5th Neuro Inspired Computational Elements Workshop (NICE 2017)

Machine Intelligence for Mobile Augmented Reality

- Requirements in HW & SW towards Commercialization -

Hyong-Euk (Luke) Lee, Ph.D.

Principal Researcher

March 6, 2017

SAIT, Samsung Electronics Co.

SAMSUNG

Contents

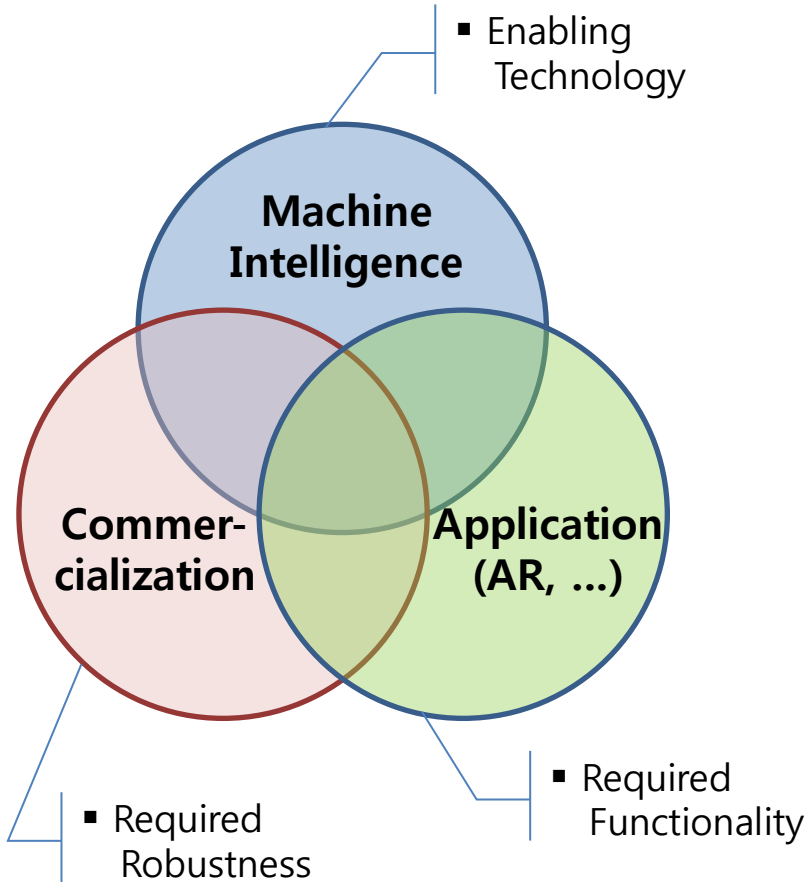
- ☐ Introduction
- ☐ A brief overview of cognitive applications
- ☐ The issues and requirements for mobile augmented reality
  : accuracy, response time, and h/w acceleration
- ☐ The functional requirements for future applications
- ☐ Concluding remarks

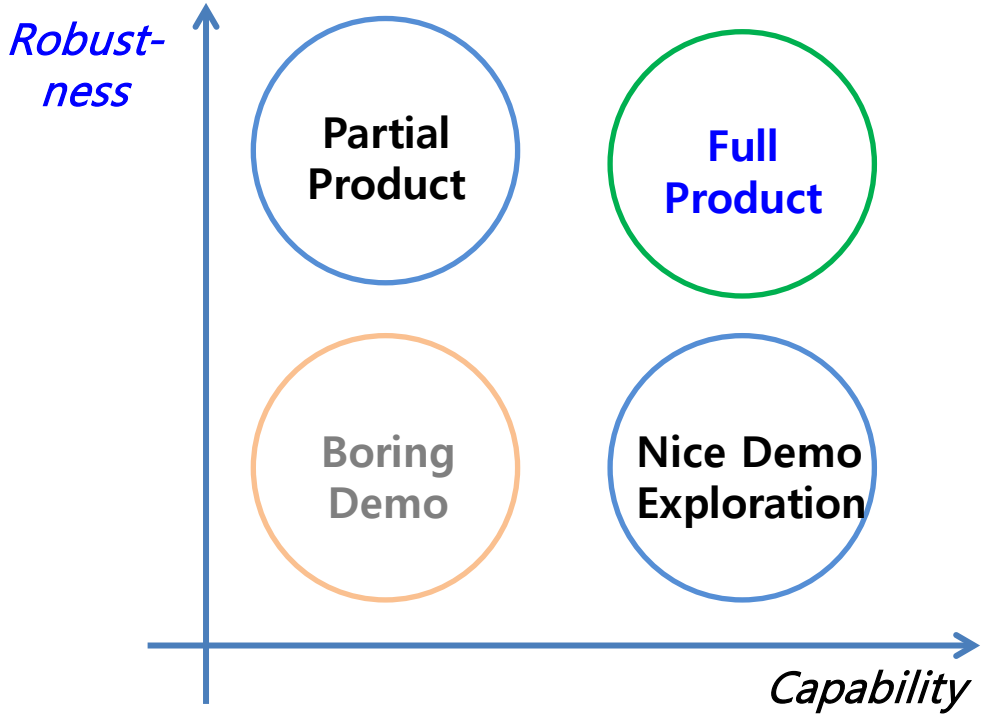
1. Introduction (1/3)

creation+

What do we need to consider?

Keywords:





[Ref. Invited talk by Dimitro Dolgov (Waymo/Google) in AAAI 2017, "The Consilience of Natural and Artificial Reinforcement Learning"

How to achieve robustness?
→ Concrete problem formation (target func. & eval. criteria) is important!

Example. Problem Formulation (1) - Function

Changes in the two major capabilities for smartphone:

Examples of Smartphone

Imaging

Graphics

	Galaxy S1 (2010)	Galaxy S7 (2016)
Imaging	***quality matters!*** 5M pixels (S1), Low illumination? (1/3.2”, S2)	***quality matters!*** 12M pixels, Low illumination! (1/2.5”, S7)
		management matters! Too many pictures! Make it fun(SNS..)!
Graphics	***quality matters!*** GPU 3.2 GFLOPS	***quality matters!*** GPU 519.2 GFLOPS
	efficiency matters! 480x800 (WVGA) / 3.97” Disp., + (Low quality) UI & Simple Game	***efficiency matters!*** 1440x2560 (WQHD) / 5.1” Disp., + (High quality) Game & VR

YEAR

- The target functions (capabilities) are usually defined by the expected UX. (based on the user expectations, market trend analysis, competitors, …)

1. Introduction (3/3)

Example. Problem Formulation (2) – Specification : *procedures in graphic app.*

Trends

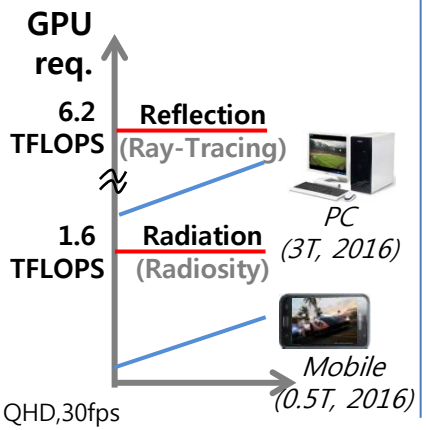
Graphic Quality: Mobile vs. Console =~10yrs GAP

PS3 (446.8GFLOPS, '06)

Galaxy S5 (150GFLOPS, '14)
※ G-S7: 519 GFLOPS

Function

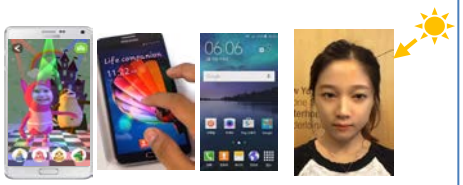
graphic rendering w/ indirect lighting



Specification #1

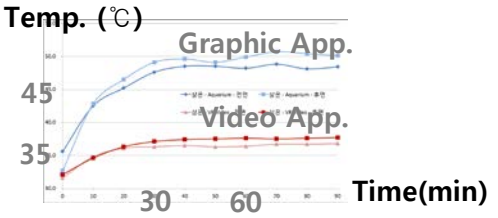
[Basic FPS]
- For video: 30fps
- For game: 60fps
- For VR: ≥30fps/eye (c.f. 90fps@PC)

[Application-specific]
- FPS & loading time:
  1) Indep. App.: ~60fps ≤400ms
  2) Home/Lock-screen UI: ≥60fps (no-drop), ≤100ms@page-turn
  3) Camera after-effect <10ms @ mem. <50MB



Specification #2

100% GPU operation (1000mA@Note4)
→ Temp. Increase (@VR)
: Current Req. <700mA



Smartphone + VR

→ Implementation: SW algorithm to reduce calculation to catch up the HW perf. gap,
+ Low-level optimization/HW-acceleration for power consump. reduction.

2. Brief Overview on Cognitive Applications (1/3)

creation+

☐ **Machine Intelligence could be used in a wide variety of Samsung applications**

Mobile Biz.
: Smartphone, Tablet, …

- Identification
- Authentication
- Location-based Service

Display/Home Biz.
: Smart TV, Home Appliances, …

- Personalized Multimedia
- Security/Surveillance
- Home-assistive robot/ Companion for elderly

Semiconductor Biz.
: Mobile AP, IoT

- AP, VPU
- Neural Processor
- IoT

Car Component Biz.
: Connectivity, HUD, …

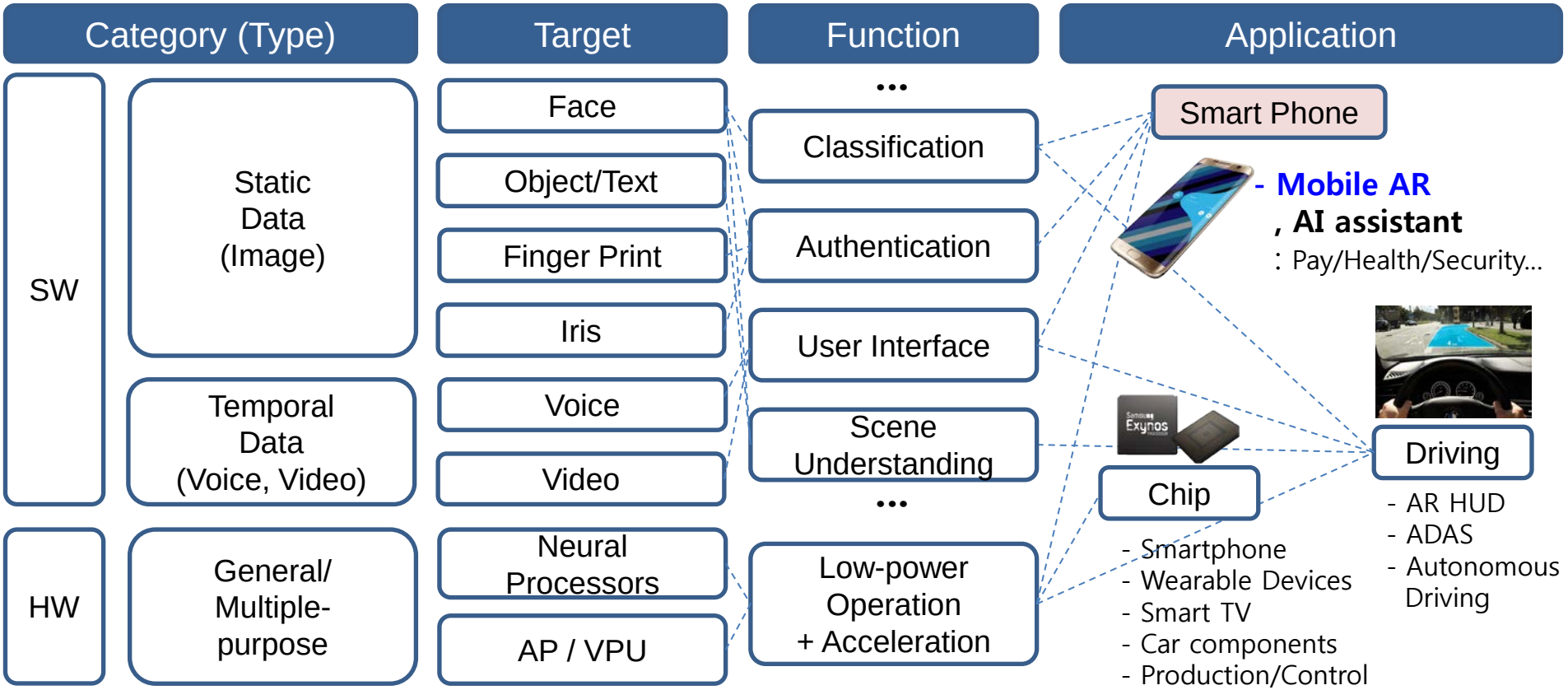
- User Interface
- Authentication/ Connectivity
- Co-pilot

☐ **In early stage of cognitive applications were focused in its 'recognition' capability**
: examples – finger print, facial expression, voice recognition, etc.

2. Brief Overview on Cognitive Applications (2/3)

creation+

- ☐ **Static** (Image) → **Temporal** (Voice, Video) **Data**
- ☐ **SW-only** → **HW-combined** (GPU-accelerated, VPU/AP)
- ☐ **Non-accurate** / **Specific** / **Not-necessarily Practical** (image classification)
  - → **Accurate** / **Specific** / **Practical** (authentication)
  - → **Accurate** / **General** / **Practical** (mobile AR/AI assistant)



2. Brief Overview on Cognitive Applications (3/3)

creation+

☐ **Augmented Reality**

: [Def.] a live direct or indirect view of a physical, real-world environment whose elements are augmented (or supplemented) by computer-generated sensory input such as sound, video, graphics or GPS data. (from wikipedia)

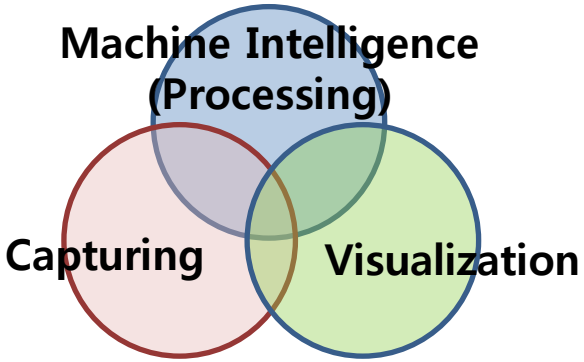



☐ **Basic Philosophy**

1) Augmented reality in terms of **augmenting 'human sensing & intelligence'**!

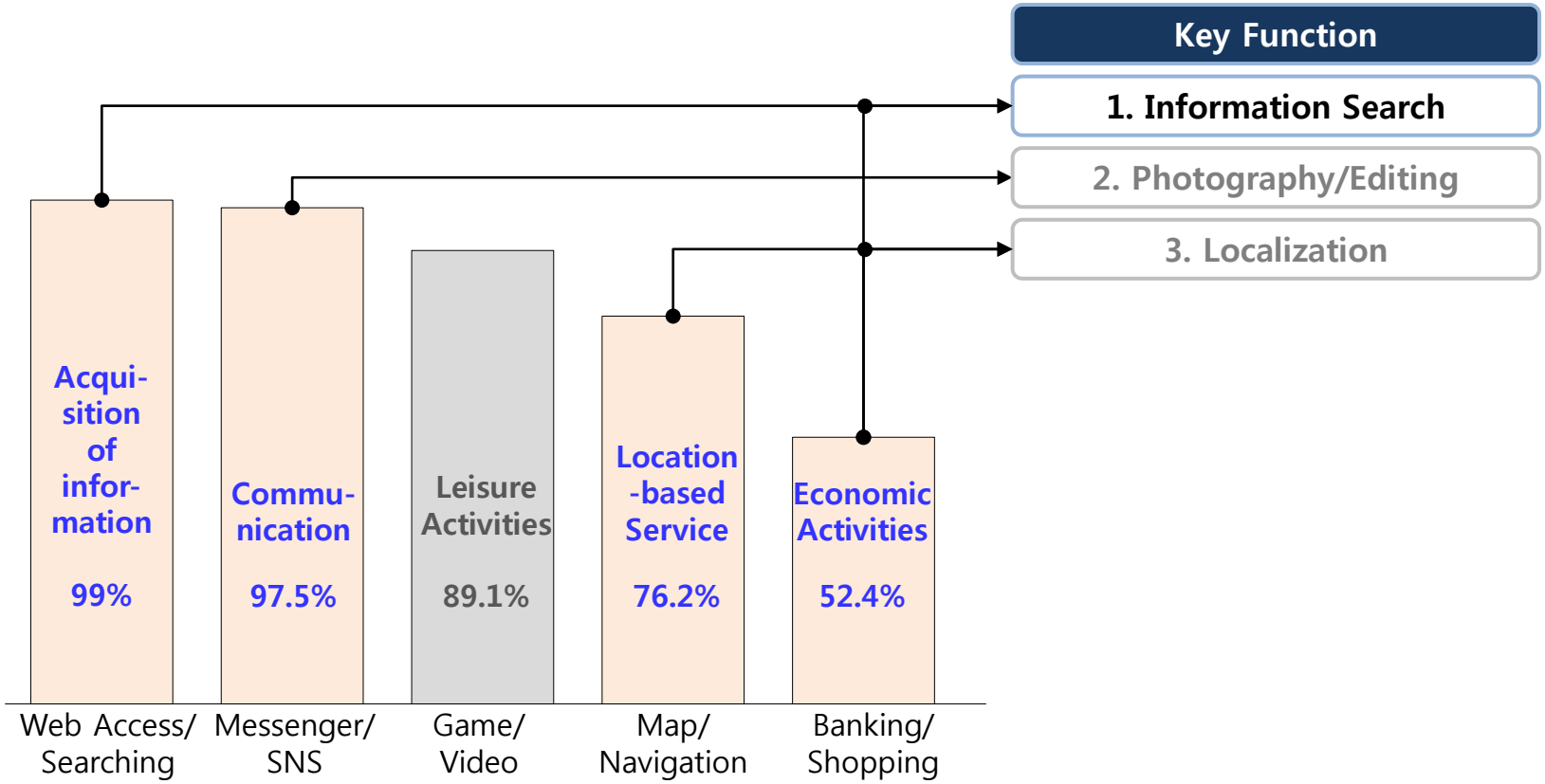
2) Smartphone itself is a nice device for personal AR!

(except some specific application like AR-HUD)



3. Mobile Augmented Reality – Scenario (1/3)

☐ **Where can we find a chance for 'practically useful' AR?**
: Insight from user's behavioral pattern



[Purpose of Mobile Internet Usage *]

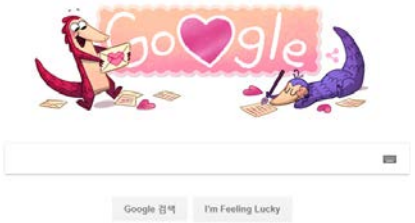
→ *Visual Search*

* 2014 Survey on Mobile Internet Usage, Korea Internet & Security Agency

3. Mobile Augmented Reality – Scenario (2/3)

☐ ***Visual search*** **can provide a 'new functionality' for searching activities**

If I know the keyword,



Text

Voice

But what if we don't know the keyword ?

Complex Keywords?

3. Mobile Augmented Reality – Scenario (3/3)

☐ **One potential scenario**

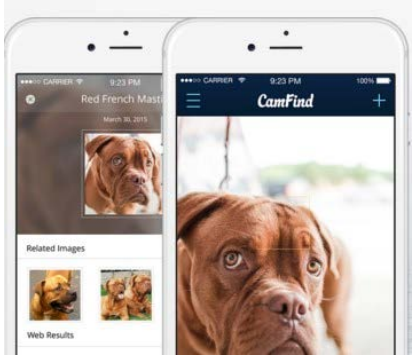
: *Product Visual Search - O2O (online-to-offline) …. → AI Assistant (+Voice/Text)*

☐ **Major requirement – Accuracy**

: inaccurate recognition → # of users will be rapidly reduced!

[Wine Recog. App.]

[car – voice recognition]



[CamFind App.]

3. Issues and Requirement (1) – Accuracy (1/3)

☐ **Function: Product Information Recognition**

- Technical Issues : Inter-Class Separability vs. Intra-Class Separability

☐ **Additional technical issues**

: (Environmental Condition) Illumination, Variable orientation, …

: (Maintenance) Product Information Update, Labeling, …

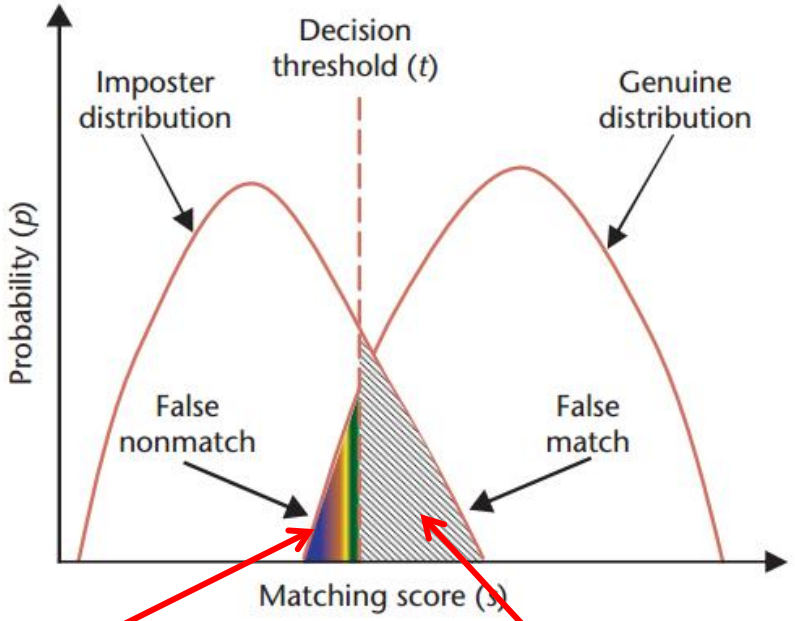
☐ **What is important in AR – visual search?**

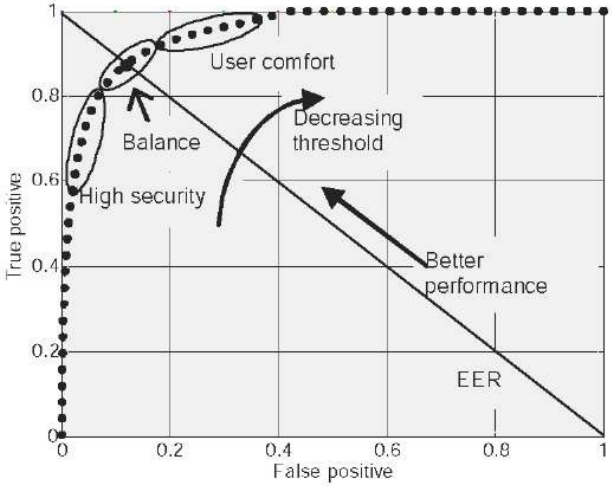
: Fine-grained recognition for object recognition

+ Property recognition for visual search (color, material property, …)

3. Issues and Requirement (1) – Accuracy (2/3)

- **Evaluation Criteria - FAR (False Acceptance Rate / Type 2 Error):**
  - measures the percent of invalid inputs that are incorrectly accepted





[FAR/FRR]*

High FRR : uncomfortable!! **High FAR : unsecure!!**

* http://what-when-how.com/artificial-intelligence/biometric-security-technology-artificial-intelligence/

3. Issues and Requirement (1) – Accuracy (3/3)

creation+

- **(Minimum) Requirement for Face Recognition**

  - **Authentication : 97%@FAR 1%**→ 99%@FAR 1%**, 100ms~1s, 50MB**
    : (Ref) [Finger Print] 96% @ FAR 1% =~ 85% @ FAR 0.1%
    [Iris] 99.4%@ FAR 1% =~ 94% @ FAR 0.1%,
    [Iris/Finger + α (Combined)] 90% @ FAR 1/10M)

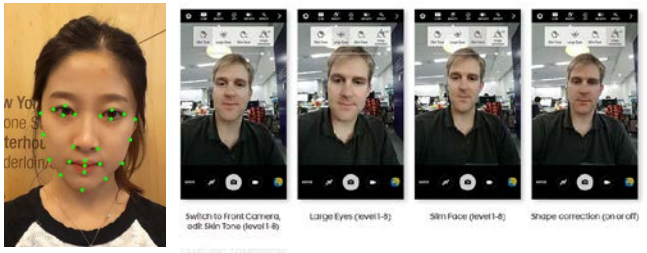
  + Liveness?
  + Secure storage?

  - **cf. the other applications:**
    . Image Classification (Gallery) : 90%@Recall 75% (2D Face)
    . Image Editing (Face Detection) : N/A (FRR than FAR), <10ms
    . Voice Recognition: ~@SNR 5dB

Accuracy | Speed | Memory | Power

Authentication

Auto-Tagging



Camera App.

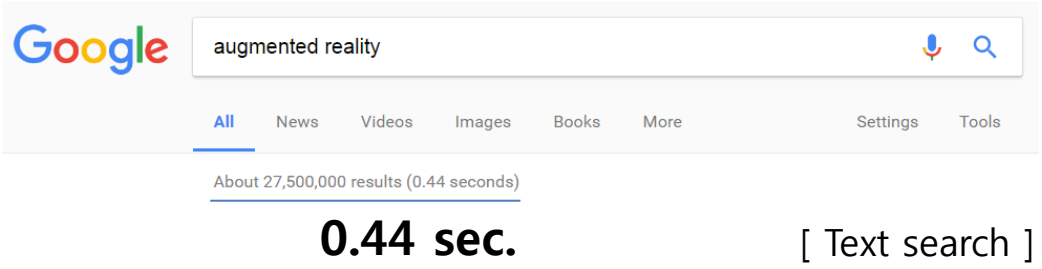
3. Issues and Requirement (2) – Response time

☐ **Response Time – the basic advice [Miller 1968; Card et al. 1991]:**

- **0.1 second** is about the limit for having the user feel that the system is reacting **instantaneously**, meaning that no special feedback is necessary except to display the result.
- **1.0 second** is about the limit for the **user's flow of thought** to stay uninterrupted, even though the user will notice the delay. Normally, no special feedback is necessary during delays of more than 0.1 but less than 1.0 second, ……
- **10 seconds** is about the limit for **keeping the user's attention** focused on the dialogue. For longer delays, users will want to perform other tasks while waiting for the computer to finish, so they should be given feedback indicating when the computer expects to be done….

☐ **Web-based Application Response Time [Jakbob, Usability Engineering, 1993]:**

- 0.1 second: Limit for users feeling that they are directly manipulating objects in the UI.
- 1.0 second: Limit for users feeling that they are freely navigating the command space without having to unduly wait for the computer.



0.44 sec.

[Text search]

0.85 sec.

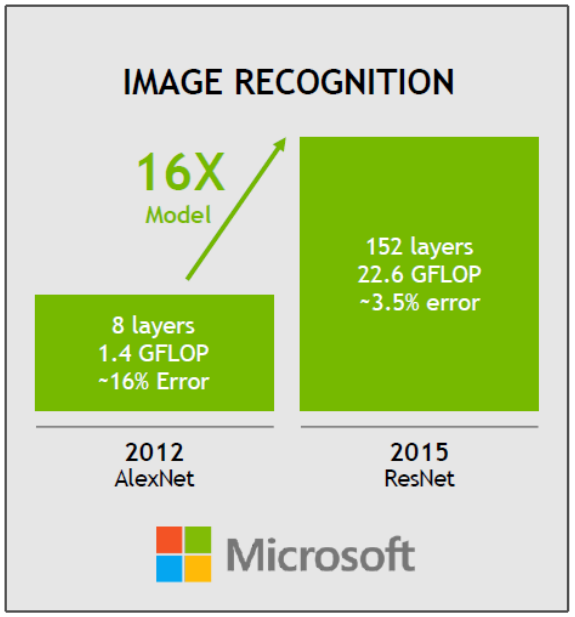
[Image search]

* Jakob nielson, "Usability Engineering", 1993

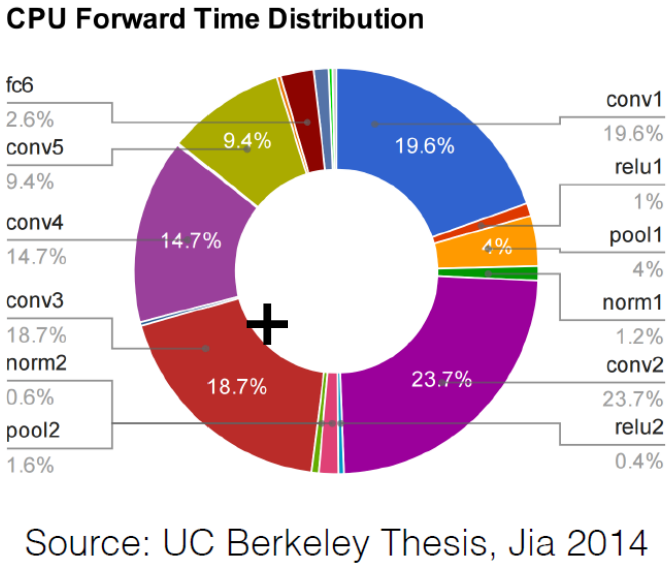
3. Issues and Requirement (3) – HW acceleration/power (1/3)

☐ Computational Cost – Can SW itself solve the problem?

[Model Complexity*]



[Convolutional Computation]



Source: UC Berkeley Thesis, Jia 2014

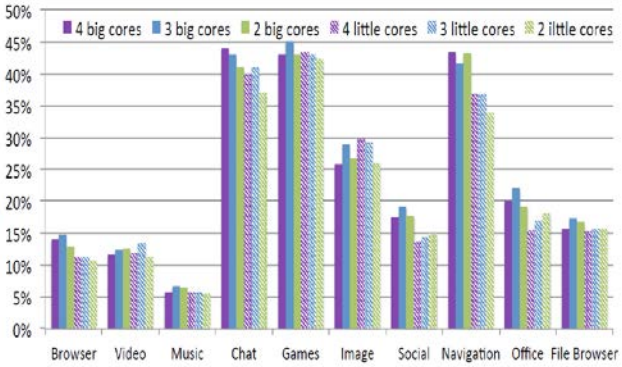
→ HW acceleration is required.

: GPU?

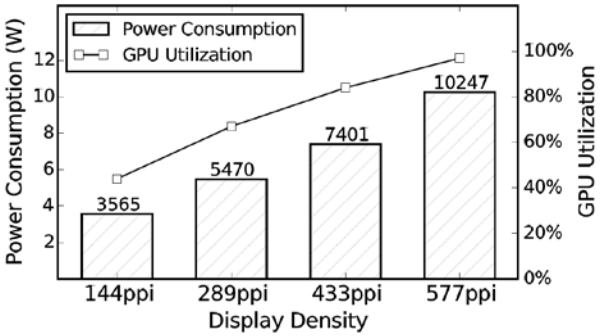
*Bryan C. (NVidia), "More than a GPU: Platforms for Deep Learning", Samsung AI Summit (1/9)

3. Issues and Requirement (3) – HW acceleration/power (2/3)

☐ **CPU vs. GPU** : FLOPs/BYTE (similar), FLOPs/Cycle (GPU>CPU), … ,
However, GPU is still busy & CPU-GPU Communication is also an issue.



[GPU Utilization of Different Category of Apps]
(Exynos 5410 SoC, under various CPU workload)



[System power and GPU utilization, Galaxy S5/Q Adreno 420]
: 577 ppi = 2560x1440 res @ 3D Graphics Rendering

☐ **Power & Efficiency**

ARITHMETIC

Mixed precision for training

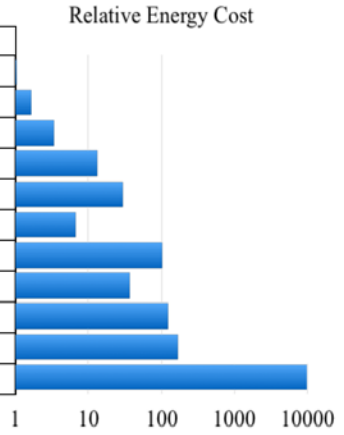
FP32 + FP16

Lower precision integer for inference

Int8

Relative Energy Cost

Operation:	Energy (pJ)
8b Add	0.03
16b Add	0.05
32b Add	0.1
16b FP Add	0.4
32b FP Add	0.9
8b Mult	0.2
32b Mult	3.1
16b FP Mult	1.1
32b FP Mult	3.7
32b SRAM Read (8KB)	5
32b DRAM Read	640



1) Cao Gao Et.al, "A Study of Mobile Device Utilization", IEEE Int'l Symposium on Performance Analysis of Systems and Software, 2015
2) Songtao He Et.al, "Optimizing Smartphone Power Consumption through Dynamic Resolution Scaling", ACM MobiCom 2015
3) Mark Horowitz, "Computing's Energy Problem (and what we can do about it)", ISSCC 2014

3. Issues and Requirement (3) – HW acceleration/power (3/3)

- **GPU → VPU → GPU+VPU → Integrated SoC / Neural Processors** (+Memory)

Pre-training vs. Continuous learning …

Fast Inference | **On-Chip Learning**

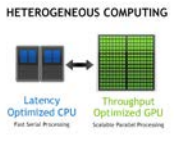
Size, Accuracy, Variance of initialization …

Digital Memory | **Analog Memory**

Type of Data …

ANN | **SNN**

\+ **High-bandwidth Access to CPU Memory**



\+ **Scalability**

\+ ….

[c.f. scaling product – Nvidia]

Time-to-market/ Application	[Circuit Type] Approach	Algorithm	Memory
Short-term / ADAS, …	**[Digital Circuit]** - GPU-based initial learning (recognition acceleration) + On-chip fine tuning	**ANN**	**SRAM+ DRAM**
Mid-term / …	**[Mixed (Digital + Analog) Circuit]** - NVM-based DL acceleration (ANN) : On-chip learning (minimize circuitry with analog resistance)	ANN	PCM
Short-term for ANN Long-term for SNN / Visual Processing, Voice Recognition	**[Digital Circuit]** -ANN to SNN converting -SNN learning algorithm	**ANN** ↔ **SNN**	SRAM+ DRAM
Mid-term / Neural Processor	**[Digital Circuit]** - Ultra Low-Power Event-based Recognition Processor (Inference Acceleration)	SNN	SRAM
Long-term / Neural Processor	**[Analog Circuit]** - NVM-based DL acceleration (SNN/RBM) : On-chip learning (minimize circuitry with analog resistance)	**SNN** (RBM)	**PCM**

4. Functional Requirement for Future Applications (1/3)

creation+

- **AR in terms of 'Information Retrieval' – augmenting human intelligence**

  - **Step 1. Request & Answer**
    : Technology for 'convenient' interaction
    (e.g. text → voice → visual input)

  - **Step 2. Active Feed**
    : Technology for 'selective' information collection
    (e.g. news/video feed based on preference, product recommendation)

  - **Step 3. Interactive Agent**
    : Technology for 'real-time assistive' information search
    (e.g. conversational AI towards AI Assistant)

- **What will be essential?** ***The problem is getting closer to "Open-ended" one!***
  : "**Reasoning capability & continuous learning**"
  - We can't learn everything only with the collected data
    → Effective exploration based on learned/common sense knowledge is essential!
  - Knowledge could be modified, continuously & in parallel.
    → Explainable (Transferrable) AI, based on knowledge representation, is desired.

4. Functional Requirement for Future Applications (2/3)

creation+

- The average elapsed time between **key algorithm proposals and corresponding advances was about 18 years**, whereas the average elapsed time between **key dataset availabilities and corresponding advances** was less than 3 years, or about 6 times faster

[Ref] AAAI Invited talk by Xavier Amartriain/Quora

Year	Breakthrough in AI	Datasets (First Available)	Algorithms (First Proposal)
1994	Human-level spontaneous speech recognition	Spoken Wall Street Journal articles and other texts (1991)	Hidden Markov Model (1984)
1997	IBM Deep Blue defeated Garry Kasparov	700,000 Grandmaster chess games, aka "The Extended Book" (1991)	Negascout planning algorithm (1983)
2005	Google's Arabic- and Chinese-to-English translation	1.8 trillion tokens from Google Web and News pages (collected in 2005)	Statistical machine translation algorithm (1988)
2011	IBM Watson become the world Jeopardy! Champion	8.6 million documents from Wikipedia, Wikitionary, Wikiquote, and Project Gutenberg (updated in 2005)	Mixture-of-Experts algorithm (1991)
2014	Google's GoogLeNet object classification at near-human performance	ImageNet corpus of 1.5 million labeled images and 1,000 object categories (2010)	Convolution neural network algorithm (1989)
2015	Google's Deepmind achieved human parity in playing 29 Atari games by learning general control from video	Arcade Learning Environment dataset of over 50 Atari games (2013)	Q-learning algorithm (1992)
Average No. of Years to Breakthrough		**3 years**	**18 years**

- Data analysis enables revealing the problem, including the unconsidered cases, while evaluation criteria guides direction.
  : It is very important, however, is it still valid for open-ended problem?

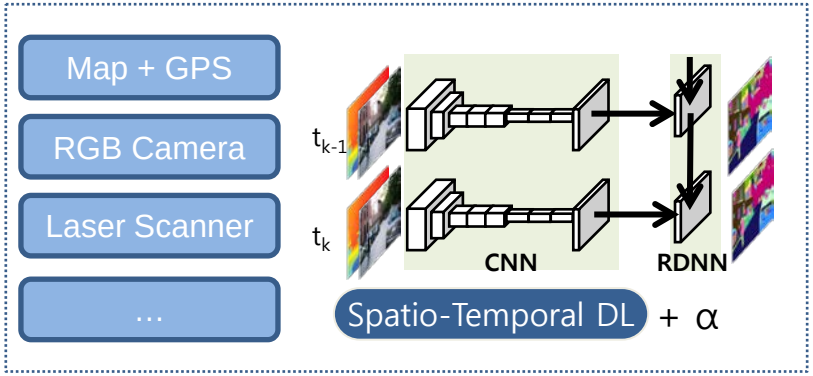
4. Functional Requirement for Future Applications (3/3)

creation+

- **Example: Scene Understanding in Autonomous Driving** – *An Open-Ended Problem*

  - It is difficult to handle every corner cases!!

    → Reasoning enables the best actions,
    based on the hypothesis, not by simple interpolation.



Can we learn 'underlying' rule of a driver?

* Note. Remembering everything could pretend to be intelligent,
in spite of poor reasoning capability.

Can we make a system learn' yield' in driving?

→ The important things are
"extracting underlying rules"
& "common sense reasoning"

Concluding Remarks

creation+

- **The requirements for MI applications have been discussed for mobile AR and the related cognitive applications**

  - accuracy, response time, and h/w acceleration and power consumption
  - application-specific accuracy requirement of recognition

- **The Next Challenges**

  : "(common sense) reasoning" and "continuous learning" will be essential towards handling open-ended problems

  - reasoning provides the best action based on its knowledge-based hypothesis

Appendix

1. Remarks on Reasoning Capability (1/2)

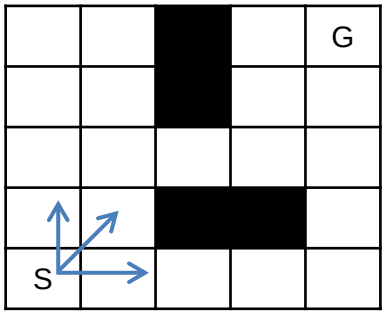
creation+

☐ **2 Examples : Can you distinguish btw 'intelligent' vs. 'pretending to be intelligent*'?**

* Note. Remembering everything could pretend to be intelligent, in spite of poor reasoning capability.

1) [Action Selection] Two mechanisms in conventional reinforcement learning

: In early stage of learning, # of exploration is more than # of exploitation.



1) Exploitation – mainly by probability/rewards
   : simple reasoning helps!

2) Exploration – mainly by random access
   : <u>common sense and high-level reasoning help</u>!

→ It could be a measure of intelligence in terms of unsupervised learning.

2) [Continuous Learning & Fast Decision] Fast mapping (in linguistics) **

: The child (2~3 yrs old), who knows the word 'puppy' as a name of dog,
can point out a picture of dog even when hearing 'doggy' for the first time.
→ by means of 'reasoning', based on the knowledge!

** Dogs have been recognized to have this capability (Science, 2004)
- J. Kaminski et. al, 'Word learning in a domestic dog: evidence for "fast mapping"', Science 2004 (Jun 11; 304(5677): 1682-3)

2. Remarks on Reasoning Capability (2/2)

creation+

☐ **Recapping the two point of desired functions,**

1) Meaningful extraction of implicit rules

2) Intelligent action selection

☐ **The potential items to be investigated are**

1) Clarification of 'common sense' as a set of specified functions and relation

(e.g. learning hierarchical SDR as reconfigurable knowledge representation)

2) Flexible association of the existing knowledge

(e.g. hippo campus modeling)

Q. Eventually, can we make a system learn' yield' in autonomous driving?

[Supplementary #1] Fast Mapping Capability in a Dog

One of the experiments :

☐ **Step 1. Rico (a dog) has been trained to learn 200 words to pick up the corresponding object.**

: Rico can pick up the object which is told to do.

☐ **Step 2. 7 learned objects (among the 200 words that it has learned)**

and 1 unlearned object has been displayed in front of Rico.

☐ **Step 3. The new word (which is corresponding to the unlearned object) is spoken to Rico.**

☐ **Result : Rico could pick up the unlearned object!**

- Rico understood that there was one unlearned object quickly,
  then it concluded that the new word could be matched to the object based on reasoning
- And then, this experience could be a seed for (unsupervised) learning the new word.

[REF] J. Kaminski et. al, 'Word learning in a domestic dog: evidence for "fast mapping"', Science 2004 (Jun 11; 304(5677): 1682-3)