Visual Semantic Complex Network for Web Images

Shi Qiu[1], Xiaogang Wang[2], and XiaoouTang[1]

[1]Department of Information Engineering, [2]Department of Electronic Engineering, The Chinese University of Hong Kong

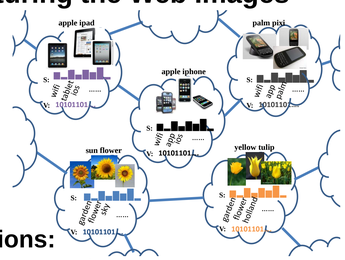
1. Main Idea

How to model the relevance of web images?

- Textual information + keyword index
- Visual information + ANN algorithms
- Only effective for images within a small local region in textual/visual feature space

Proposed Idea: Structuring the Web Images

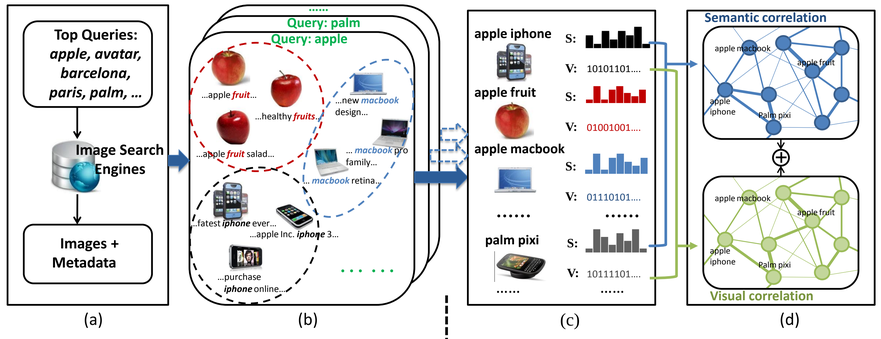
Build a **V**isual **S**emantic **C**omplex **N**etwork:



1) Semantic concepts: compact image clusters (elementary units)

2) Inter-concept correlations: connect relevant semantic concepts

2. VSCN Construction



Semantic Concepts Discovery

INPUT: 2k top keywords from Image SE
OUTPUT: 33,240 semantic concepts
+ 10M exemplar images

Inter-concept Correlations

Textual dscp. → Semantic correlation
Visual dscp. → Visual correlation
K-nearest-neighbor network

4. CBIR with the VSCN

Idea: relevant images are connected through VSCN

Methods

(b) Initial ranking list $\{(I_k, d_k)\}_{k=1}^{N_I}$

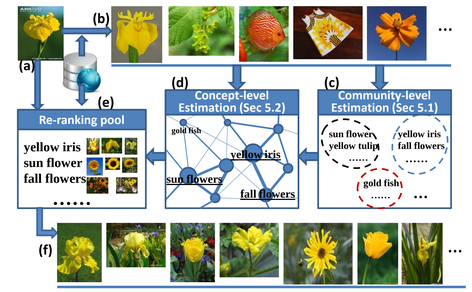
(c) Community Estimation

$$s(T_i) = \sum_{k=1}^{N_I} exp(\frac{-d_k}{\sigma}) \cdot \chi[c(I_k), T_i]$$

(d) Concept Estimation via RW

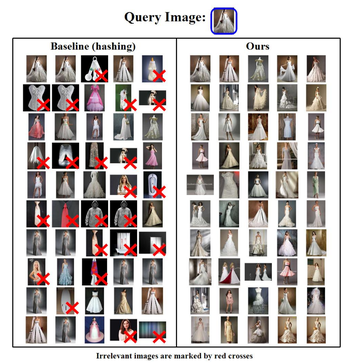
$$s(c_i') = \sum_{k=1}^{N_I} exp(\frac{-d_k}{\sigma}) \cdot \mathbf{1}[c(I_k) = c_i']$$

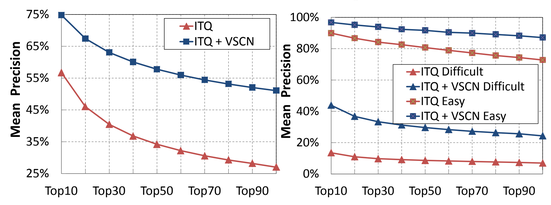
$$p^{n+1} = \alpha \mathbf{P}^T p^n + (1-\alpha)\pi, \quad \pi(i) = s(c_i') / \sum_i s(c_i')$$



Experimental Results

- ITQ-hashing [3] as the baseline
- 10K query images collected from Google
- Two subsets with *Difficult* & *Easy* queries
- Retrieval results labeled by annotators





[3] Y. Gong and S. Lazebnik. Iterative quantization: A procrustean approach to learning binary codes. In CVPR, 2011.

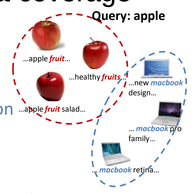
Semantic Concepts Discovery

- Query expansion with visually coherent keywords
- Less ambiguity, more consistency & coverage

Algorithm 1 Concept Discovery through Query Expansion

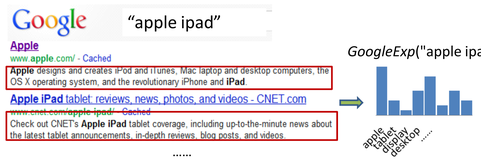
Input: Query q, image collection $\mathcal{I}_q$, surrounding texts $\mathcal{T}_q$.
Output: Learned concept set $\mathcal{C}_q = \{c_i\}_{i=1}^{M_q}$.

1: **Initialization:** $\mathcal{C}_q := \emptyset$, $r_I(w) := 0$.
2: **for all** images $I_k \in \mathcal{I}_q$ **do**
3: Find the top K visual neighbors, denote as $\mathcal{N}(I_k)$ — V-coherent region
4: Let $W(I_k) = \{w_{I_k}^i\}_{i=1}^{T}$ be the T most frequent words in the surrounding texts of $\mathcal{N}(I_k)$.
5: **for all** words $w_{I_k}^i \in W(I_k)$ **do**
6: $r_I(w_{I_k}^i) := r_I(w_{I_k}^i) + (T - i)$. — Update V-coherent scores for textual words
7: **end for**
8: **end for**
9: Combine q and the M_q words with largest $r_I(w)$ to form $\mathcal{C}_q$.



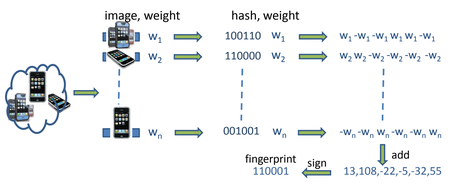
Semantic Correlation

- Google Kernel [1]
- Determine correlations from the rich web context



Visual Correlation

- Measured by the visual similarity between the exemplar image sets
- Sim-hashing [2]: generate binary signatures for image sets

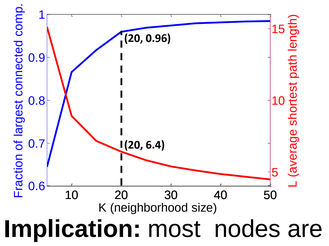


[1] M. Sahami and T. D. Heilman. A web-based kernel function for measuring the similarity of short text snippets. In Proc. WWW. ACM, 2006.
[2] G. Manku , A. Jain, and A. Das Sarma. Detecting near-duplicates for web crawling. In Proc. WWW. ACM, 2007.

3. Exploring VSCN Structures

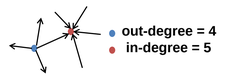
Connectivity

$$L = \frac{1}{|V|(|V|-1)} \sum_{v_i, v_j \in V, v_i \neq v_j} d(v_i, v_j)$$

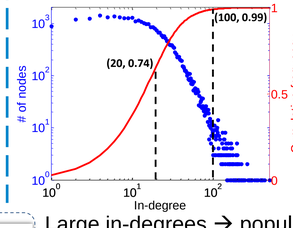


Implication: most nodes are reachable within a few hops.

In-degree Distribution

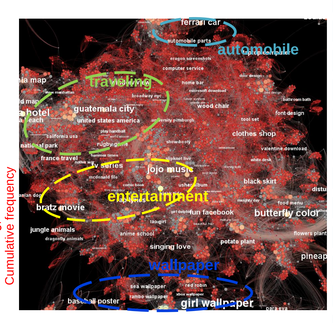


In-degree reflects the density around a node



Large in-degrees → popular / representative concepts
Zero in-degrees → uncommon / outlier concepts

Concept Community



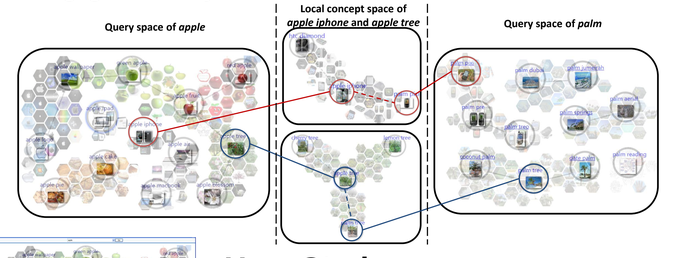
Implication: Concept communities aggregate relevant images, which is beneficial for image retrieval.

More details, demos, and data can be found at *http://mmlab.ie.cuhk.edu.hk/project_VSCN.html*

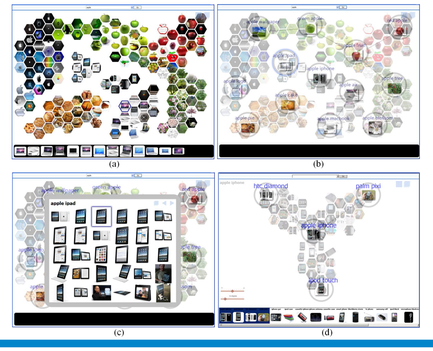
5. Image Browsing with the VSCN

Idea: image browsing guided by the VSCN

A novel browsing scheme that bridges different local spaces



Interface



User Study

- Interactive navigational image search (finding target image in mind)
- Results