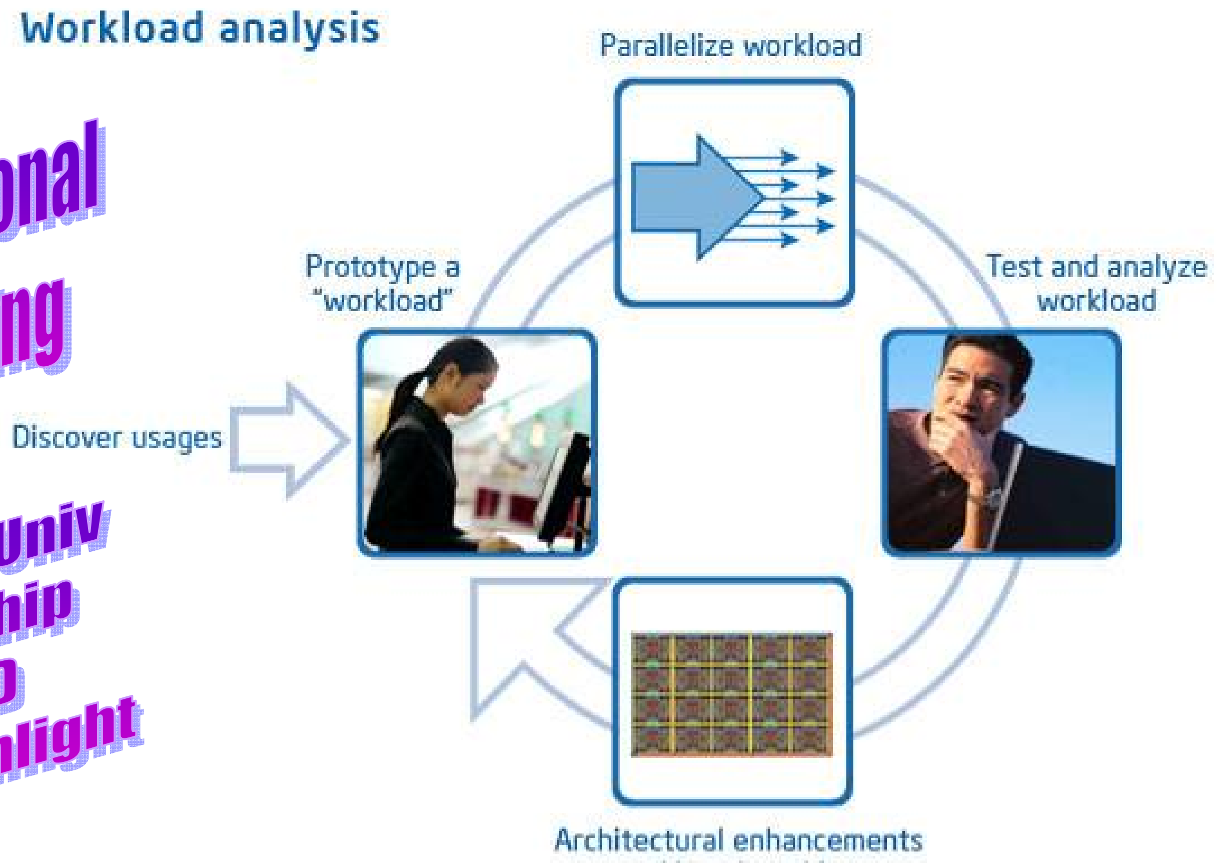
Where: *park*

What: *swimming*

View: *closeup shot*

Tera-Scale Applications Research

Future Personal Video Mining

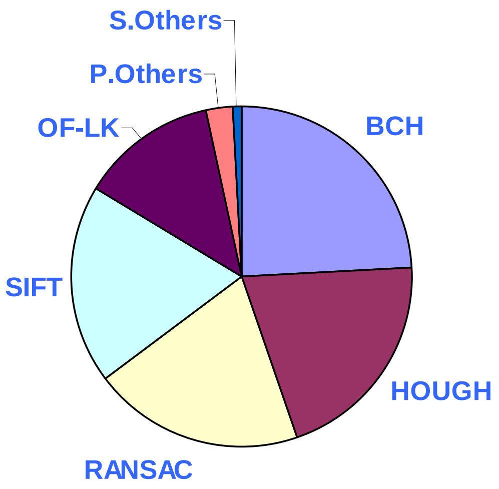


TsingHua Univ Partnership

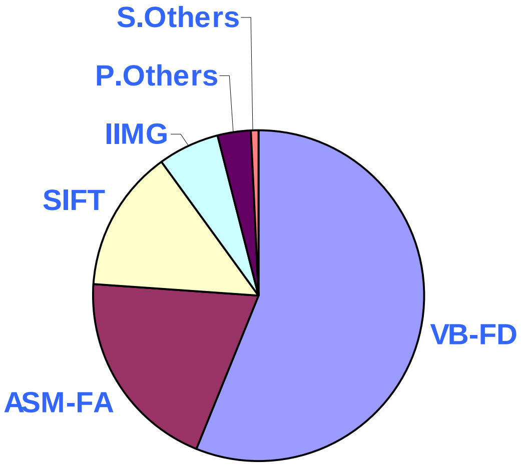
TRECVID

Soccer Highlight

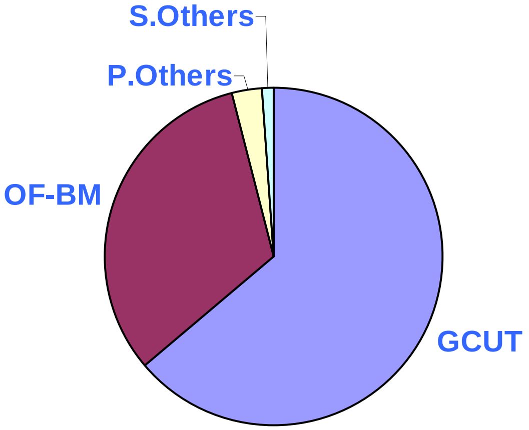
Profiling of Video Mining System



Sports Video Analysis



Video Cast Indexing



Salient object detection

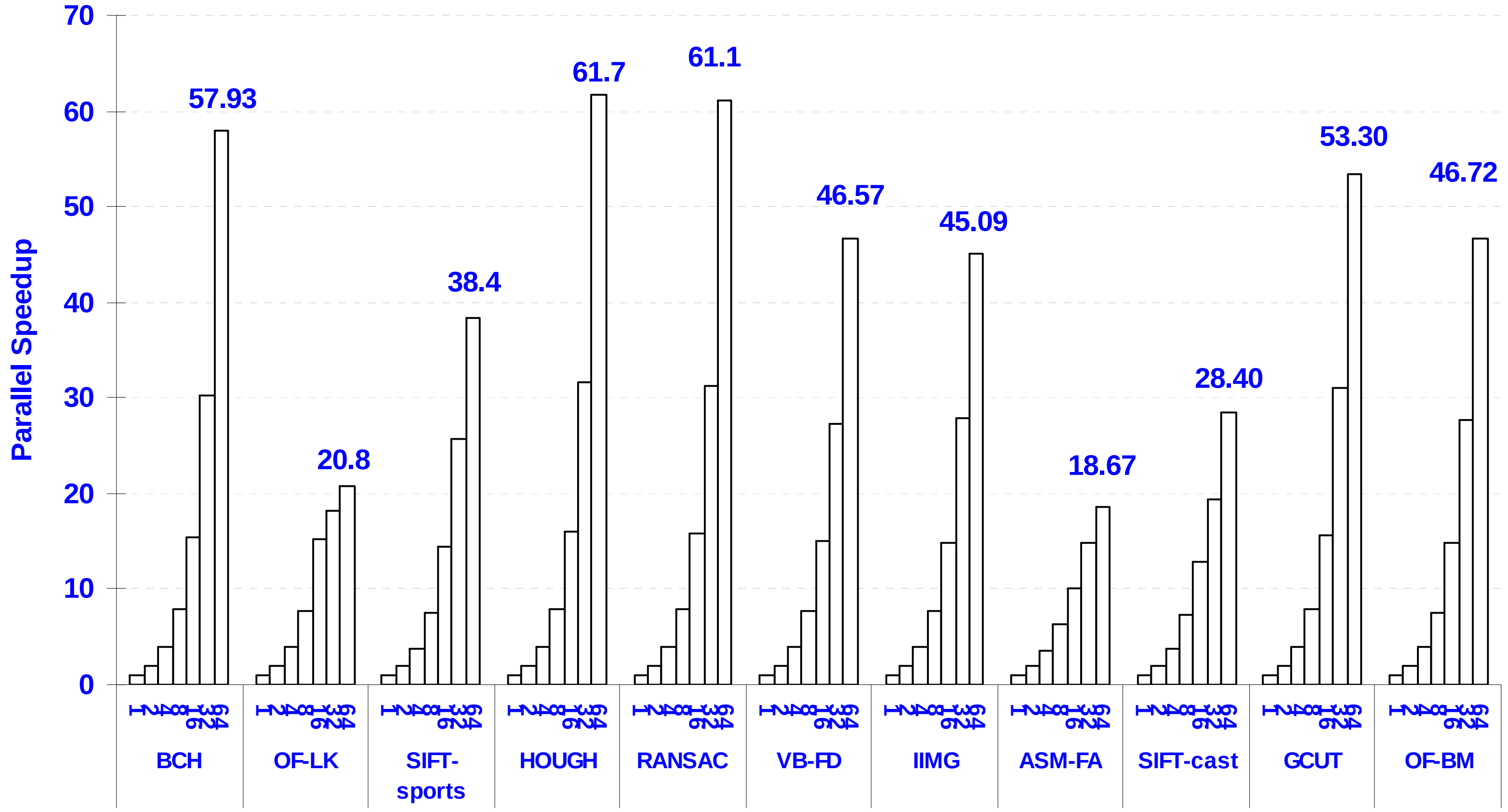
Parallelize codes accounting for more than 99% of total execution time

Parallelization: Coarse-Grain vs. Fine-Grain

Granularity	**Coarse**	**Fine**
Parallelization	Between frames	Tiling within frame
Memory BW Requirements	High	Low
Programmability	Easy	Difficult

- ***Future system may need to support both***
- ***Parallelization is not as easy as it looks (even for coarse-grain)***

Parallel Scaling Performance
--- Fine-grain parallelization



- Most algorithms scale very well up to 64 cores in simulation
  - Many useful feedbacks on multi-core architecture
- The full applications achieve 47x, 37x, and 53x speedup on 64 cores

Summary

- Future CPU performance increases will be primarily achieved through multi-core parallelism

- Intel Tera-scale research aims to enable a wide range of compelling, compute intensive applications including visual computing.

- Intel is driving research as well as industry and academic collaboration to solve HW and SW challenges.

- The Intel China Research Center is developing advanced, highly parallel computer vision algorithms which could enable compelling multi-media search, editing and facial recognition applications.

intel™

Leap ahead™