

# CONTENT-BASED IMAGE RETRIEVAL IN AUGMENTED REALITY

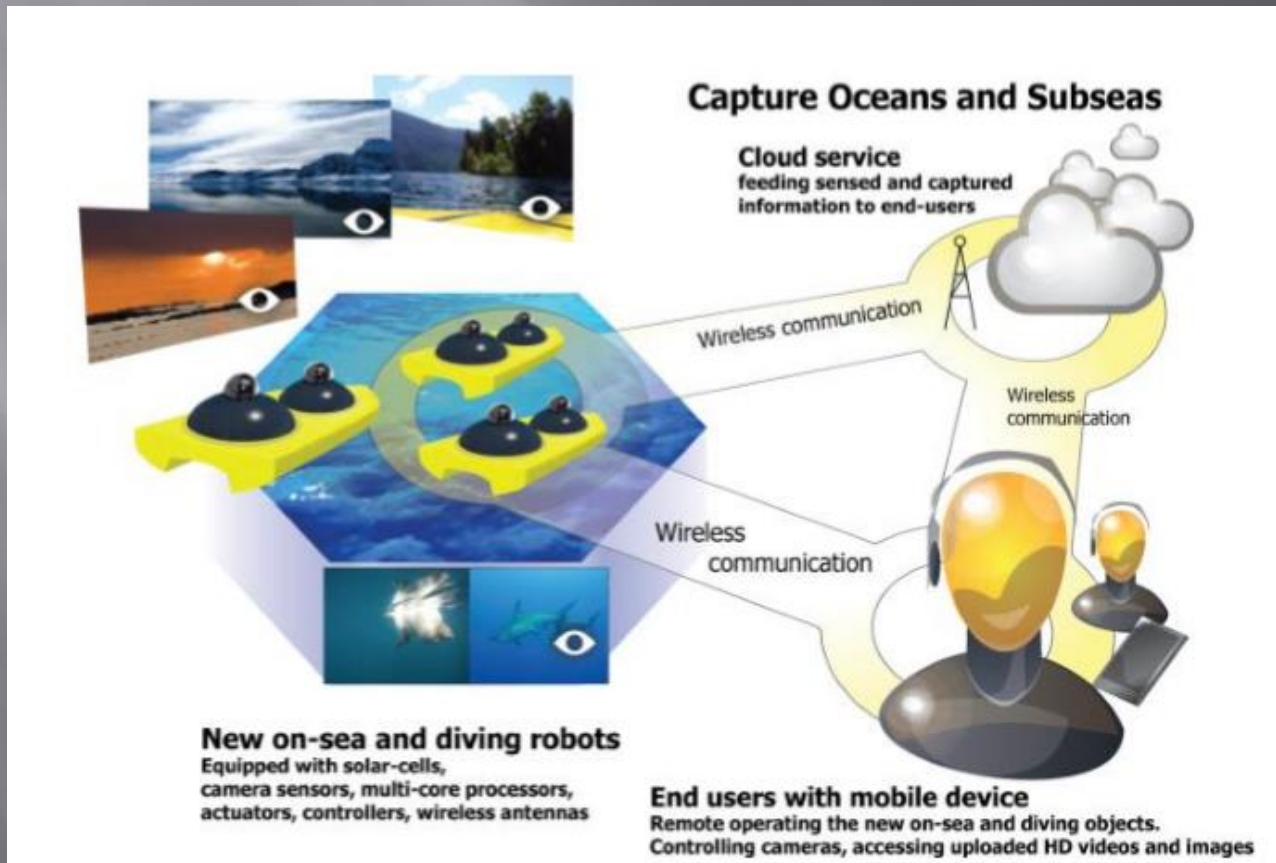
Leszek Kaliciak, Hans Myrhaug, Ayse Goker  
Ambiesense Ltd, Scotland

# Content

- ▣ Ocean monitoring robot
- ▣ Image retrieval
  - Textual features
  - Visual features
  - Similarity measurement
  - Fusion of feature spaces
- ▣ Developed hybrid models
  - Tensor based to capture correlation
  - Adaptivity
- ▣ Augmented Reality User Interface

# Ocean monitoring use-case: Horizon 2020: Cerbero

- ▣ New type of marine robots with surface and underwater surveillance capabilities



# Specifications

- ▣ Smart video-sensing unmanned vehicles with immersive environmental monitoring capabilities
- ▣ Can capture live videos and images of the local on-sea and subsea surroundings
- ▣ Can be remote controlled within wireless reach and visible sight
- ▣ Also capable of self-operation and navigation

# Specifications

- ▣ Robots can perform on-the-fly data analysis and fusion in order to make decisions (e.g. manoeuvre) and adapt to changing environment
- ▣ Sensed data can be stored locally or streamed to a cloud service from where relevant information can be retrieved
- ▣ 100% battery driven, solar and wind charged



# Some Existing ROVs, AUVs



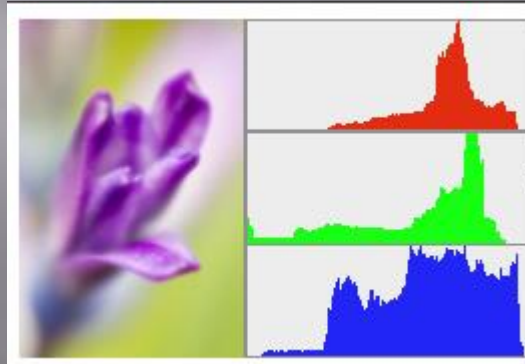
# Image Retrieval

- ▣ Usually based on Vector Space Model
- ▣ Visual content and image tags represented as vectors
- ▣ Query represented as vector
- ▣ Angle or distance between vectors -> similarity (one feature space)
- ▣ Top ranked images presented to user (based on similarity scores)

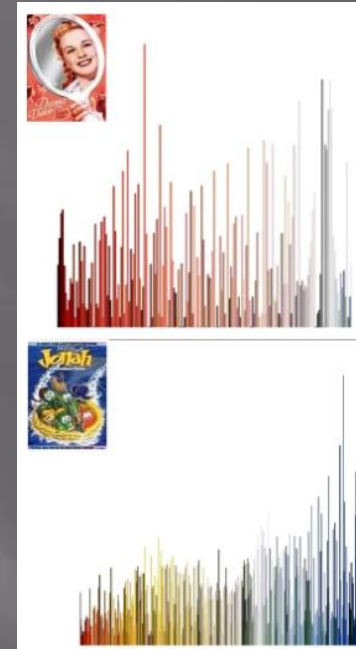
$$sim(a, b) = \frac{\langle a | b \rangle}{\|a\| \cdot \|b\|}$$

$$sim(a, b) = \sqrt{\sum_i (a_i - b_i)^2}$$

# Global Visual Features (low-level)



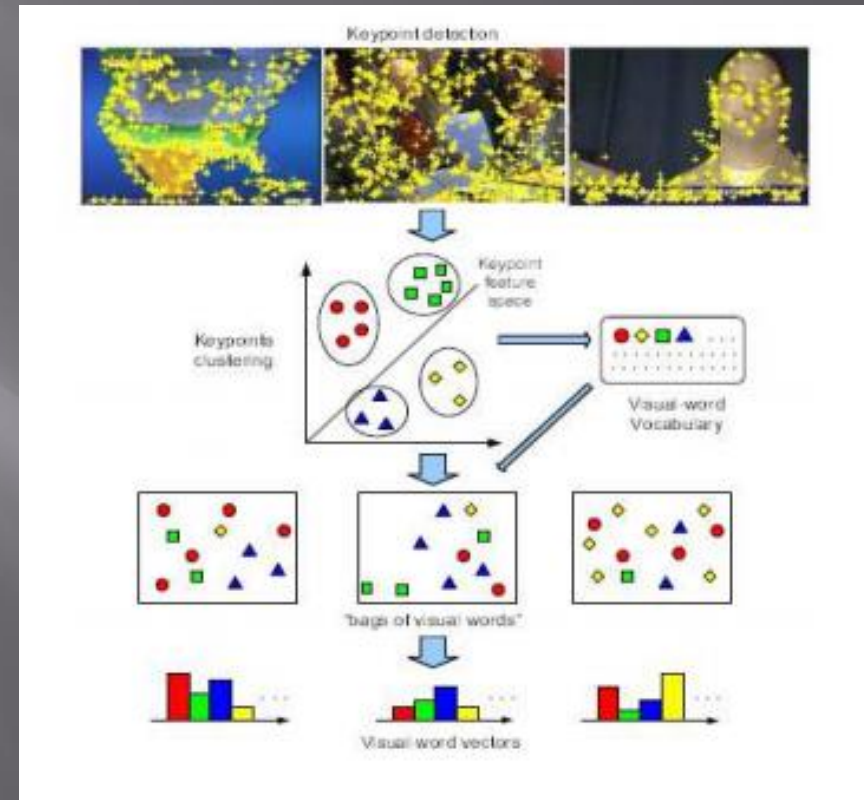
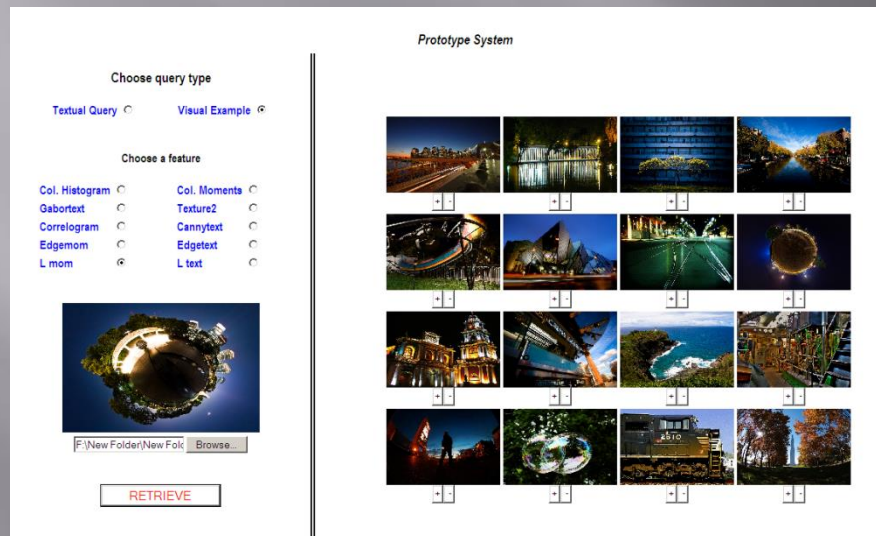
$$C_{\Delta x, \Delta y}(i, j) = \sum_{p=1}^n \sum_{q=1}^m \begin{cases} 1, & \text{if } I(p, q) = i \text{ and } I(p + \Delta x, q + \Delta y) = j \\ 0, & \text{otherwise} \end{cases}$$





# Visual Features (mid-level)

## □ Bag of Visual Words



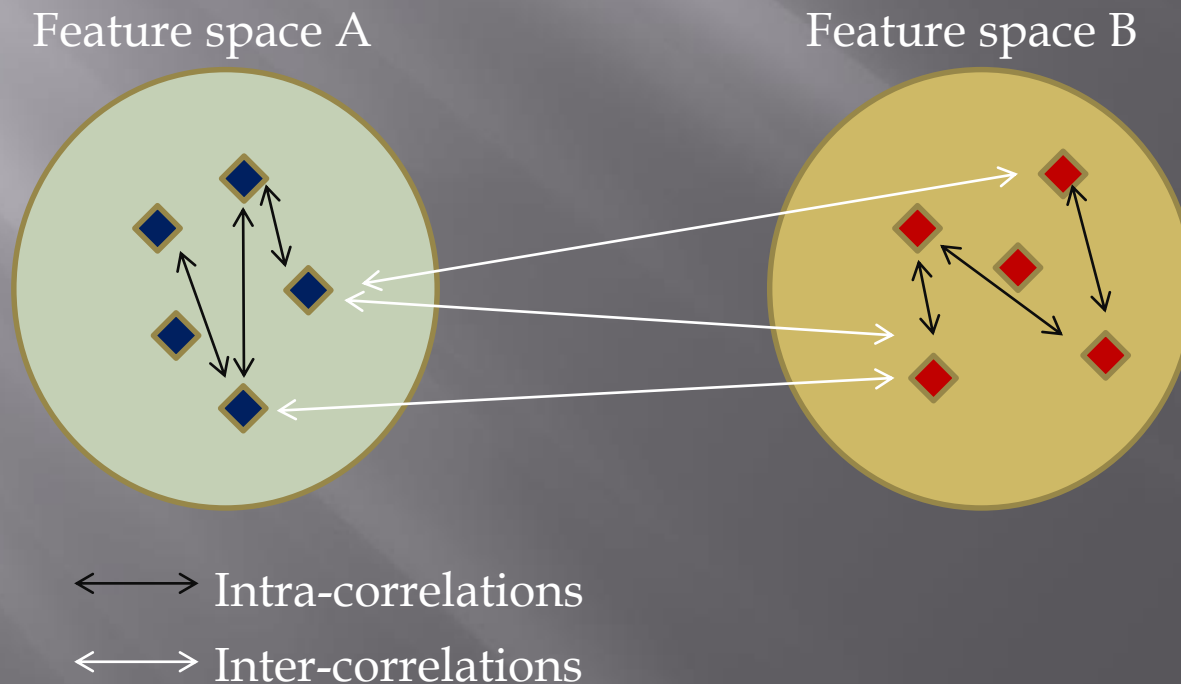
- (+) some ability to recognize objects
- (-) visual words have no semantic meaning

# Visual Features (high level)

- ▣ Grouping of visual words
- ▣ Segmentation-based
  
- ▣ (+) closest to human perception
- ▣ (-) not yet scalable to large data collections and generic image retrieval

# Our Hybrid Models

- ▣ Fusion of feature spaces improves the retrieval results
- ▣ We use tensors to fuse the feature spaces



# Adaptivity

- ▣ We measure the strength of the relationship between query and its context
- ▣ Weak relationship - context becomes important. We adjust the probability of the original query terms; the adjustment will significantly modify the original query
- ▣ Strong relationship - context will not help much. The original query terms will tend to dominate the whole term distribution in the modified model. The adjustment will not significantly modify the original query

# User Interface in VR

## Ocean monitoring + potentially Martian Rover use-case





Thank you