Image Retrieval using Salient Regions with Vector-Spaces and Latent Semantics

Jonathon S. Hare & Paul H. Lewis
{jsh02r, phl}@ecs.soton.ac.uk
Intelligence, Agents, Multimedia Group, School of Electronics and Computer Science, University of Southampton, Southampton, UK

MOTIVATION

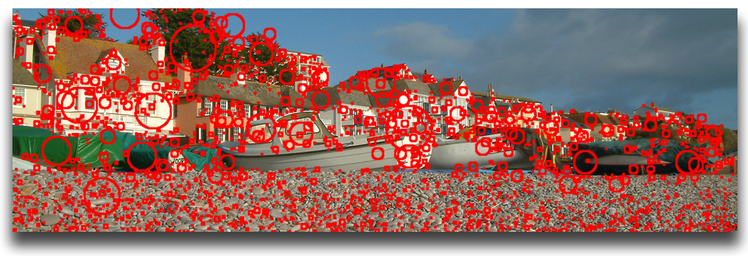
THIS WORK HAS BEEN MOTIVATED BY OUR PREVIOUS WORK ON CBIR USING SALIENT REGIONS WITHIN A VECTOR-SPACE MODEL.

OUR PREVIOUS WORK HAS INVESTIGATED TWO AREAS OF IMAGE RETRIEVAL;
- PERFORMANCE OF SALIENT REGION DETECTORS & HOW RETRIEVAL BASED ON SALIENT REGIONS CAN OUTPERFORM OTHER TECHNIQUES, ESPECIALLY IN THE PRESENCE OF IMAGE TRANSFORMS AND ROTATIONS.
- QUERY BY IMAGE CONTENT WITH QUERY IMAGES OF DEGRADED QUALITY (I.E. QUERIES CAPTURED BY A CAMERA PHONE).
  - WE USED SALIENT REGIONS WITH QUANTISED LOCAL DESCRIPTORS IN A VECTOR-SPACE.
  - THE IMAGE ON THE RIGHT IS TYPICAL OF THE KIND OF QUERY WE DEALT WITH; OUT-OF-FOCUS, AND OVEREXPOSED.

LATENT SEMANTIC INDEXING (LSI) IS A LOGICAL EXTENSION TO USING A VECTOR-SPACE MODEL. THE AIM OF THIS WORK WAS TO INVESTIGATE WHETHER THE LSI MODEL BRINGS ANY BENEFITS OVER THE VECTOR-SPACE MODEL.

THIS POSTER GIVES AN OVERVIEW OF OUR APPROACH TO MODELLING IMAGE CONTENT WITH VECTOR-SPACES AND LATENT SEMANTIC ANALYSIS.

SALIENT REGIONS



PREVIOUS WORK HAS SHOWN THAT CBIR USING SALIENT INTEREST REGIONS OUTPERFORMS GLOBAL IMAGE DESCRIPTORS.
- SALIENT REGION APPROACHES AVOID THE PROBLEMS INHERENT WITH SEGMENTATION BASED RETRIEVAL APPROACHES.

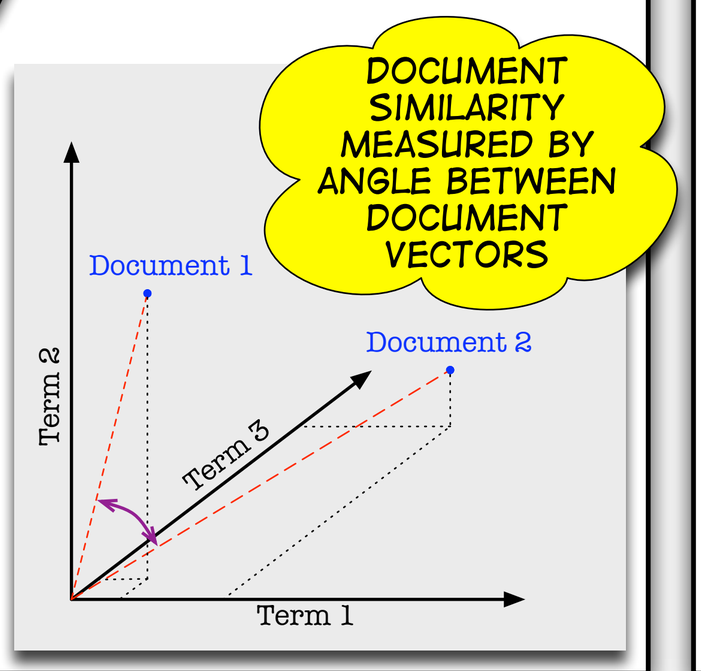
FOR OUR ALGORITHM, WE SELECT SALIENT REGIONS BY DETECTING SCALE-SPACE PEAKS IN A MULTI-SCALE DIFFERENCE-OF-GAUSSIAN PYRAMID.
- REGIONS SELECTED THIS WAY HAVE BEEN SHOWN TO BE VERY ROBUST.

INFORMATION RETRIEVAL TECHNIQUES

VECTOR-SPACE MODEL

THE VECTOR-SPACE MODEL IS CONCEPTUALLY SIMPLE;
- DOCUMENTS ARE PARSED INTO INDIVIDUAL TERMS.
  - TERMS THEN UNDERGO A PROCESS CALLED STEMMING.
  - EACH STEMMED TERM IS THEN REPRESENTED BY A UNIQUE IDENTIFIER FOR THAT TERM.
- THE NUMBER OF OCCURRENCES OF EACH TERM IN THE DOCUMENT IS COUNTED AND ARRANGED IN A VECTOR OF TERM-FREQUENCIES.
  -THE VECTOR OFTEN HAS A WEIGHTING APPLIED TO IT.

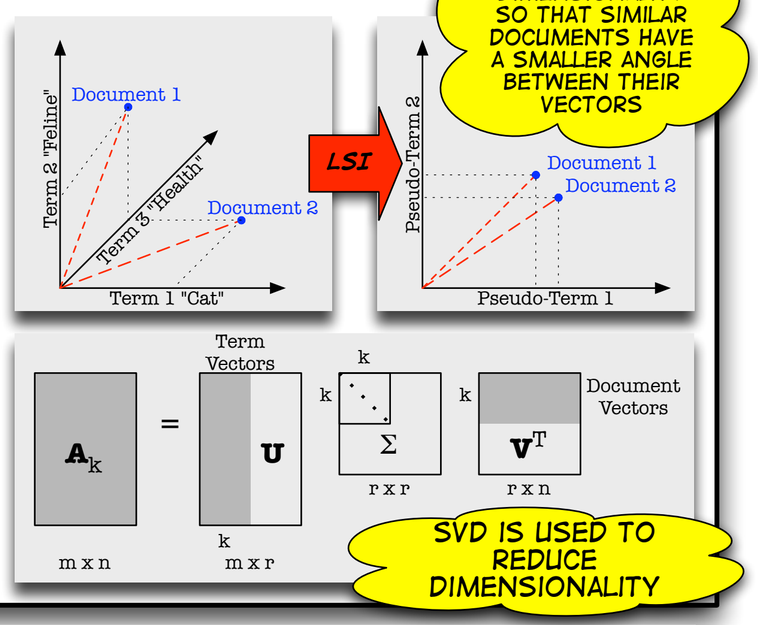
- DOCUMENTS CAN BE CONSIDERED TO BE SIMILAR IF THE ANGLE BETWEEN THEIR VECTORS IS SMALL.



THE VECTOR-SPACE TECHNIQUE DEPENDS ON A LEXICAL MATCH BETWEEN THE WORDS IN THE DOCUMENTS FOR THEM TO BE CONSIDERED SIMILAR.
- THERE IS OFTEN DIVERSITY IN THE WORDS USED TO DESCRIBE A DOCUMENT.
  - THIS MAKES THE LEXICAL METHODS INCOMPLETE AND IMPRECISE.

THERE IS AN IMPLICIT HIGHER-ORDER STRUCTURE IN THE ASSOCIATION OF TERMS WITH DOCUMENTS.
- THIS STRUCTURE CAN BE EXPLOITED BY USING THE K LARGEST SINGULAR VALUES FROM A SINGULAR VALUE DECOMPOSITION (SVD) OF THE TERM BY DOCUMENT MATRIX.

LATENT SEMANTIC INDEXING



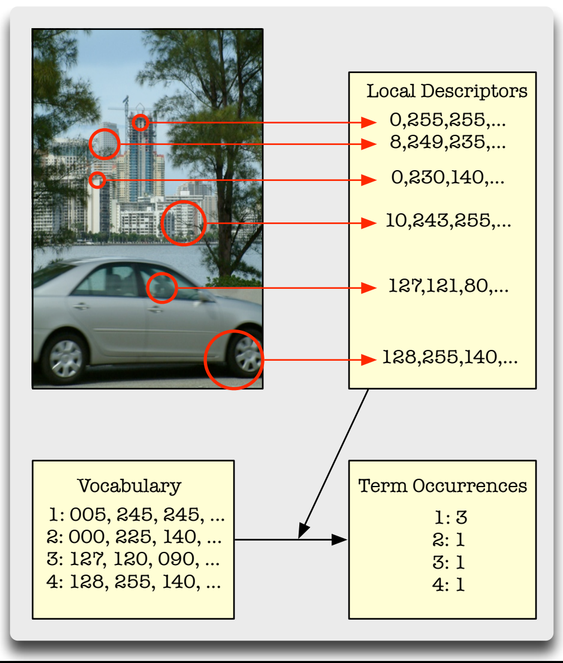
MODELLING IMAGES AS WORDS

IMAGES ARE MODELLED AS VISUAL TERMS USING LOCAL DESCRIPTORS FROM THEIR SALIENT REGION REPRESENTATIONS.
- SIFT (SCALE INVARIANT FEATURE TRANSFORM) DESCRIPTOR IS USED BECAUSE IT HAS BEEN SHOWN TO BE VERY ROBUST.

EACH LOCAL DESCRIPTOR IS VECTOR-QUANTISED INTO A VISUAL TERM. THE VECTOR-QUANTISER WAS TRAINED BY CLUSTERING LOCAL FEATURE DESCRIPTORS USING K-MEANS.

THE VISUAL TERMS ARE COUNTED AND ASSEMBLED INTO TERM VECTORS.
- THE VECTORS ARE THEN WEIGHTED.

THE VECTORS CAN BE USED IMMEDIATELY FOR VECTOR-SPACE QUERYING, OR LSI CAN BE APPLIED.



MEASURING RETRIEVAL PERFORMANCE

ASSUME A TEST SET OF IMAGES WITH LABELS (SUCH AS "TREES", "BUSHES", "CLEAR SKY", ETC.).
- REASONABLE TO EXPECT THAT THE IMAGES RETURNED BY THE RETRIEVAL SYSTEM SHOULD HAVE THE SAME LABELS AS THE QUERY IMAGE.

THIS IS MEASURED THROUGH 'SEMANTIC RELEVANCE':

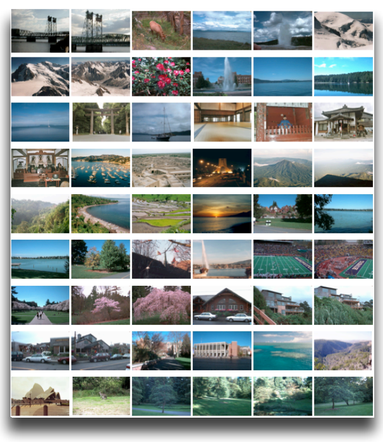
$$R_{semantic} = \frac{|A \cap B|}{|A|}$$

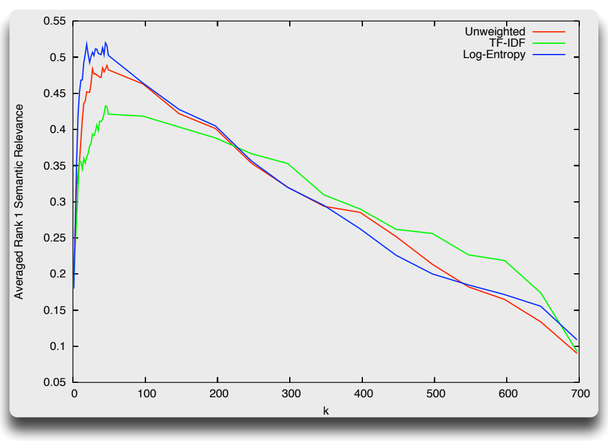
- ***A*** AND ***B*** REPRESENT THE SETS OF LABELS IN THE QUERY AND RESULT IMAGE RESPECTIVELY.

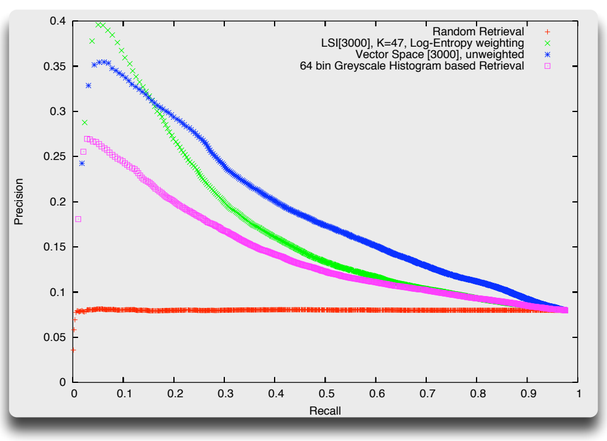
PRECISION AND RECALL CAN BE CALCULATED BY QUANTISING THE 'SEMANTIC RELEVANCE' ABOVE AND BELOW A THRESHOLD, ***Z***, TO GET THE TRUE RELEVANCE, $V_{N,Z}$:

$$V_{n,Z} = \begin{cases} 0 \text{ if } R_{semantic} < Z \\ 1 \text{ otherwise} \end{cases}$$

EXPERIMENTAL RESULTS







ALL 697 SEMANTICALLY MARKED IMAGES FROM THE WASHINGTON DATASET WERE USED.
- EACH IMAGE INDEXED USING THE TWO ALGORITHMS WITH TF-IDF AND LOG-ENTROPY WEIGHTING SCHEMES.
THE FIRST GRAPH ABOVE, SHOWS THE OPTIMAL K-VALUE FOR LSI. THE SECOND SUMMARISES THE BEST PERFORMING WEIGHTED VERSIONS OF THE TWO MODELS IN THE FORM OF PRECISION-RECALL PLOTS ($V_{N,Z}$ = 0.5).

FUTURE WORK

FUTURE DIRECTIONS CONSIST OF A NUMBER OF TASKS:

THE LABELS DESCRIBING EACH OF THE IMAGES NEED OVERHAULING BY APPLYING A WELL-DEFINED VOCABULARY.
- THIS SHOULD GIVE A BETTER INDICATION OF HOW SEMANTICALLY RELEVANT ONE IMAGE IS TO ANOTHER.

ISSUES OF COMPUTATIONAL COMPLEXITY NEED ADDRESSING.
- WHAT IS THE TRADE-OFF BETWEEN ON- AND OFF-LINE PROCESSING FOR THE TWO MODELS?

INVESTIGATING THE EFFECT OF USING STOP-WORDS ON THE PERFORMANCE OF THE VECTOR-SPACE TECHNIQUE.

INVESTIGATING THE USE OF OTHER LOCAL FEATURE DESCRIPTORS, AND THE COMBINATION OF DIFFERENT DESCRIPTORS.

I A M
Electronics and Computer Science
University of Southampton
MOTOROLA LABS