Exciting Media

Limsoon Wong
Plan

• I will discuss some made-in-Singapore innovations on the handling and processing of native media
  – New things that you can do with texts
  – New things that you can do with images
  – New things that you can do with audio
  – New things that you can do with video
New things that you can do with texts
Intelligent information extraction, improve safety

- Benefits
  - OHD: Check 100% of MSDS (currently < 10%) with same manpower
  - Chemical Suppliers: Savings in distribution of MSDS as it is online
  - End users of chemicals: Better quality MSDS, improved safety

- Extract chemical safety information from Materials Safety Data Sheets (MSDS)
- Check for conformance to standards
**How is it done?**

**What's needed to get it done?**

Erythropoietin stimulates transcription of the TAL1/SCL gene and phosphorylation of its protein products.

Activation of the TAL1 (or SCL) gene, originally identified through its involvement by a recurrent chromosomal translocation, is the most frequent molecular lesion recognized in T-cell acute lymphoblastic leukemia. The protein products of this gene contain the basic-helix-loop-helix motif characteristic of a large family of transcription factors that bind to the canonical DNA sequence CANNTG as protein heterodimers. TAL1 expression by erythroid cells in vivo and in chemical-induced erythroleukemia cell lines in vivo suggests that erythropoietin may influence whether the erythroid spleenic erythroid leukemia cell line.

These results were replicated with the TAL1 locus in Friend mouse erythroleukemia cell line, a member of a
New things that you can do with images
Make computers easier to use

- Abstraction of image content allows interpretation & matching in semantic space

- Visual query language allows specification of what and where

Search photos by visual keywords
How is it done?
What’s needed to get it done?

- Trained visual keywords for semantic detection and summarisation
- Automatic indexing using such keywords

Faces:
Crowd:
Buildings:
Foliage:
Let machines perceive as we do

- **Perceptual visual quality according to characteristics of human vision**

  - Adoption in video coding results in efficiency & quality improvement (other systems make compromise betw. the two)

<table>
<thead>
<tr>
<th>Knowledge &amp; Task</th>
<th>Visual Contents</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>High level contents</td>
<td>Low level Contents</td>
<td></td>
</tr>
</tbody>
</table>

- **Visual Attention**
- **Eye Fixation & Movement**
- **Retina & Vision Path**
- **Perceptual Quality**

<table>
<thead>
<tr>
<th>Comparison with MOS (Mean Opinion Score)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Correlation</td>
</tr>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td>PSNR metric</td>
</tr>
<tr>
<td>New metric</td>
</tr>
</tbody>
</table>

- Higher PSNR!
- New perceptual metric perceives correctly.
- Better accuracy
- Better consistency
Protect authenticity and integrity of data in a robust way

- Third Party Publication: Sign Once, Verify Many Ways
New things that you can do with audio
Let machines listen as we do

Automatic speech recognition

Business Logic

Text-to-speech

PSTN

Speech enhancement, Noise reduction

Multilingual voice mining, Speech & dialogue processing

Text categorization, Natural language processing

Semantic outputs
To build a mobile audio industry
... analogous to the graphics industry

- Synthesis-directed analysis of sounds
  - how would you model a lion’s roar?
- Algorithms for synthetic sound generation
- Tools for sound model automation and support
- Cross-platform audio synthesis engine with a small footprint and low compute requirements

<table>
<thead>
<tr>
<th>Sound Technologies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>---------------------</td>
</tr>
<tr>
<td>Recordings</td>
</tr>
<tr>
<td>Current Sound Models</td>
</tr>
<tr>
<td>New Sound Models</td>
</tr>
</tbody>
</table>
Advanced Audio Zip
Scaleable-to-Lossless Audio Codec

State-of-the-Art

<table>
<thead>
<tr>
<th>AAZ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Encode ONE bit-rate, decode ANY bit-rate</td>
</tr>
<tr>
<td>Dead zones in quality due to unstable QoS resulting in mismatch bit-rates</td>
</tr>
</tbody>
</table>

AAZ has been adopted by ISO as an MPEG4 international standard
New things that you can do with video
Make home videos more fun:
www.muvee.com

- Select video frames
- Cut to music
- Decide on transitions

Turn home video into high quality MTV automatically
Prevent drowning, save lives

- Drowning Early Warning System
- tracks people in dynamic aquatic conditions
- intelligently detect water crises situations
Watch video anywhere, any time, on any device!

And improve quality at the same time!
How is it done?
What’s needed to get it done?

- **Efficient implementation of MPEG-4 coding algo**
  - Fast hexagonal-search motion estimation algo
  - Fast block matching using partial distortion comparison
  - Fast DCT/Quantization algo
  - Adv-detection of zero-DCT-coefficients (ADZDC technique)
- **From 1 fps to 20 fps**

- **Adaptive rate/error control algorithms**
  - Adaptive to available channel bandwidth
  - Adaptive to available computational resources
  - Adaptive to bit error rate
More intelligent CCTV, improve homeland security

Scene/event analysis:
- Loitering
- Abandoning object
- Potential theft
- People tracking

Behavior Learning:
- Crowd Behaviour Analysis
  - Fighting
  - Rioting
  - Unusual crowding

Real-time notification
Prioritised viewing
Event indexing

Bayesian Event Mapping

Indoor & Outdoor CCTV
Improve sophistication of our media industry

Existing pains

- "Intrusive" ads
  - pops out during play!

Tennis TV
- manual

Tagged to camera h/w
- costly
- done once at source

Non-intrusive virtual contents insertion

- Non-intrusive insertions
  - detects non-play segment
  - non-interfering insertion

Enhanced tennis TV
- auto-tracking
- super-resolution

Software detection:
- performed any-time
- demographic Ads
- cheap
Super Resolution: How is it done?

• Robust Scene Modeling and Camera calibration
  – Given a 2D court model of 3D scene that camera is capturing, identify 3D object positions robustly and accurately

• Super-Resolution Image Reconstruction from Video
  – Given a low resolution image sequence of an object far away from the camera, reconstruct a larger resolution image sequence
  – This is essentially an ill-posed problem, but we can apply domain info such as motion, pose, etc, to seek a good solution

• Robust Object and Landmark Detection
  – Real-time
  – Geometric invariance

• Deployment and Application Constrains
  – Real-time
Highlight Detection: How is it done?

- **Audio**
  - Detect raised pitch in commentator

- **Visual**
  - Detect goal-mouth appearances

- **Results**
  - 5-15% of original video length
  - 100% goal detected

- **Application**
  - Set-top-box / digital video recorder (client-side)
  - News summary (broadcast server-side)
Virtual Content Insertion: How is it done?

1. Sport Events
2. Live Broadcast / Repeat Telecast
3. Receive / Decode / Process

Determine
(1) WHEN is a good time to insert content
(2) WHERE is a Good place to insert content
Acknowledgements

All the innovations shown in this talk were made by researchers at

Institute for Infocomm Research
A*STAR