On MPEG Work Towards a Standard for Visual Search

Yuriy A. Reznik

yreznik@ieee.org

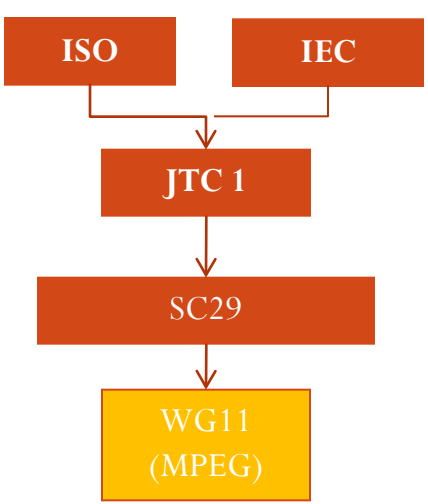
MPEG-7 CDVS, 8th FP7 Networked Media Concentration meeting, Brussels, December 13, 2011.

Outline

- Context
  - MPEG
  - Previous efforts: MPEG-7, MPQF, Visual Signatures
- Compact Descriptors for Visual Search
  - Applications, Scope and Objectives, Requirements, CfP
  - Standardization Timeline
- Evaluation framework
  - Databases
  - Evaluation criteria
- Responses & next steps

What is MPEG?

- Informal title:
  - Moving Picture Experts Group (MPEG)
- Formal title:
  - ISO/IEC JTC1 SC29 WG11 (Coding of Moving Pictures and Audio)
- Parent SDOs:
  - ISO: International Organization for Standardization
  - IEC: International Electrotechnical Commission
  - JTC 1: Joint Technical Committee One
  - SC29: Study Committee 29: Coding of Audio, Picture, Multimedia and Hypermedia Information
    - Members: National Bodies (25 voting, 15 observers)

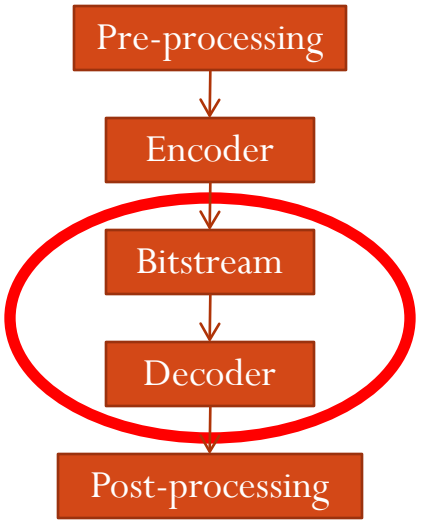


MPEG families of standards

- MPEG-1 (1992)
  - Video CD, MP3
- MPEG-2 | H.262 (1994)
  - DVDs, cable TV, etc.
- MPEG-4 (1998)
  - MP4, AAC
  - Part 10 (2003) – AVC aka H.264 (2 Emmies)
  - Blu-Ray discs, HDTV, mobile TV, Internet
- MPEG-7 (2002)
  - content description interface
- MPEG-21 (2001-5)
  - multimedia framework
- New standards:
  - MPEG-A (app. formats), B (systems), C (video), D (audio)…

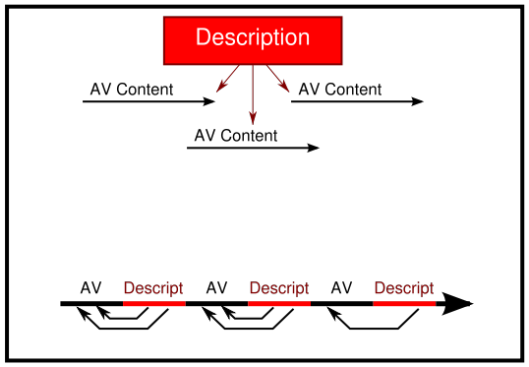
MPEG media coding standards

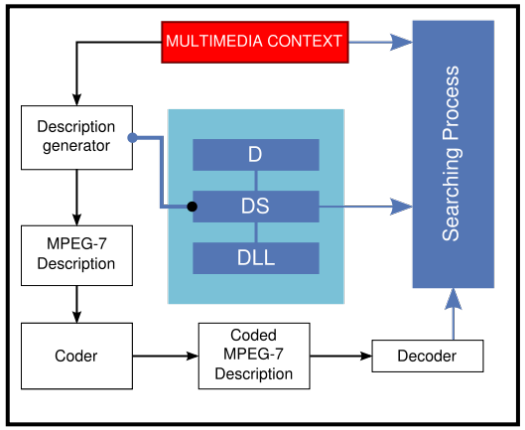
- Normative parts (MPEG-1,2,4):
  - **Bitstream syntax**
  - **Decoder**
  - Conformance (Part 4)
  - Reference software (Part 5)
- Not standardized:
  - Encoder
  - Particular (optimized) algorithms in decoder
  - Pre- or post-processing algorithms
- Why:
  - The smaller the standard the better
  - Promotes competition for best implementations in industry
  - Allows gradual improvements over time / increases longevity of the standard



MPEG media description standards

- MPEG-7:
  - Defines **description** of media content
    - Complementary to MPEG-1-4 standards
    - Collection of generic tools for multiple applications
  - Parts:
    - **Systems** (Part 1)
    - **Description definition language (DDL)** (Part2)
    - **Descriptors (D)** (Parts 3,4)
    - **Description Schemes (DS)** (Part 5)
    - Reference software (Part 6)
    - Conformance (Part 7)
    - Extraction and use of descriptors (Part8, TR)
    - **Query format** (Part 12)





MPEG-7 descriptors

- Examples of MPEG-7 descriptors (1998):

Type	Feature	Descriptors
Visual	Basic structure	Grid layout, Histogram
	Color	Color space, Dominant color, Color histogram, Color quantization
	Texture	Spatial intensity distribution, Homogeneous structure
	Shape	Object bounding box, Region-based shape, Contour-based shape, 3D shape
	Motion	Camera motion, Parametric object motion, Motion activity, Motion trajectory
Audio	Speech annotation	Lattice of words and phonemes
	Timbre	Ratio of even to odd harmonics
	Melody	Melodic contour and rhythm

- Implementation challenges:
  - High-level or abstract nature of some of the descriptors (e.g. parametric motion)
  - Extraction from media content may require human assistance

Visual Search

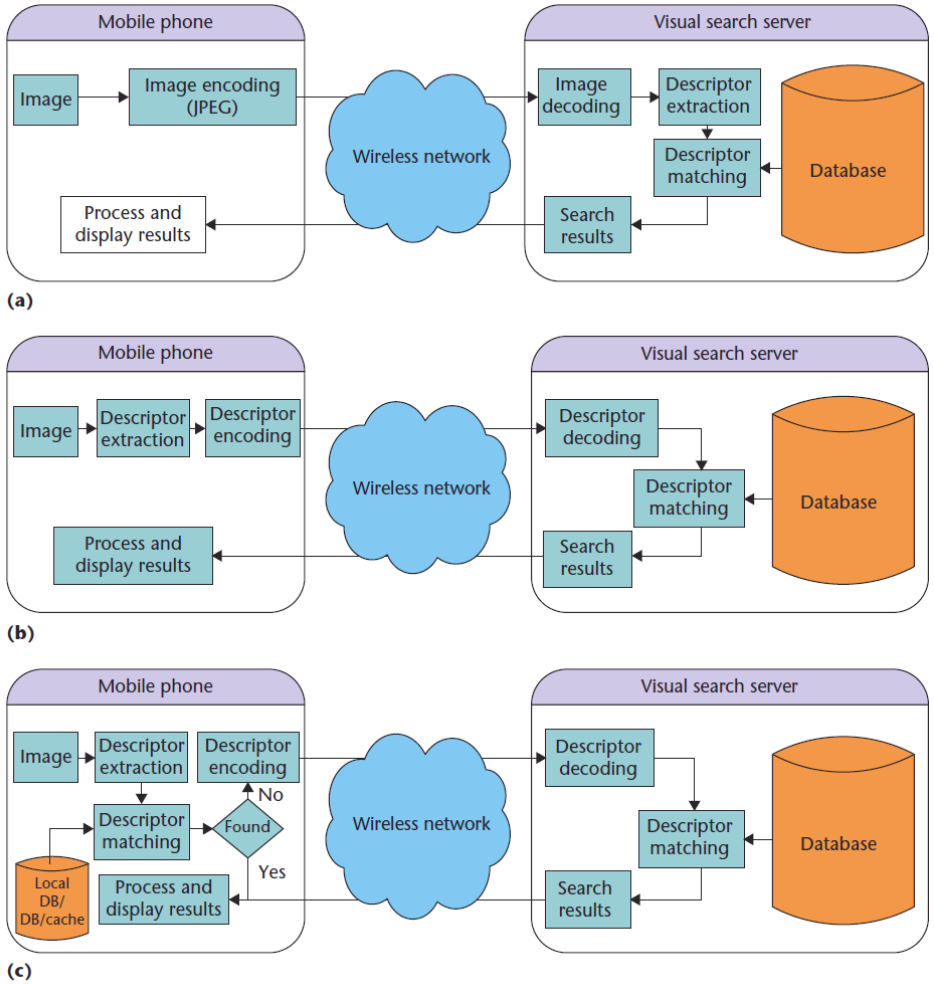
- Capability to initiate search queries using an image of an object of interest
  - Can be implemented in mobile phones
  - Mobile Visual Search
- Existing applications:
  - Google Goggles
  - Bing mobile app
  - Point & Find
  - Snaptell
  - Kooaba
- Essential functionality for Mobile Augmented Reality applications
  - Layar
  - more to come

Mobile Visual Search Architectures

- Several possible designs:
  a) Send image
  b) Send descriptors
  c) Perform local search first.

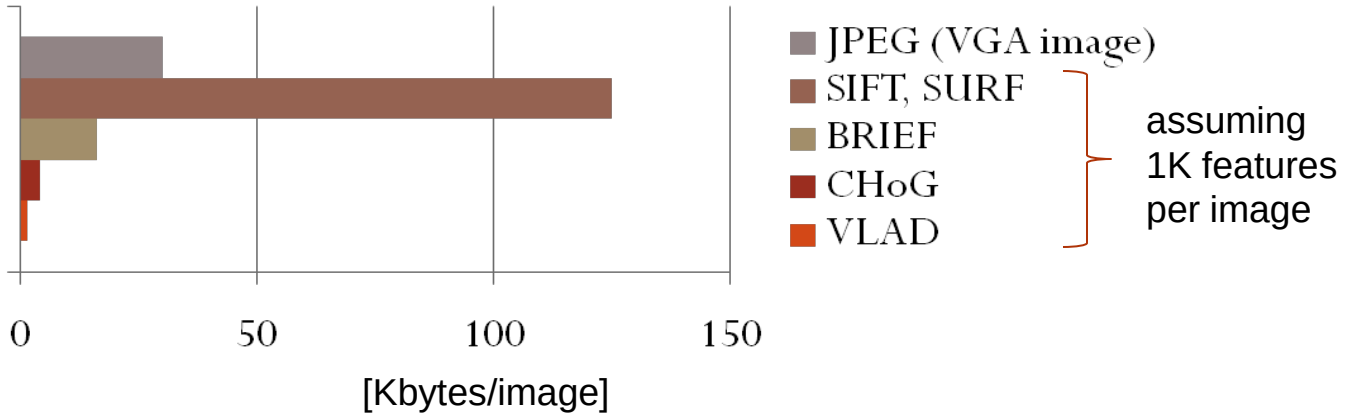
- To enable architectures b) and c) descriptors must be compact.

- Extraction must be fully automated!

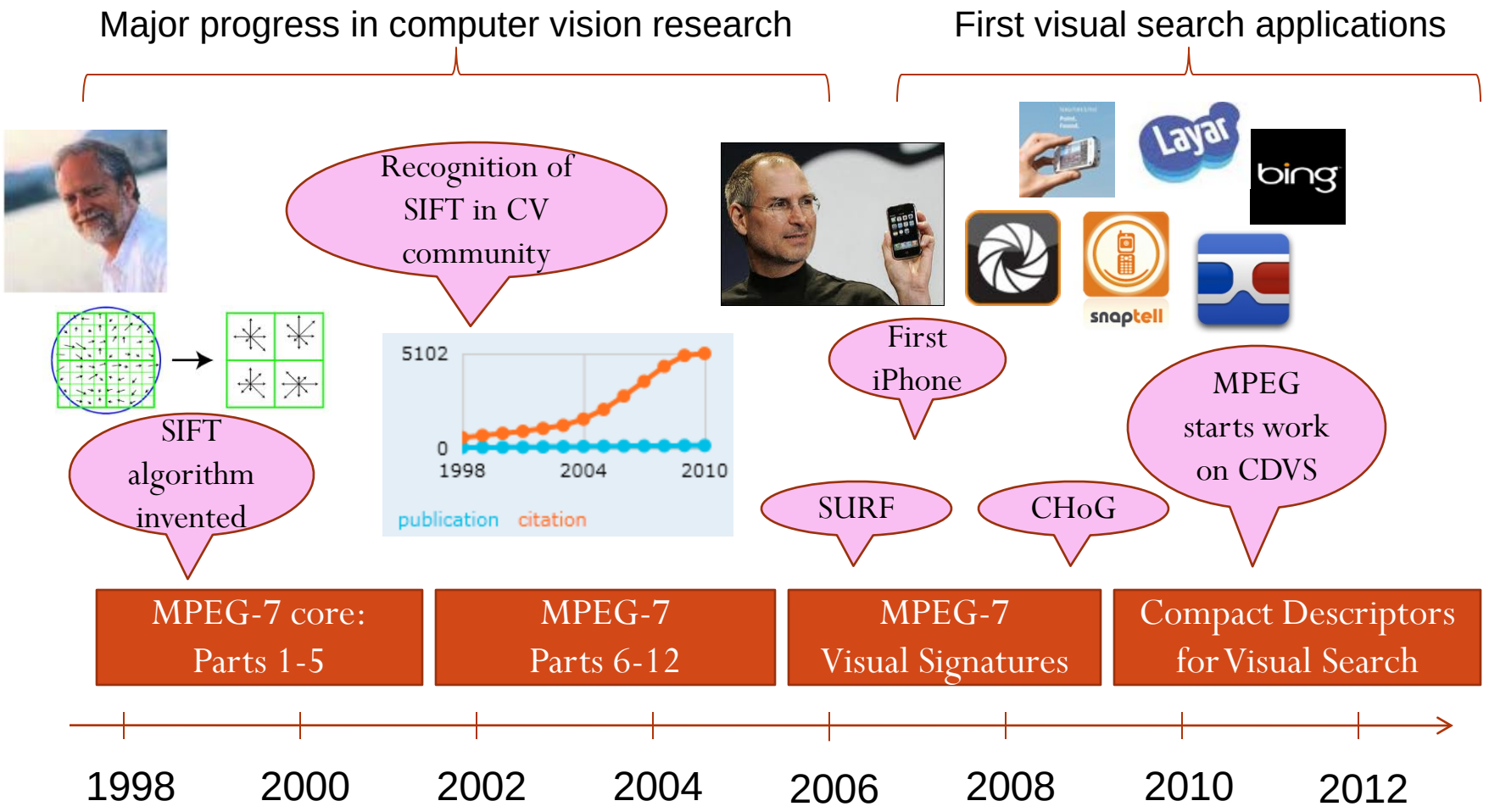


Suitable existing technologies

- SIFT and variants:
  - SIFT: D. Lowe, Distinctive image features from scale-invariant keypoints, IJCV'04.
  - SURF: H. Bay, *et.al.*, Speeded up robust features, ECCV'06
  - BRIEF: M. Calonder, *et.al.*, Binary Robust Independent Elementary Features, ECCV'10
  - CHoG: V. Chandrasekhar, *et.al.*, Compressed Histograms of Gradients: a low-bitrate feature descriptor, CVPR'09, CVPR'10, IJCV'11
  - VLAD: H. Jegou, *et.al.*, Aggregating local descriptors into a compact image representation, CVPR'10.
- Bitrates:



Timeline of events



MPEG work towards a standard

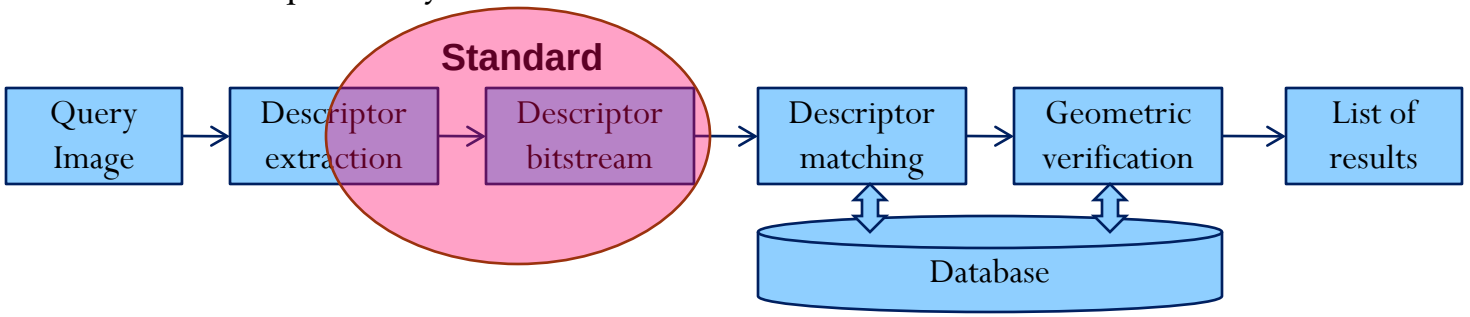
- Stanford's Workshop (Dec 2009)
- 91st MPEG Meeting (Jan 2010 Kyoto, JP)
  - First look at the topic
- 92nd MPEG Meeting (April 2010, Dresden, DE)
  - Context & Objectives
  - Title: "Compact Descriptors for Visual Search"
- 93rd MPEG Meeting (July 2010, Geneva,CH)
  - Workshop on Visual Search
  - Draft Requirements
- 94th MPEG Meeting (Oct 2010, Guangzhou)
  - Draft CfP, Draft Evaluation Framework
- 95th MPEG Meeting (Jan 2011, Daegu, KR)
  - Published CfP

WORKSHOP ON MOBILE VISUAL SEARCH

Stanford University
Frances C. Arrillaga Alumni Center
Thursday, December 3rd, 2009. [Group photo]

Scope and Objectives

- **Objectives**
  - Define a standard that:
    - enables design of visual search applications
    - minimizes lengths of query requests
    - ensures high matching performance (in terms of reliability and complexity)
    - enables interoperability between search applications and visual databases
    - enables efficient implementation of visual search functionality on mobile devices
- **Scope**
  - It is envisioned that (as a minimum) the standard will specify:
    - **bitstream of descriptors**
    - **parts of descriptor extraction process** (e.g. key-point detection) needed to ensure interoperability.



Requirements

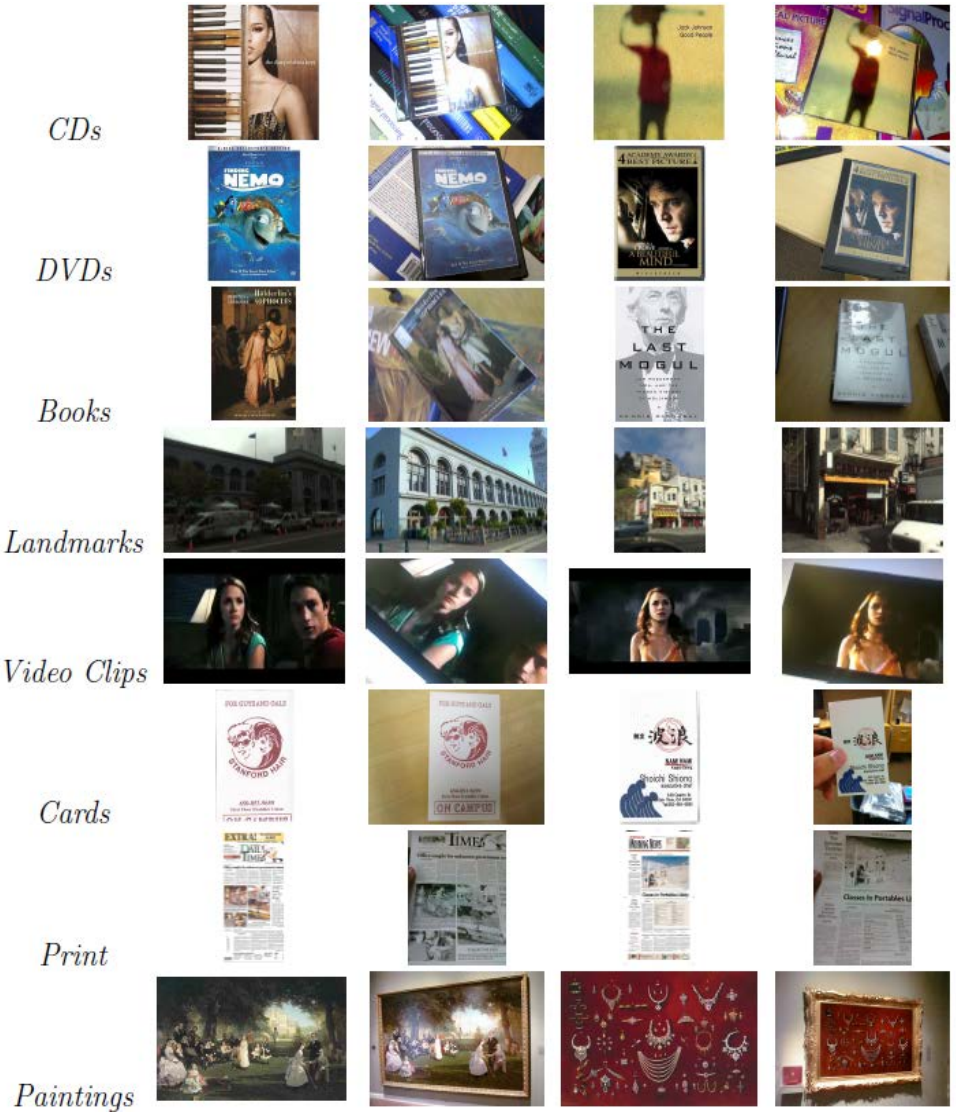
- **Robustness**
  - High matching accuracy shall be achieved at least for images of textured rigid objects, landmarks, and printed documents. The matching accuracy shall be robust to changes in vantage point, camera parameters, lighting conditions, as well as in the presence of partial occlusions.
- **Sufficiency**
  - Descriptors shall be self-contained, in the sense that no other data are necessary for matching.
- **Compactness**
  - Shall minimize lengths/size of image descriptors
- **Scalability**
  - Shall allow adaptation of descriptor lengths to support the required performance level and database size.
  - Shall enable design of web-scale visual search applications and databases.

Requirements (2)

- **Image format independence**
  - Descriptors shall be independent of the image format
- **Extraction complexity**
  - Shall allow descriptor extraction with low complexity (in terms of memory and computation) to facilitate video rate implementations
- **Matching complexity**
  - Shall allow matching of descriptors with low complexity (in terms of memory and computation).
  - If decoding of descriptors is required for matching, such decoding shall also be possible with low complexity.
- **Localization:**
  - Shall support visual search algorithms that identify and localize matching regions of the query image and the database image
  - Shall support visual search algorithms that provide an estimate of a geometric transformation between matching regions of the query image and the database image

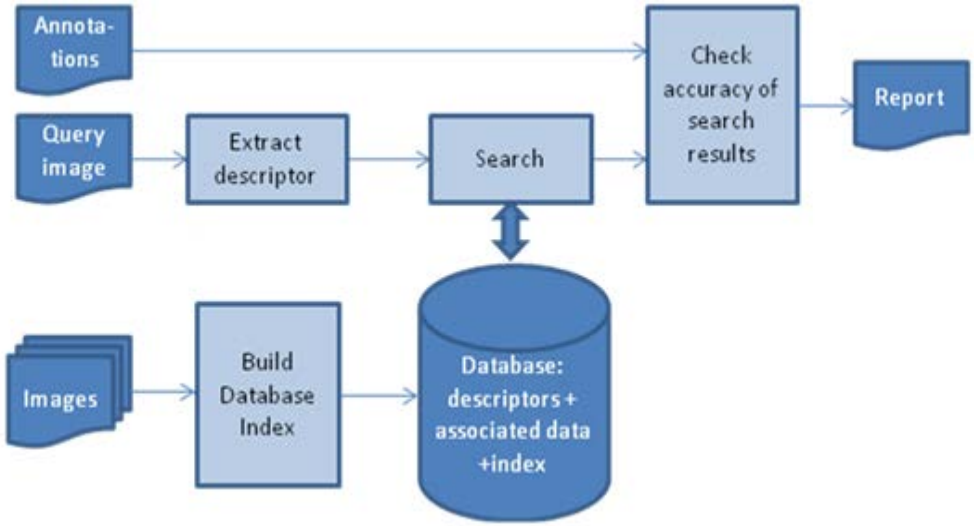
Databases

- Existing:
  - Zurich Buildings
  - University of Kentucky
- MPEG contributions:
  - Stanford university
  - ETRI
  - Telecom Italia
  - Telecom SudParis
  - Peking University
  - Huawei
- 30K annotated images + 1M distractors (FLICKR)



Evaluation Framework

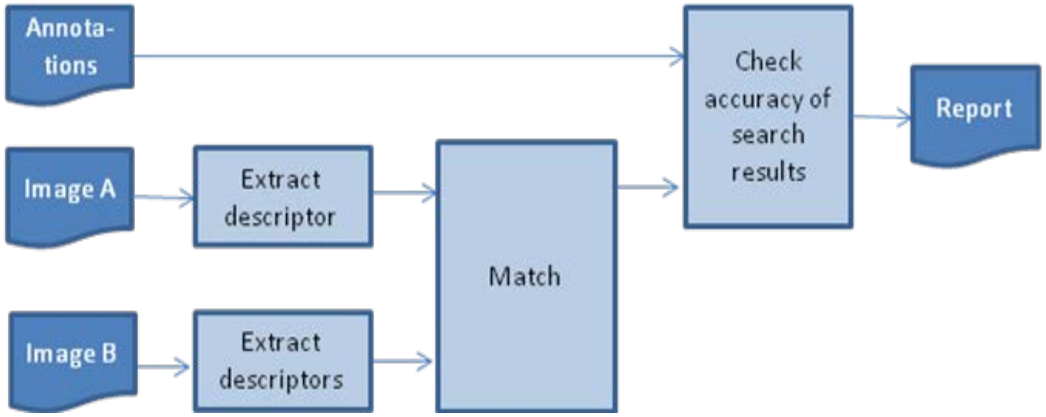
- Retrieval experiment:



- Ground truth annotations include:
  - lists of matching images corresponding to each query image
- Report to include:
  - MAP (mean average precision), accuracy for a top match

Evaluation Framework (2)

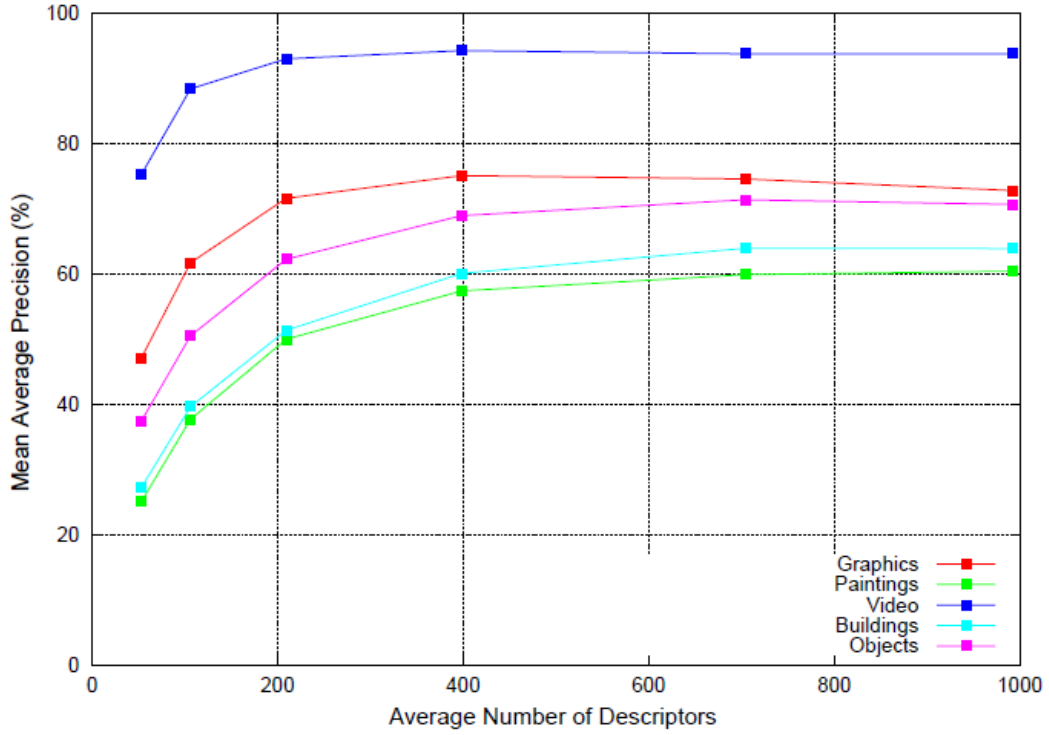
- Pair-wise matching experiment:



- Ground truth annotations include:
  - lists of matching images and non-matching images
  - localization information
- Report to include:
  - success rate at target false alarm rate; localization precision

Crosscheck of Evaluation Framework

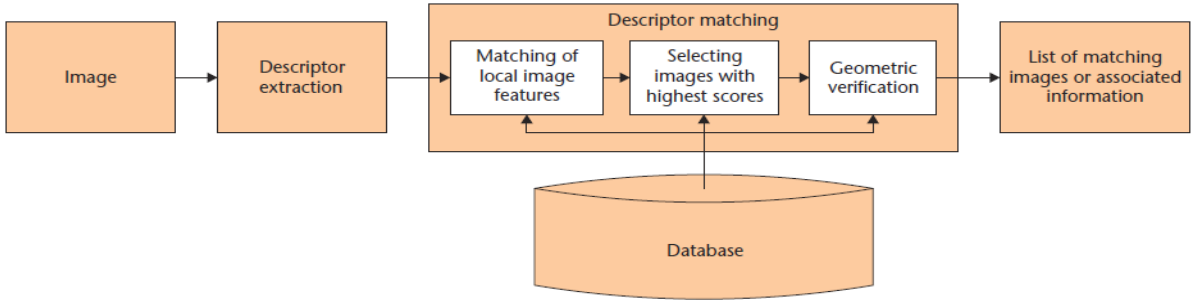
- Retrieval results using uncompressed SIFT features:



- Some datasets (buildings, paintings) are very challenging!

Responses to Call for Proposals

- 11 responses overall
  - 7 industrial labs, 3 universities (few joint proposals)
  - 6 countries: US, UK, Italy, Japan, Korea, China
- Most are based on local image features + BoF retrieval pipeline

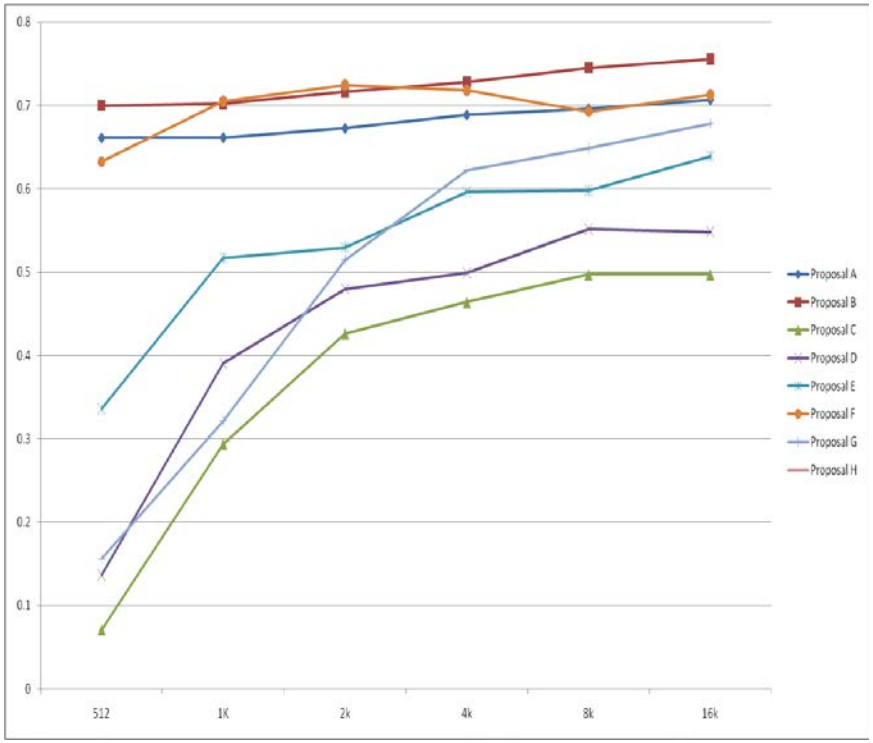


- Some innovations:
  - Novel designs of key-point detectors (some proposals support affine invariance)
  - New ways of achieving rotation invariance (radial gradient transform)
  - Mix of local + global descriptors, etc.

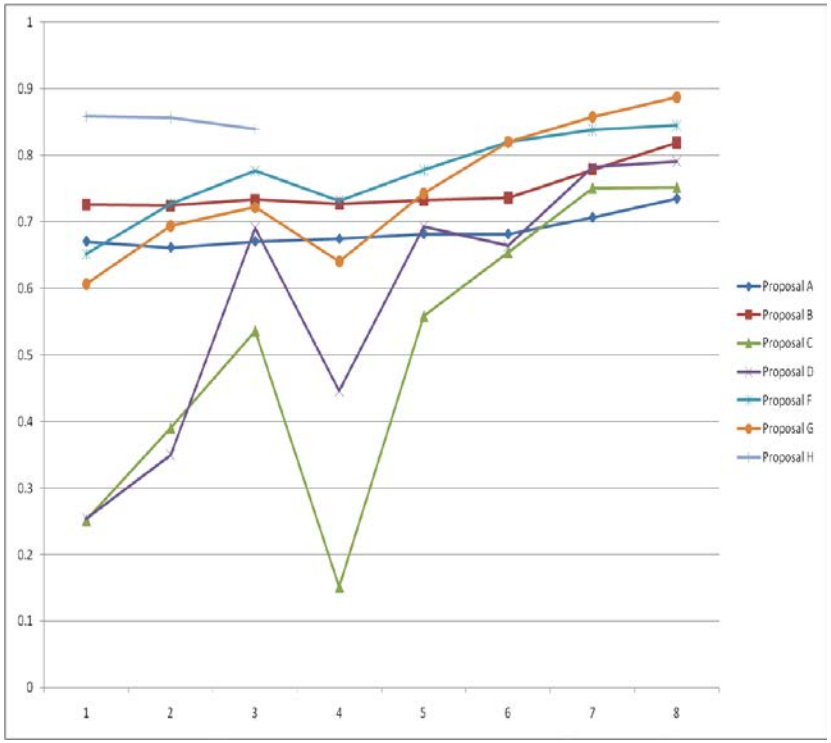
Performance of proposals

- Results for "Buildings" dataset:

Retrieval experiments
[mean average precision]



Pair-wise matching experiments:
[success rate @ 1% false alarm]



Timeline

Meeting / Location/ Date		Action	Comments
96th meeting, Geneva	Mar 25, 2011	Final CfP published Available databases and evaluation software.	
97th meeting, Torino	July 18-23, 2011	Last changes in databases and evaluation software	
98th meeting, Geneva	Nov 26 – Dec 2, 2011	Evaluation of proposals	11 proposals received, selected test model under consideration
99th meeting, San Jose	Feb 10, 2012	WD1	Working Draft 1
100th meeting, Geneva	March 4, 2012	WD2	Working Draft 2
101st meeting, Stockholm	Jul 20, 2012	CD	Committee draft – stabilized and complete specification
103rd meeting, Geneva	January, 2013	DIS	last changes made

Conclusions

- The first major standardization effort in domain of Visual Search has been successfully launched:
  - 11 responses from leading industrial and academic labs
  - solid performance of top few proposals
  - selected Test Model Under Consideration
- The project is now entering collaborative stage:
  - Progression according to MPEG's "core experiment" process
    - New or improved component technologies can be incorporated
- Target completion date: January 2013