Searching Images with MPEG-7 [& MPEG-7-like] Powered Localized dEscriptors

The SIMPLE answer to effective Content Based Image Retrieval

C. Iakovidou, N.Anagnostopoulos, A. Ch. Kapoutsis, Y. Boutalis and S. A. Chatzichristofis
Democritus University of Thrace, Greece.

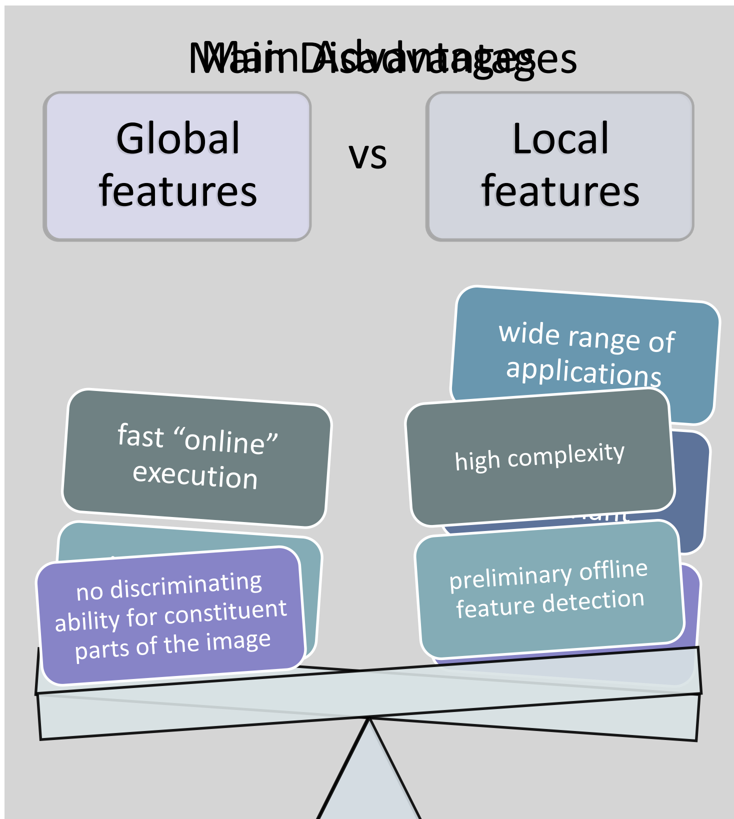
presentation outline

Why?	How?	Worth it?
what type of features form an efficient descriptor global vs local features motivation	provide core technical implementation details experimental set-up	experimental results evaluation advantages limitations

SIMPLE

the SIMPLE idea

the battle of global vs local features for CBIR tasks



What is considered “similar”?

- What does the user rate as effective image retrieval?
- How should we vectorize the features he’s looking for?
- Should we look further for new techniques?

SIMPLE

the SIMPLE idea

motivation and related work (aka so many methods, so little time)

Revisit well established methods from both major image description tactics (global features – local features).

Mix and match. Take a fresh outlook on original thoughts, combine and test strategies based on what we know today about retrieval systems.

Simplify! Produce and test the new descriptors in a straightforward fashion, so we can get some insight on what works together and what doesn't.

the SIMPLE family of descriptors

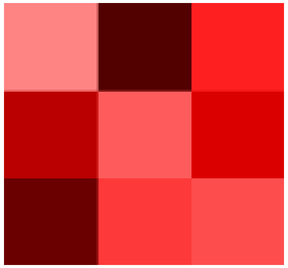
the SIMPLE family

implementation strategy

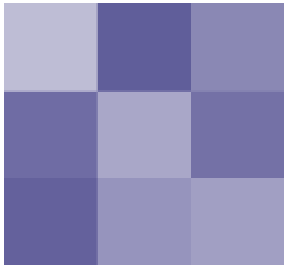
- Images are meaningful when discriminating foreground from background.
- Localized texture information is essential
- Localized color information highly boost retrieval performance
- Image features need to be quantized for faster vector distance measurements
- Compact overall representations.

Texture and Color <u>are not</u> orthogonal properties

All reds

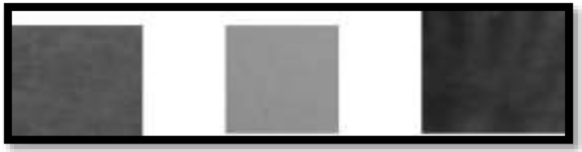


All blues



All edges

All uniform



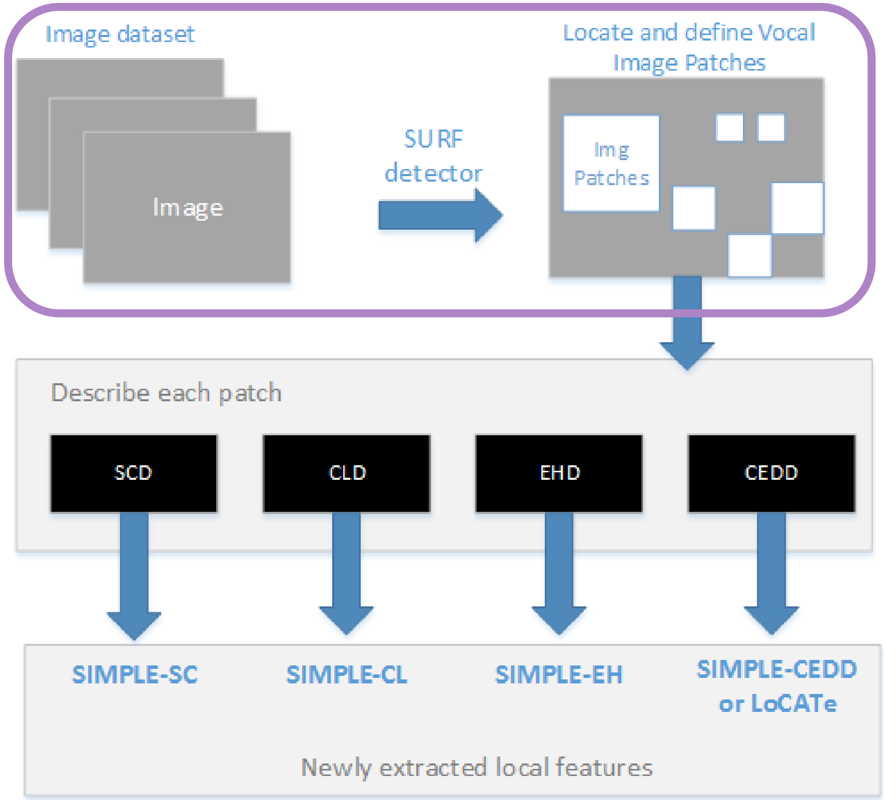
the SIMPLE family

implementation details (POI detection)

- Employ the **SURF** detector
- Utilize achromatic information
- Locate salient image patches of blob-like structures in multiple scales using the Hessian matrix and integral images.

1. Robustness to image transformations
2. Fast execution
3. Easily adapted to parallel processing since each Hessian image can be independently generated

1. Detecting Salient Image patches



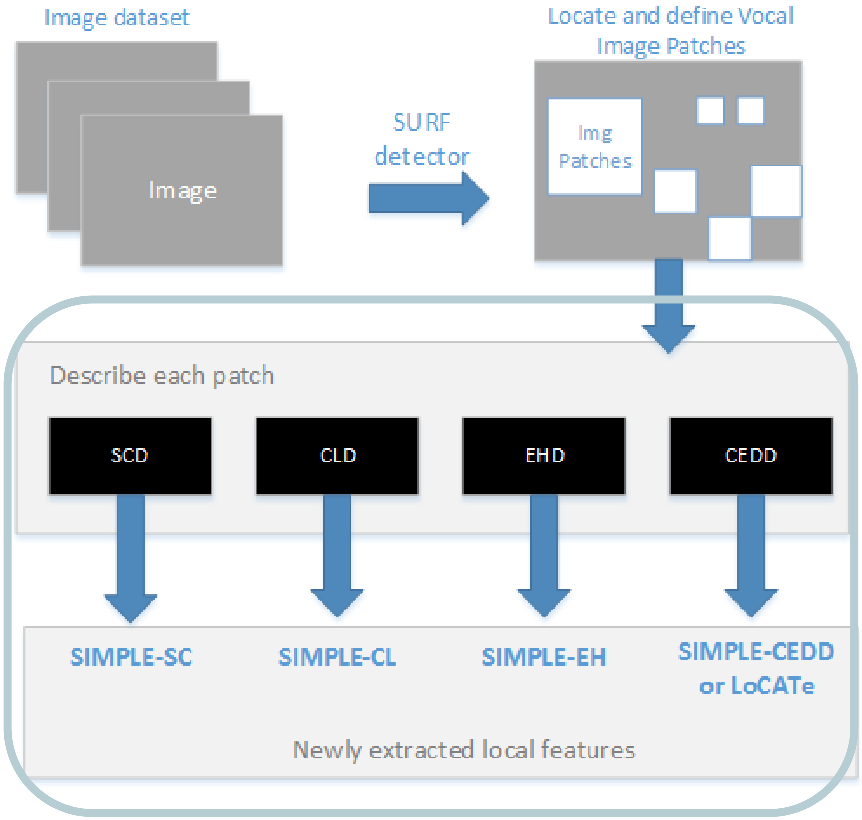
SIMPLE

the SIMPLE family

implementation details (POI description)

We obtained image patches from the whole collection where we know something "interesting" is happening texture wise. (blob-like responses)

2. Describing Salient Image patches



Without actually vectorizing these responses we employed:

- two color based descriptors
  (MPEG-7 SCD, CLD)
- one edge based descriptor
  (MPEG-7 EHD)
- one descriptor that combines color and texture information
  (MPEG-7-like CEDD)

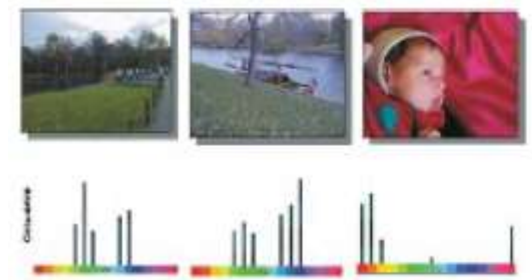
SIMPLE

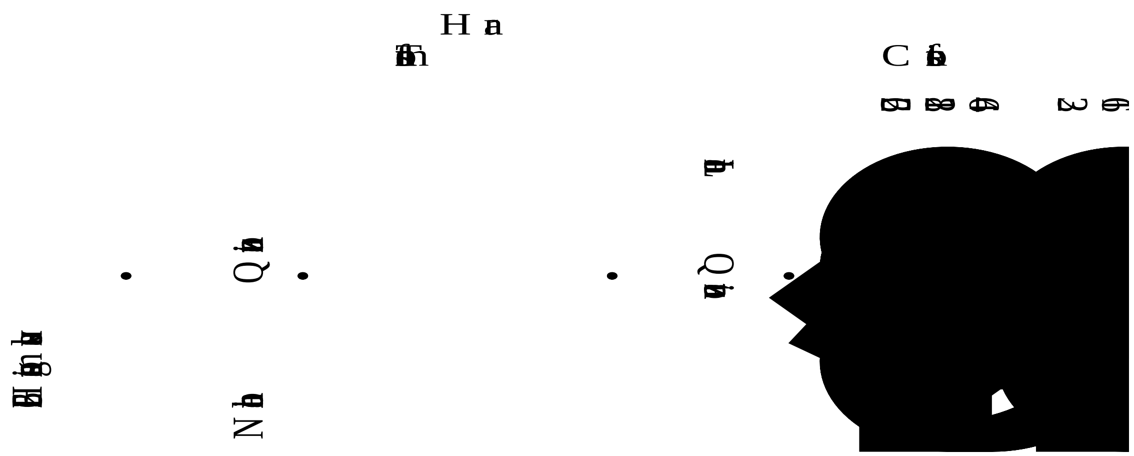
the SIMPLE family

implementation details (POI description)

MPEG-7 Scalable Color Descriptor

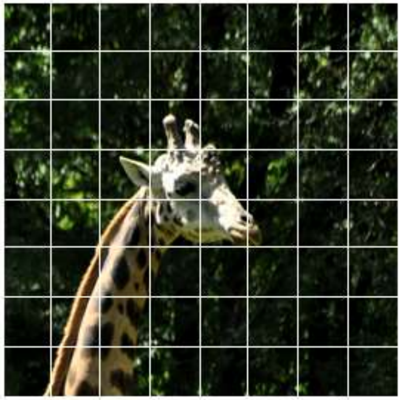
is a color histogram in a fixed HSV color space achieved through a uniform quantization of the space to 256 bins. An encoding step is performed by a Haar transform, for compression. Then, a number of coefficients is used to represent the descriptor. Its representation is scalable in terms of bin numbers and bits used for accuracy. **We followed the default proposed setting of 64 coefficients.**

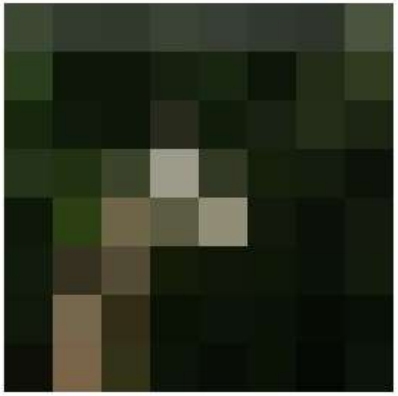


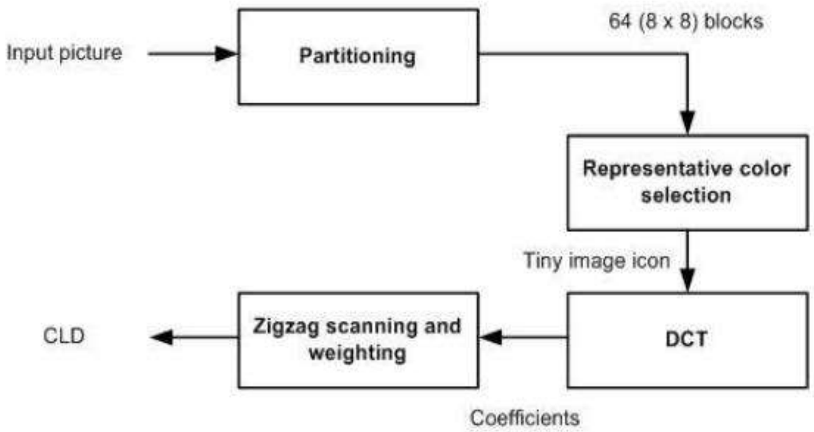


MPEG-7 Color Layout Descriptor

The descriptor **represents** **the spatial distribution of the color** **in images in a compact form**. The image is divided into (64**)** **8 x 8 discrete blocks** and their representative colors in the YCbCr space are extracted. The descriptor is obtained by applying the discrete cosine transformation (DCT) on every block and using its coefficients. **The produced descriptor is a 3 x 64 bin (64-Y, 64-Cb, 64-Cr)** representation of the image





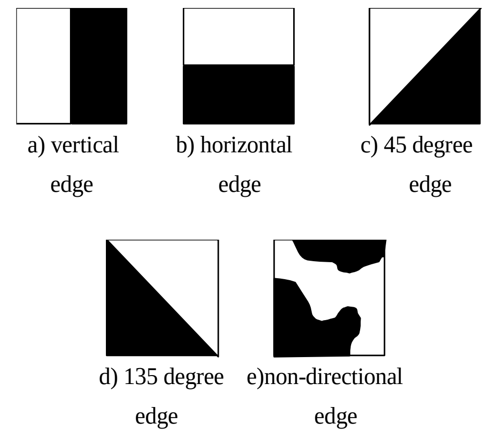


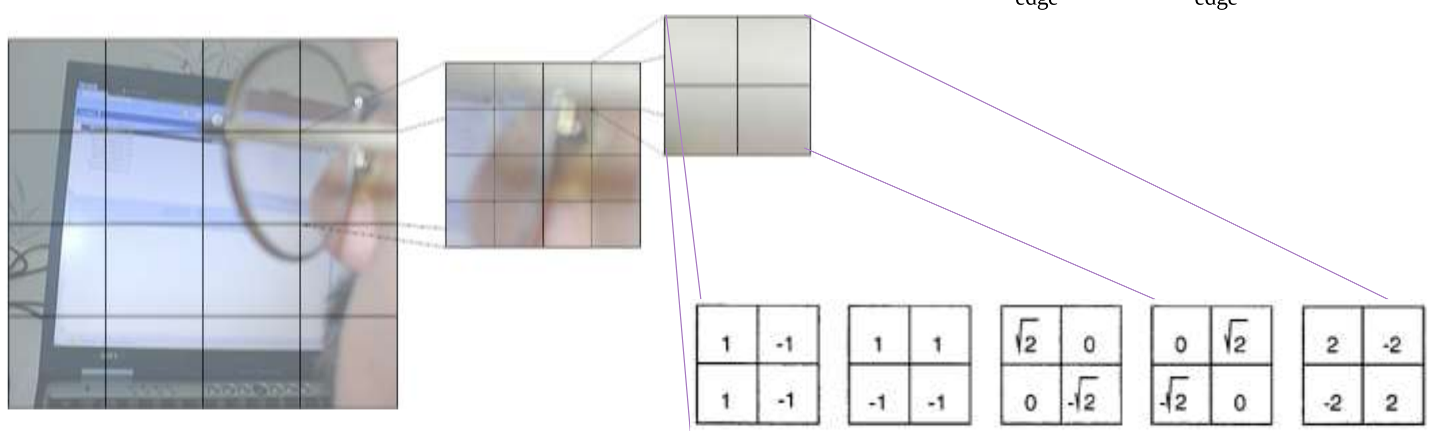
the SIMPLE family

implementation details (POI description)

MPEG-7 Edge Histogram Descriptor

The descriptor represents **the spatial distribution of five types of edges in the image.** A given image is first subdivided into 4 x 4 subimages, and the local edge histogram of five broadly grouped edge types (vertical, horizontal, 45 diagonal, 135 diagonal, and isotropic) is computed. Each edge histogram consists of five bins (one for every edge type). An image subdivided in 16 blocks produces an 80-bins edge descriptor.



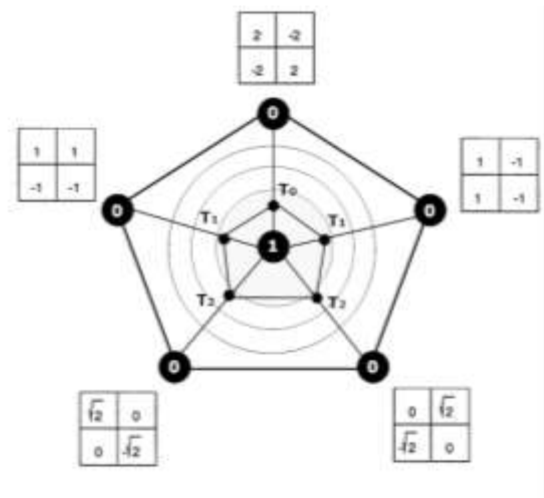


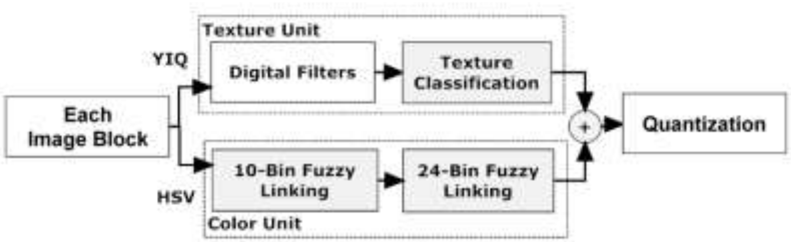
the SIMPLE family

implementation details (POI description)

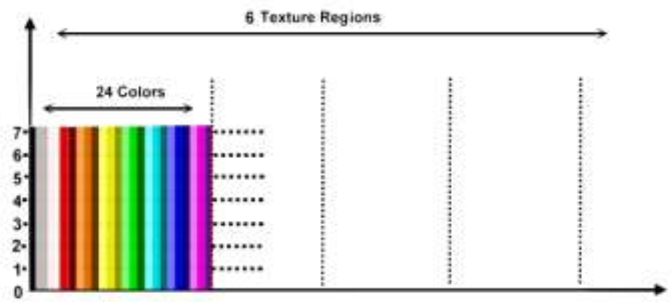
MPEG-7-like Color Edge Directivity Descriptor

CEDD is originally a global descriptor that divides an image into <u>1600 rectangular image areas.</u> Those Image-Blocks are then handled independently to extract <u>their color information</u> (through a two staged Fuzzy Histogram Linking procedure that produces a 24-bin color histogram of pre-set colors) and <u>texture information</u> (employing the five digital filters proposed by the MPEG-7 EHD and using a heuristic fuzzy pentagon diagram to threshold the normalized maximum responses so as to form a 6- bin texture vector). **<u>The obtained vectors are combined in the end</u> to form the 144 bins CEDD descriptor.**





(a)



(b)

the SIMPLE family

implementation

+ Detect regions in multiple scales, that are interesting texture-wise

+ Describe them with 4 different global-features' methods

+ Produce 4 new local features for Image retrieval:

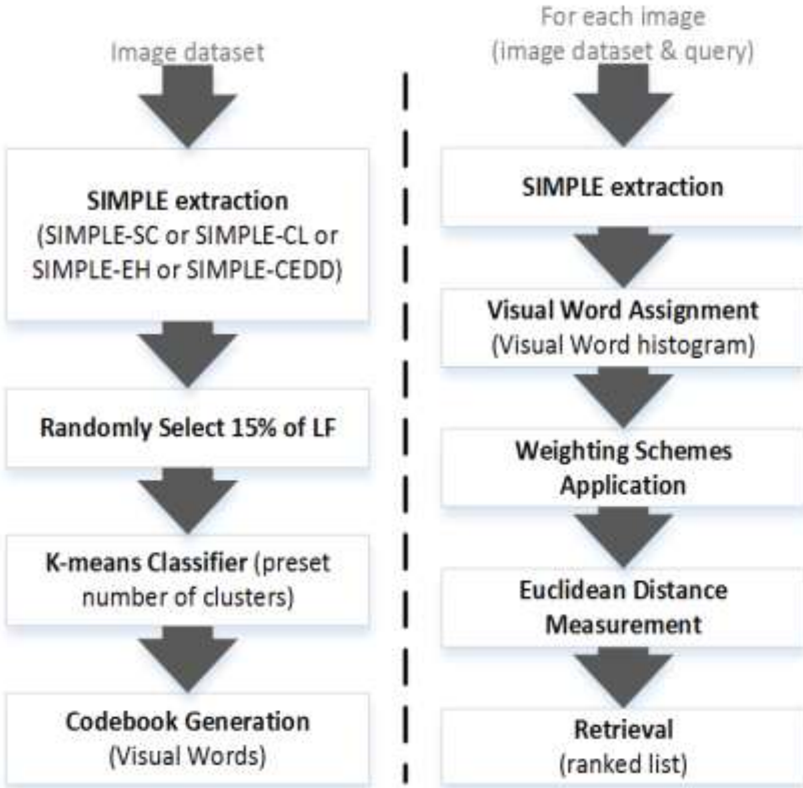
SIMPLE-SC, SIMPLE-CL, SIMPLE-EH, SIMPLE-CEDD
All compact and quantized

testing the SIMPLE descriptors for image retrieval

the SIMPLE family

retrieval system

Bag-of-Visual-Words framework



1. Extract the SIMPLE local features
2. Forward 15% to K-means classifier
3. Prepare the codebooks (32, 128, 512, 2048)
4. Assign VW to all images
5. Employ 8 tf.idf weighting schemes
6. Perform retrieval, ranking results based on the lowest Euclidean distance

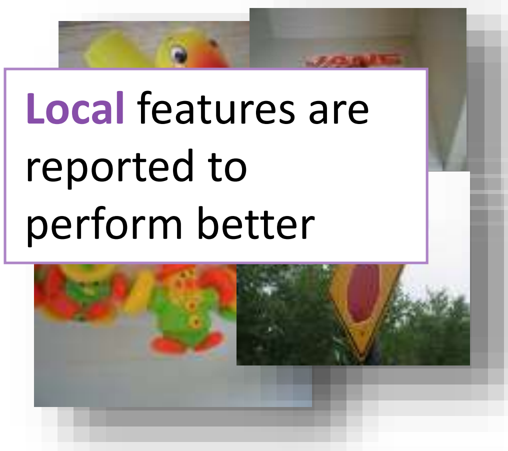
SIMPLE

experimental set-up

image collections, codebook sizes, evaluation metrics

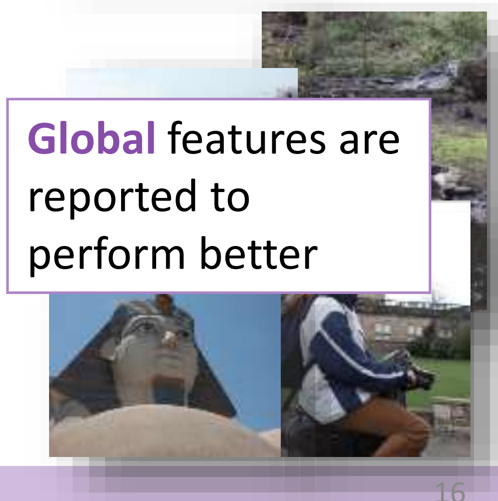
UKBench Image Collection

- Consists of **10200 images arranged in 2250 groups of four images** per group.
- Each group includes depictions of a single object.
- Only images of the same group are considered relevant.
- **The first 250 images of the first 250 groups were used as queries.**



UCID Image Collection

- Consists of **1338** uncompressed Tagged Image File (TIF) format images.
- It covers a variety of topics, including natural scenes and man-made objects.
- **Manual relevance assessments** among all database images are provided.
- the **ground truth consists of images with similar visual concept to the query image**.



experimental set-up

image collections, codebook sizes, evaluation metrics

Codebooks

Four different codebook sizes

- **32** VW,
- **128** VW,
- **512** VW,
- **2048** VW

Evaluation Metrics

- Mean Average Precision (**MAP**) (max at 1)
- MPEG-7 Average Normalized Modified Retrieval Rank (**ANMRR**) (max at 0)
- Precision-at-*K* (P@K) (max at 1)
  P@4 (UKBench)
  P@10 (UCID)

Total Number of experiments

- **Local features** (SURF, SIFT, ORB, BRISK, Oppo. SIFT)
  (**4** SIMPLE + **5** LFDescr) x **4** codebooks x **8** weighting schemes= **288**
- **Global features**
  7 GFDescr

Total of 295 x 3 evaluations= 885 retrieval evaluations

SIMPLE

ntal
ults

llection

other
for 3

CL also

the

Descriptor	Size	WS	MAP	P@4	ANMRR
SIMPLE-SC	512	l.t.c.	0.9145↑”	0.8960↑”	0.0713↑”
SIMPLE-CEDD	512	l.t.c.	0.8964↑”	0.8670↑”	0.0879↑”
SIMPLE-SC	128	l.t.c.	0.8941↑”	0.8640↑”	0.0858↑”
SIMPLE-SC	2048	l.t.c.	0.8730↑”	0.8180↑”	0.0871↑”
SIMPLE-CEDD	128	l.n.c.	0.8665↑”	0.8260↑”	0.1104↑”
SIMPLE-CL	512	l.t.n.	0.8446↑’	0.7710-	0.1333-
SIMPLE-CEDD	2048	l.t.c.	0.8280-	0.7580-	0.1207↑’
SURF(baseline)	512	l.n.n.	0.8159	0.7730	0.1535
SIMPLE-CL	128	l.n.n.	0.8112	0.7640	0.1576
CEDD	Global		0.8026	0.7630	0.1690
SIMPLE-SC	32	l.n.n.	0.7956	0.7420	0.1672
SIMPLE-CEDD	32	l.n.n.	0.7806	0.7250	0.1771
SIMPLE-CL	2048	l.n.c.	0.7693	0.6890	0.1706
SURF	128	l.n.n.	0.7634	0.71[illegible]	[illegible]
Oppo. SIFT	128	n.n.c.	0.7475	0.70[illegible]	[illegible]
SIMPLE-EH	128	n.n.n.	0.3972	0.3760	0.5590
BRISK	128	l.n.n.	0.3904	0.3570	0.5694
ORB	32	n.n.n.	0.3880	0.3560	0.5656
SIMPLE-EH	32	n.n.n.	0.3570	0.3330	0.5987
BRISK	32	n.n.n.	0.3550	0.3190	0.5979
BRISK	512	l.n.n.	0.3463	0.3240	0.6166
Tamura [42]	Global		0.3130	0.2950	0.6582
BRISK	2048	n.n.c.	0.3096	0.2900	0.6524

Our best performing **SIMPLE descriptor** improves

- MAP by 12%
- P@4 by 16% and
- ANMRR by 53%.

SIMPLE

Descriptor	Size	WS	MAP	P@10	ANMRR
SIMPLE-CEDD	2048	l.t.c.	0.7811↑"	0.2590↑"	0.1892↑"
SIMPLE-SC	2048	l.t.c.	0.7718↑"	0.2550↑"	0.1968↑"
SIMPLE-CEDD	512	l.t.c.	0.7635↑"	0.2531↑"	0.2054↑"
SIMPLE-SC	512	l.t.c.	0.7648↑"	0.2515↑"	0.2010↑"
SIMPLE-CEDD	128	l.n.n.	0.7332↑"	0.2447↑"	0.2260↑"
SIMPLE-SC	128	l.t.c.	0.7275↑"	0.2382↑	0.2355↑"
SIMPLE-CL	2048	l.t.c.	0.7161↑"	0.2393↑	0.2502↑'
SIMPLE-CL	512	l.n.n.	0.6765-	0.2225-	0.2829-
CEDD(baseline)	Global		0.6748	0.2267	0.2823
SIMPLE-CEDD	32	l.n.n.	0.6570	0.2206	0.2954
SURF	512	l.n.n.	0.6513	0.2088	0.3113
SIMPLE-SC	32	l.n.n.	0.6450	0.2095	0.3118
SIMPLE-CL	128	l.n.n.	0.6291	0.2073	0.3288
SIFT	512	l.n.n.	0.6261	0.2034	0.3353
SURF	2048	l.n.c.	0.6259	0.2011	0.3387
Oppo. SIFT	2048	n.t.c.	0.6244	0.2050	0.3383
SIMPLE-EH	32	n.n.c.	0.4682	0.1450	0.4948
ORB	128	n.n.c.	0.4642	0.1397	0.5052
BRISK	128	l.n.n.	0.4636	0.1385	0.5070
BRISK	32	n.n.n.	0.4532	0.1370	0.5107
Color Hist.	Global		0.4443	0.1328	0.5231
Tamura	Global		0.4411	0.1317	0.5304
BRISK	2048	n.t.c.	0.4360	0.1328	0.5352
ORB	32	n.n.c.	0.4360	0.1298	0.5332
BRISK	512	l.n.n.	0.4345	0.1347	0.5352

SIMPLE-CEDD and **SIMPLE-SC** increase

- MAP by 14%,
- P@10 by 12% and
- ANMRR by 30%.

SIMPLE

conclusions and discussion

contribution, applications, open issues

- Four novel descriptors were presented in this paper and were tested in the most straightforward fashion to provide some insight on retrieval requirements.
- We believe **SIMPLE-SC and SIMPLE-CL** were successful because they provide color information with textural attention.
- **SIMPLE-CEDD** which has both local color and texture information also performs exceptionally good. Its quantization stages produce retrieval-friendly image representations.
- Some limitations concern image/patch sizes, image collection properties and the generation of the appropriate codebook
- **Further experiments** must be conducted on different collections along with comparisons to more local features to **draw solid conclusions**
- The descriptors **are easy to implement, present high retrieval performance** and **can be adopted** as local features in many other more sophisticated retrieval systems.

SIMPLE

source code

available in C#, Matlab and Java

http://tinyurl.com/SIMPLE-Descriptors

Source Code

For questions, comments, suggestions for improving or bugs reporting please send an email to Nektarios Anagnostopoulos [nek.anag<at>gmail<dot>com]

Part A: SIMPLE-CEDD (LoCATe)

A. C# Version using EmguCV (**reference implementation – Suggested**)

- DLL (32 Bit) – DLL (64 Bit) [GNU GPL]

Instructions: Decompress the file into your bin/debug or bin/release folder. Please note that it is mandatory to include together with the LoCATe_Descriptor.dll file the X32 or the X64 folder. Then add the LoCATe_Descriptor as reference into your project. To extract the list of the descriptors for an image, just call the *.extract* method.

Part B: SIMPLE-SC, SIMPLE-SC and SIMPLE-CL

C# Version using EmguCV

- SIMPLE-EH [DLL 32 bit]-[DLL 64 bit]-[Source 32 Bit]-[Source 64 Bit] [GNU GPL]
- SIMPLE-CL [DLL 32 bit]-[DLL 64 bit]-[Source 32 Bit]-[Source 64 Bit] [GNU GPL]
- SIMLE-SC [DLL 32 bit]-[DLL 64 bit]-[Source 32 Bit]-[Source 64 Bit] [GNU GPL]
- Example Application [Source Code] [GNU GPL]

Instructions:

Decompress the file into your bin/debug or bin/release folder. Please note that it is mandatory to include together with the SIMPLE.dll file the X32 or the X64 folder. Then add the SIMPLE_Descriptor

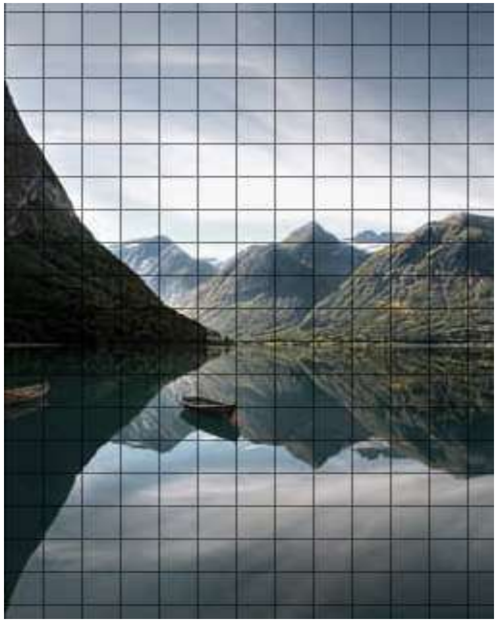
Also included in

Lucene Image Retrieval

LIRE

open source library for *CBIR*

SIMPLE



Thank you!

This research has been co-financed by the European Union (European Social Fund – ESF) and Greek national funds through the Operational Program "Education and Lifelong Learning" of the National Strategic Reference Framework (NSRF) - Research Funding Program: Heracleitus II. Investing in knowledge society through the European Social Fund.