

Burstiness in text

- ✓ Burstiness is a well known phenomenon in text retrieval
- ✓ Example of burstiness: the term "vision" in Wikipedia's article on "Computer Vision":

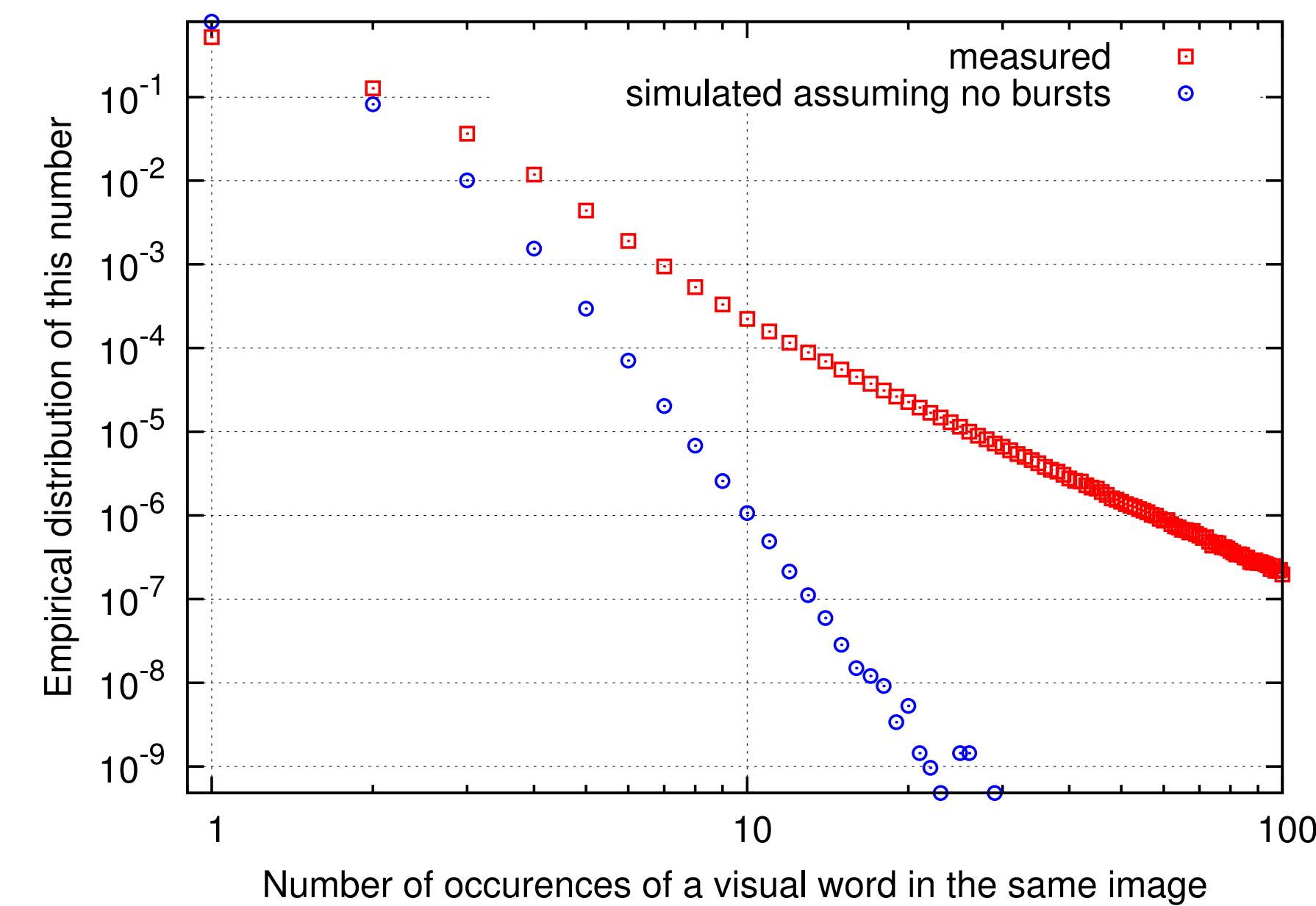
Computer **vision** is the science and technology of machines that see. As a scientific discipline, computer **vision** is concerned with the theory for building artificial systems that obtain information from images. The image data can take many forms, such as a video sequence, views from multiple cameras, or multi-dimensional data from a medical scanner. As a technological discipline, computer **vision** seeks to apply the theories and models of computer **vision** to the construction of computer **vision** systems. Examples of applications of computer **vision** systems include systems for:

- * Controlling processes (e.g. an industrial robot or an autonomous vehicle).
- * Detecting events (e.g. for visual surveillance or people counting).
- * Organizing information (e.g. for indexing databases of images and image sequences).
- * Modeling objects or environments (e.g. industrial inspection, medical image analysis or topographical modeling).
- Interaction (e.g. as the input to a device for computer-human interaction). Computer **vision** can also be described as a complement (but not necessarily the opposite) of biological **vision**. In biological **vision**, the visual perception of humans and various animals are studied, resulting in models of how these systems operate in terms of physiological processes. Computer **vision**, on the other hand, studies and describes artificial **vision** systems that are implemented in software and/or hardware. Interdisciplinary exchange between biological and computer **vision** has proven increasingly fruitful for both fields. Sub-domains of computer **vision** include scene reconstruction, event detection, tracking, object recognition, learning, indexing, motion estimation, and image restoration.

- ✓ This 236-word document contains 13 occurrences of "vision"
- ✓ The term "vision" is the 1897th most common word, its marginal probability is $4.35e-5$ (source: Gutenberg project)
⇒ the probability of observing this Wikipedia page, assuming independence of word generation, is $1.6e-36!$
- ✓ Burstiness distorts the similarity measure between documents

Evaluating the burstiness in images

- ✓ compute the marginal probabilities of visual words
- ✓ for 1 million images, count the occurrences of a word that appears at least once in the image



- ✓ The burstiness phenomenon is statistically very significant
- ✓ Continuous feature space: handling visual word bursts like bursts in text (e.g. with TF-IDF weights) is suboptimal, as this does not exploit the distances between descriptors.

Core search system

- Starting point: **Video-Google** [Sivic and Zisserman 03]:
- ✓ description: Hessian-Affine [Mikolajczyk 05] + SIFT [Lowe 04]
- ✓ bag-of-features representation of images: quantize descriptors into visual words using k-means
- ✓ weighting visual words using tf-idf
- ✓ indexing structure: inverted file (one cell per descriptor)

Hamming Embedding (HE): the quantization index associated with a visual word is refined using a binary signature. Points with the same visual word are assumed to match if the Hamming distance between their signatures is below a threshold.

- ✓ **Extension**: we refine this approach by weighting the point matching score according to the Hamming distance

Weak geometric consistency (WGC): use partial geometric information (angle and scale) for all the images

- ✓ **Extension**: scale weighting ⇒ higher weights for larger regions

Multiple assignment (MA): Descriptors are assigned to several visual words.

- ✓ performed on the query side only, to limit the memory usage

Handling Burstiness

- ✓ **types of burstiness**:
 - intra** = repeating visual word within an image
 - inter** = repeating word over images
- ✓ **handling**: downweight matching scores that involve the same query point

$$m(i, b, j) := m(i, b, j) \sqrt{\frac{m(i, b, j)}{\sum_{j'/q(b, j')=q(b, j)} m(i, b, j')}} \quad \text{intra:}$$

$$m(i, b, j) := m(i, b, j) \sqrt{\frac{m(i, b, j)}{\sum_{b'} \sum_{j'/q(b', j')=q(b, j)} m(i, b', j')}} \quad \text{inter:}$$

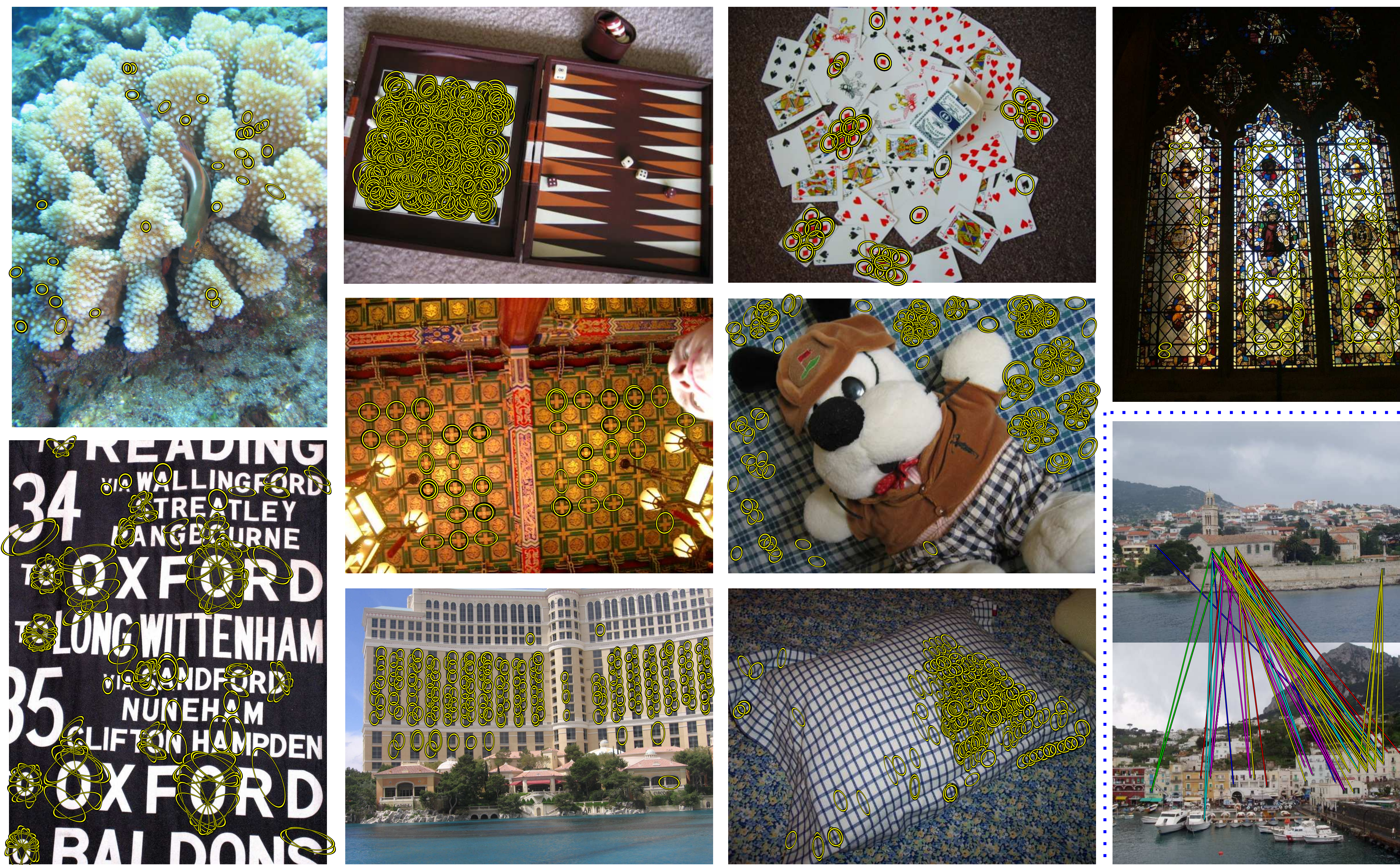
where:

$m(i, b, j)$ = matching score of query point i with point j from database image b

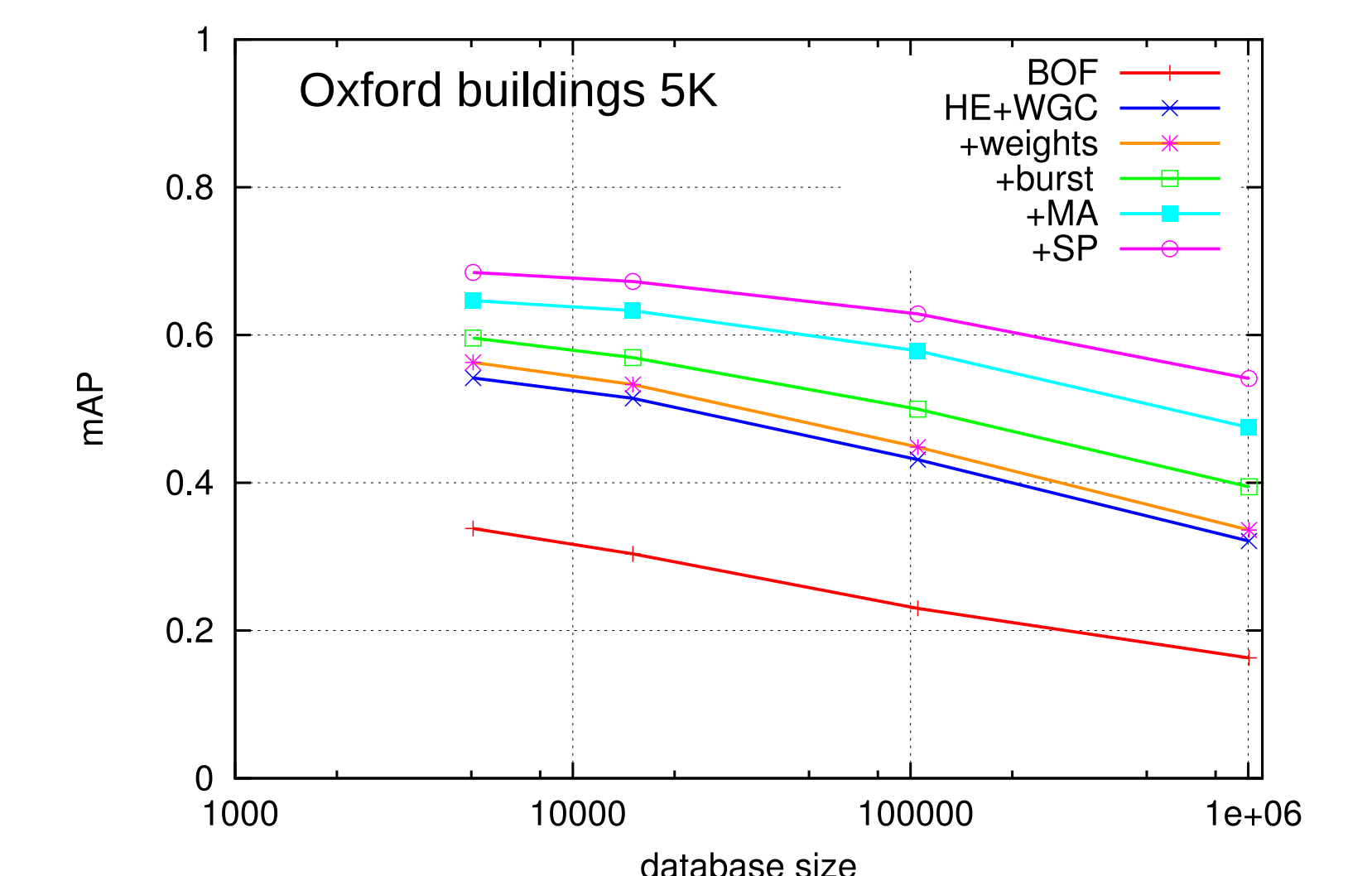
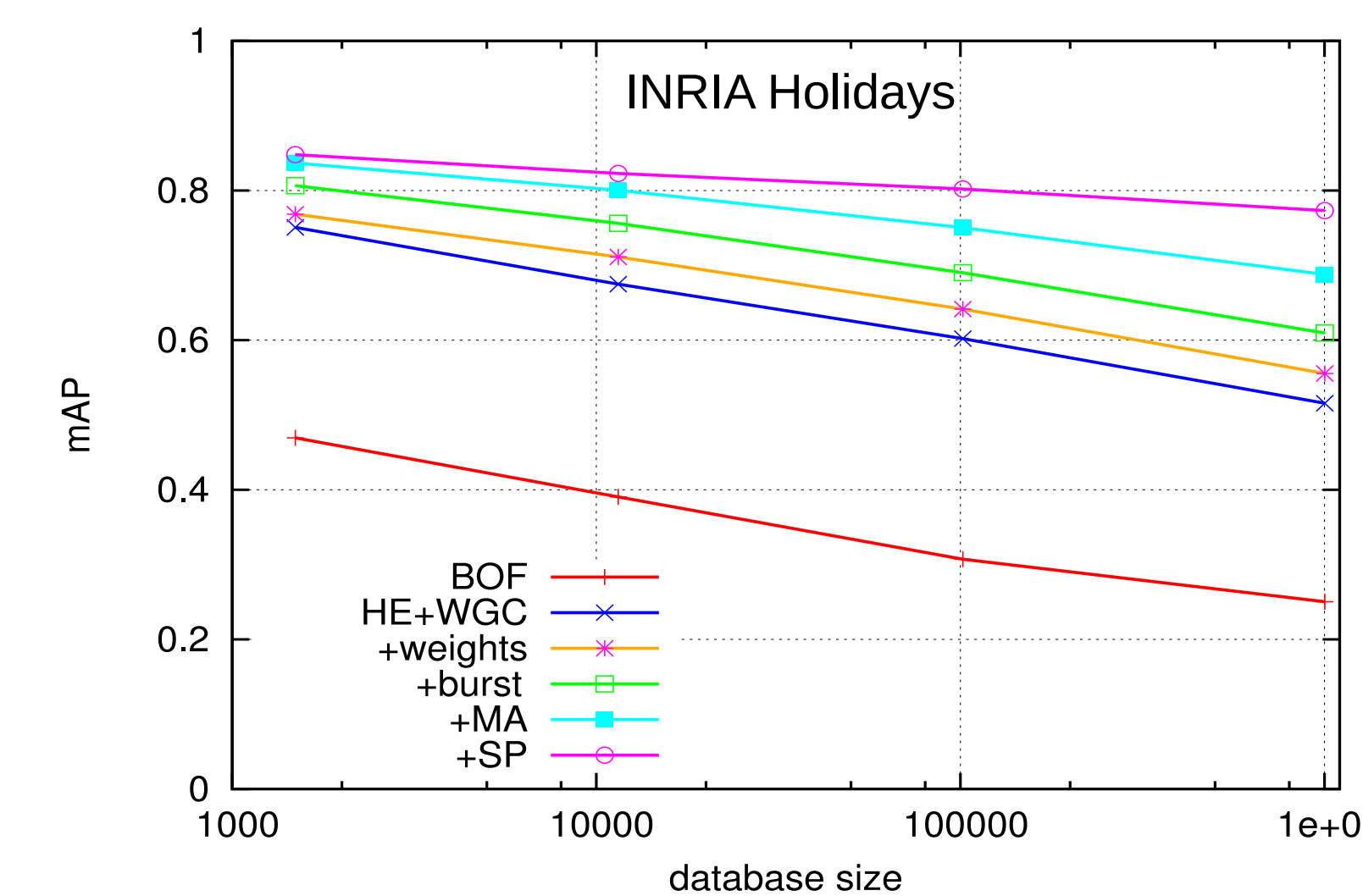
$q(b, j)$ = quantization index of point j from image b

- ✓ **results**: better than complete removal of multiple matches
- ✓ **implementation**: a post-processing of inverted file scan

Burstiness in images



Same parameters for all experiments: 20,000 visual words, 64-bit signature for HE, 10 neighbors for MA, etc



Partial solution: multiple match removal

- ✓ A point may be found to match with several points on the other image
- ✓ Burstiness: n similar occurrences of a visual word seen on two images count as n^2 matches! Images with repeating patterns score abnormally high
- ✓ Multiple match removal: keep at most one vote per query descriptor, the one with minimal Hamming distance

Comparison with the state of the art

	U. Kentucky benchmark	Oxford buildings 5K	INRIA Holidays
HE+WGC	3.36	0.542	0.751
HE+WGC	3.47	0.606	0.815
Multiple match removal	3.50	0.630	0.817
Burstiness	3.54	0.647	0.839
Burst.+spatial verification	3.64	0.685	0.848
Burst.+query expansion		0.747	
State of the art	3.68	0.718	0.751
Reported score (> is better)	CDM [Jégou et al., 2008] 4 × recall @ 4 [Nister & Stewenius 06]	Soft + query expansion [Philbin et al., 2008] mAP	HE + WGC [Jégou et al., 2008] mAP