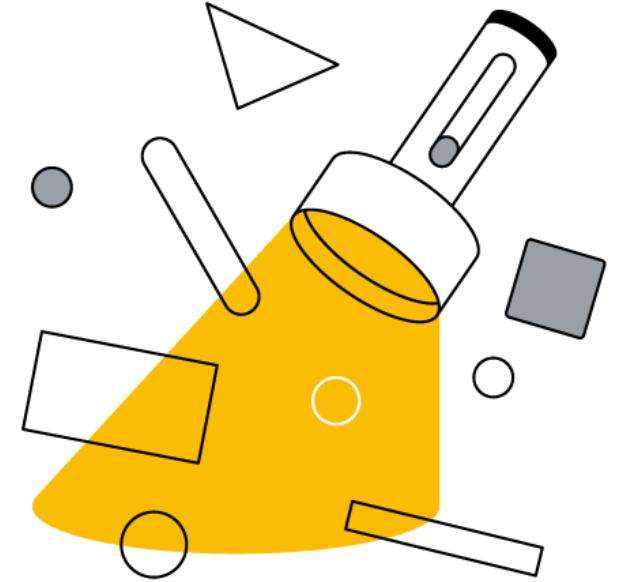


Dokončení: Data ve vědě

- repozitáře vědeckých dat
- specifické souborové formáty
- weby autorů
- datové časopisy

Rešerše dat

- Jaká data by měl dataset obsahovat?
- Jaké proměnné, aby byl užitečný?
- Jaký rozsah dat, záběr?
- Jak data vznikla a jakými úpravami prošla?
- V jakých formátech? Pro jaký SW?
- V jaké licenci by měl být dataset?
- Pro komerční nebo nekomerční využití?



Rešerše dat | Datové repozitáře

- Oborové
(např. adresář [re3data](#))
- Všeobecné
(např. [Zenodo](#) nebo [Dryad](#))
- Institucionální
(např. [Harvard Dataverse](#))



Vyhledávačka



Využijte adresář datových repozitářů [re3data](#) a pokuste se identifikovat **český** datový repozitář pro oblast **sociálních věd**.

Rešerše dat | Weby autorů

- mnozí autoři sdílí data na svých webech
- u článků vždy OA verze a odkaz na data
- většinou velmi kvalitně zpracovaná metadata
- často jediná cesta, jak se k datům dostat

Rešerše dat | Data journals

- některé odborné časopisy umožňují přiložit data
- existují ale také speciální datové časopisy
- datové články – popis data a data
- někdy např. jako [data descriptor](#)
- i zde se provádí recenzní řízení
- mohou být i smíšená periodika

Typy dat

- způsob uložení dat do souboru
- data.**XLSX**
- data.**CSV**, data.**TAB**
- data.JSON
- data.KML, data.GeoJSON – geografický rozměr
- .SAV, .DTA nebo .RDATA – různé analytické programy

Vyhledávačka



Narazili jste na výzkum, který zkoumá obavy z terorismu spojené s přijímáním uprchlíků. Rádi byste si nad daty z výzkumu provedli **svoje vlastní analýzy** – zkuste původní dataset dohledat.

DOI článku: [10.31235/osf.io/d4ewg](https://doi.org/10.31235/osf.io/d4ewg)

Analýza dat

- nadstavba prostého vyhledání dat
- vrstva práce s daty
- vyžaduje specifické dovednosti a znalosti

Hodnocení kvality dat

- data nemusí být čistá – *GI-GO*
- data vznikají v kontextu, kontext je *messy*
- kdo je měří a publikuje? – *jasný zdroj!*
- jaká jsou k nim metadata?
- dokážu zjistit, co který sloupeček znamená?
- je popsána metodologie jejich vzniku?

Čištění dat

- data nejsou vždy čistá
- především ta, která vznikají ručním zápisem
- existují nástroje, které zjednodušují jejich čištění
- práce pro datové knihovníky?



OpenRefine

Citování dat

- **jak citovat dataset?**
- [existují různé návody](#)
- ISO 690 – jako záznam v repozitáři?
- specifikace on-line zdroje: [online databáze]
- mnohdy to autoři datasetu [přímo řeknou](#)

Vyhledávání informací

Image Retrieval

19. 11. 2021

Rozehrivací vyhledávačka



Dali mu prostřední jméno Gamaliel a mezi lety 1915-1921 to dotáhl až na pozici amerického senátora (a později možná ještě dál).

Rozehrivací vyhledávačka



Dali mu prostřední jméno Gamaliel a mezi lety 1915-1921 to dotáhl až na pozici amerického senátora (a později možná ještě dál).

Najděte obrázek tohoto muže
v co největším rozlišení/kvalitě.

Opáčko...

- kolekce obrazů rostou velkým tempem
- nutnost v nich vyhledávat roste s nimi
- celá oblast hnána snahou o automatizaci
- automaticky rozpoznat co/kdo je na obrazu
- mnoho rovin využití...
- *image retrieval (ImR)*

Jak indexovat obraz?

- nutné přidat metadata
- interní/externí
- *EXIF*
- *alt*

EXIF

<http://exif.regex.info/exif.cgi>

exchangeable image file format

specifikace metadat pro uložení do obrazu

Ruční anotace

- ruční tvorba různých metadat
- archivy, muzea, knihovny, galerie
- komerční subjekty, které potřebují přesnost
- většinou kombinují ruční a automatizované přístupy

Title

Description

Object type

Type to search

Collection

Catalogue title

Type to search



Material

Type to search

Technique

Type to search

Objectnumber

Only with image

Now on display

Made between

Start year

End year

Inscription / marks

Catalogue reference

Acquisition credits

Provenance name

IMAGE RESOLUTION

PEOPLE

NUMBER OF PEOPLE

No people

One person

Two people

Group of peo...

AGE

PEOPLE COMPOSITION

ETHNICITY

Black

Caucasian

East Asian

Hispanic/Lat...

Middle Easte...

Mixed Race ...

Multi-Ethnic ...

Native Amer...

Native American/First Nations

Pacific Island...

South Asian

Southeast As...

IMAGE STYLE

COLOR

LOCATIONS



Automatická anotace

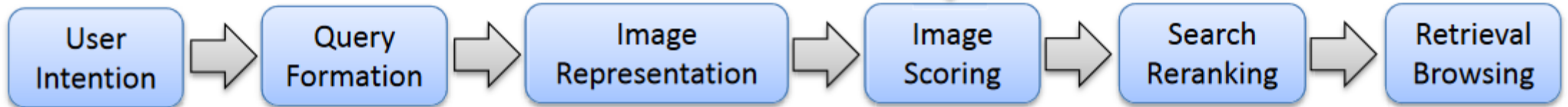
- různé úrovně složitosti
- *content-based ImR* – vyhledávání založené na obsahu
- IBM QBIC – [původní článek](#) (1993)
- automatické rozpoznávání obrazu
- počítačové vidění

In the QBIC (Query By Image Content) project we are studying methods to query large on-line image databases using the images' content as the basis of the queries. Examples of the content we use include color, texture, and shape of image objects and regions. Potential applications include medical ("Give me other images that contain a tumor with a texture like this one"), photo-journalism ("Give me images that have blue at the top and red at the bottom"), and many others in art, fashion, cataloging, retailing, and industry. Key issues include derivation and computation of attributes of images and objects that provide useful query functionality, retrieval methods based on similarity as opposed to exact match, query by image example or user drawn image, the user interfaces, query refinement and navigation, high dimensional database indexing, and automatic and semi-automatic database population. We currently have a prototype system written in X/Motif and C running on an RS/6000 that allows a variety of queries, and a test database of over 1000 images and 1000 objects populated from commercially available photo clip art images. In this paper we present the main algorithms for color texture, shape and sketch query that we use, show example query results, and discuss future directions.

Offline Stage



Online Stage



Content-based ImR

- *content-based image indexing* – indexování
 - indexování mimo textové charakteristiky
 - prvním krokem je identifikace těchto charakteristik
 - poté hledání způsobů, jak je zachytit/popsat
-
- co může charakterizovat
obrázek kromě *txt* metadat?



Content-based ImR

např. barva, textura, tvar...
získávání příznaků

Camera:	Canon EOS 400D Digital
Exposure:	0.005 sec (1/200)
Aperture:	f/11
Focal Length:	100 mm
ISO Speed:	100
Exposure Bias:	0/3 EV
Flash:	Flash fired
Orientation:	Horizontal (normal)
X-Resolution:	72 dpi
Y-Resolution:	72 dpi
Software:	Adobe Photoshop CS2 Windows
Date and Time:	2008:01:06 20:59:54
YCbCr Positioning:	Co-Sited
Exposure Program:	Manual
Date and Time (Original):	2008:01:06 14:47:06
Date and Time (Digitized):	2008:01:06 14:47:06
Shutter Speed:	500948/65536
Metering Mode:	Pattern
Color Space:	sRGB
Focal Plane X-Resolution:	4433.295 dpi
Focal Plane Y-Resolution:	4453.608 dpi
Exposure Mode:	Manual
Compression:	JPEG



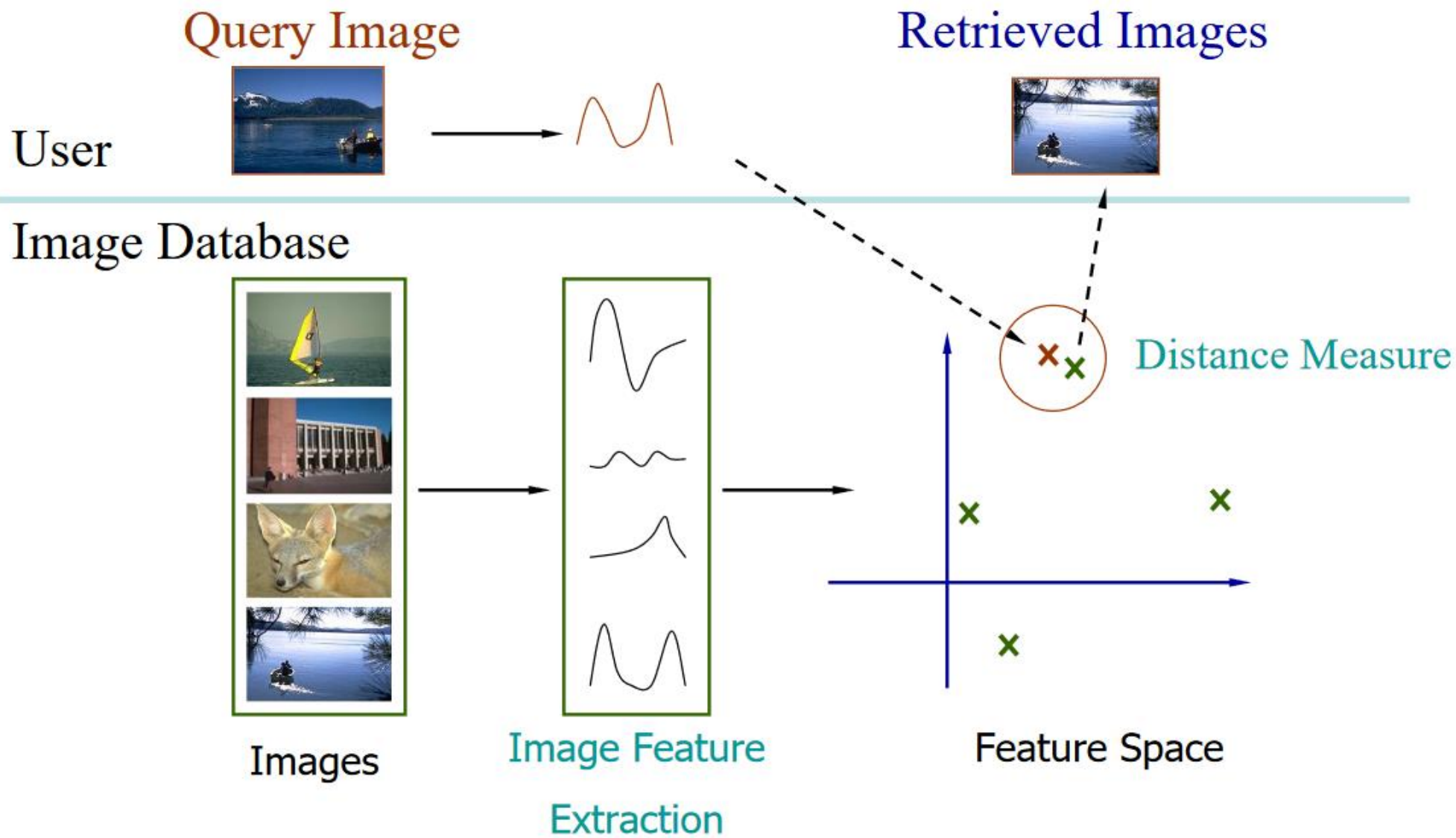


Table 1: Grouping of the features into different types. (a) color representation, (b) texture representation, (c) local features, (d) shape representation. The table also gives the time to extract the features from 10 images and to query 10 images in a 10 image database to give an impression of the computational costs of the different features (experiments were performed on a 1.8GHz machine).

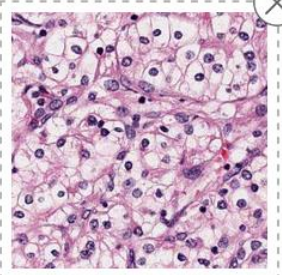
Feature name	Section	comp. measure	type	extr.[s]	retr.[s]
Appearance-based Image Features					
32×32 image	3.1	Euclidean	abcd	0.25	0.19
X×32 image	3.1	IDM	abcd	0.25	9.72
Color Histograms	3.2	JSD	a	0.77	0.16
Tamura Features	3.3	JSD	b	14.24	0.13
Global Texture Descriptor	3.4	Euclidean	b	3.51	0.16
Gabor histogram	3.5	JSD	b	8.01	0.12
Gabor vector	3.5	Euclidean	b	8.68	0.17
Invariant Feature Histograms					
w. monomial kernel	3.6	JSD	ab	28.93	0.16
w. relational kernel	3.6	JSD	ab	18.23	0.14
LF Patches					
global search	3.7	-	ac	4.69	7.13
histograms	3.7	JSD	ac	4.69+5.17	0.27
signatures	3.7	EMD	ac	4.69+3.37	0.55
LF SIFT					
global search	3.7	-	cd	11.91	9.23
histograms	3.7	JSD	cd	11.91+6.23	0.27
signatures	3.7	EMD	cd	11.91+4.50	1.03
MPEG 7: scalable color	3.8.1	MPEG7-internal	a	0.48	0.42
MPEG 7: color layout	3.8.2	MPEG7-internal	ad	0.20	0.33
MPEG 7: edge histogram	3.8.3	MPEG7-internal	b	0.16	0.43

https://en.wikipedia.org/wiki/List_of_CBIR_engines

Luigi Luigi-Genome

Search Image

Search with demo images

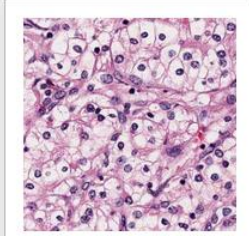
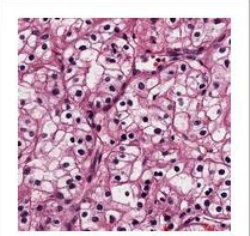
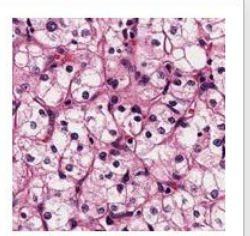
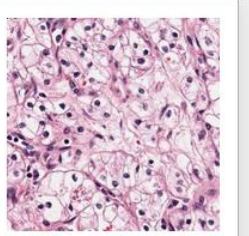
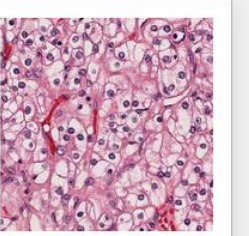
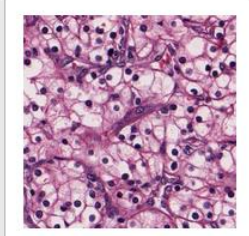
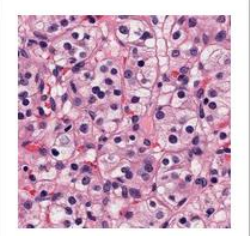
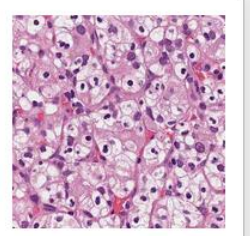
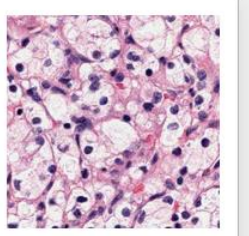
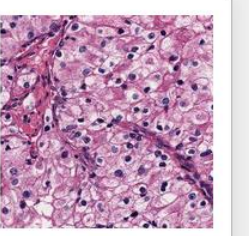


Filename: 05.jpg

Search

Results of image retrieval

Select a body location/system

 <p>Kidney renal clear cell ca... Similarity: 0.999</p> <p>TCGA-B0-4814-01Z-00-DX1</p>	 <p>Kidney renal clear cell ca... Similarity: 0.961</p> <p>TCGA-B0-4821-01Z-00-DX1</p>	 <p>Kidney renal clear cell ca... Similarity: 0.954</p> <p>TCGA-B0-4852-01Z-00-DX1</p>	 <p>Kidney renal clear cell ca... Similarity: 0.952</p> <p>TCGA-CZ-4863-01Z-00-DX1</p>	 <p>Kidney renal clear cell ca... Similarity: 0.951</p> <p>TCGA-B0-5106-01Z-00-DX1</p>
 <p>Kidney renal clear cell ca... Similarity: 0.950</p> <p>TCGA-BP-4327-01Z-00-DX1</p>	 <p>Kidney renal clear cell ca... Similarity: 0.950</p> <p>TCGA-EU-5907-01Z-00-DX1</p>	 <p>Kidney renal clear cell ca... Similarity: 0.949</p> <p>TCGA-A3-3362-01Z-00-DX1</p>	 <p>Kidney renal clear cell ca... Similarity: 0.948</p> <p>TCGA-EU-5906-01Z-00-DX1</p>	 <p>Kidney renal clear cell ca... Similarity: 0.947</p> <p>TCGA-B4-5838-01Z-00-DX1</p>

Color-based ImR

- analýza na nejnižší úrovni
 - *barevné momenty* (ze statistických centrálních m.)
 - *barevné histogramy, color signatures,...*
 - nenesou žádnou informaci o prostorovém rozložení
 - stane se něco, když obrázek otočím?
-
- proto se využívají tzv. *spatial color features*
 - např. *barevné korelogramy*, *vektory koherence barev,...*

barevné momenty



Index Image



Test Image 1



Test Image 2

$$\begin{bmatrix} 0.1016 & 0.1149 & 0.1779 \\ 0.8583 & 0.1139 & 0.0563 \\ 0.6416 & 0.2994 & 0.0974 \end{bmatrix}$$

Index Image

$$\begin{bmatrix} 0.1718 & 0.0986 & 0.1400 \\ 0.7619 & 0.1508 & 0.0455 \\ 0.7062 & 0.2242 & 0.0772 \end{bmatrix}$$

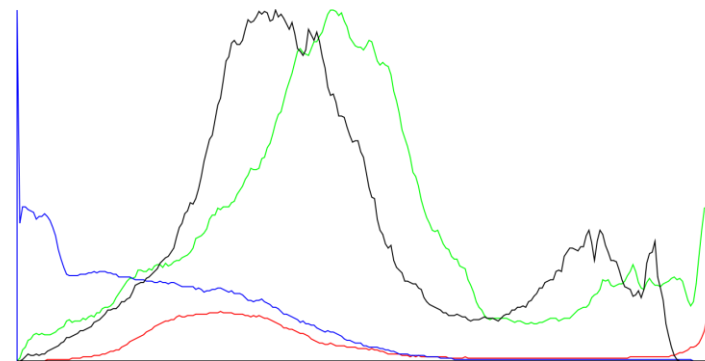
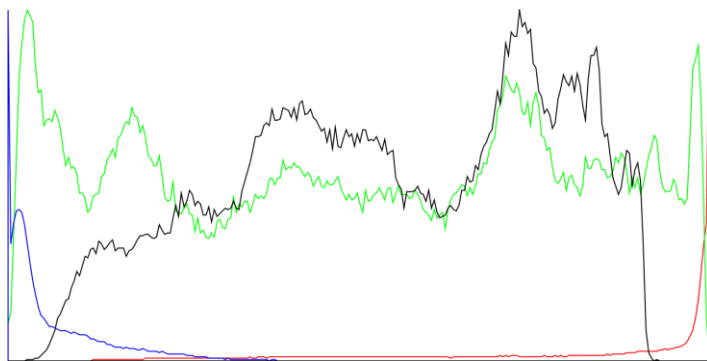
Test Image 1

$$\begin{bmatrix} 0.1878 & 0.1671 & 0.2331 \\ 0.2462 & 0.2281 & 0.2492 \\ 0.6052 & 0.3532 & 0.1534 \end{bmatrix}$$

Test Image 2

$$d_{mom}(Index, Test1) < d_{mom}(Index, Test2)$$

barevný histogram vs. korelogram



vektor koherence barev



Histogram rank: 50. CCV rank: 26.



Histogram rank: 35. CCV rank: 9.

Indexace podle textury

- lokální binární vzor
- co-occurrence matrix

Indexace podle tvaru

- edge histogram
- image moments

ISSN:0975-9646

Neetesh Prajapati et al. / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 7 (1) , 2016, 407-412

Edge Histogram Descriptor, Geometric Moment and Sobel Edge Detector Combined Features Based Object Recognition and Retrieval System

Neetesh Prajapati
M. Tech Scholar
VNS college, Bhopal

Amit Kumar Nandanwar
Assistant Professor
VNS college, Bhopal

G.S. Prajapati
HOD
VNS college, Bhopal

Abstract: Shape is one of the high level features that play an important role in the object recognition and perception. Object shape features provide a powerful clue to object identity. In this paper, we have implemented the approach by combining three feature descriptor i.e. edge histogram descriptor, geometric moment, and Sobel edge detector techniques recognize the objects in the images that is invariant with the changes , scaling, rotation, and orientation. Since edges play an important role in image perception and it is frequently used as a feature descriptor in image retrieval so we select the Edge Histogram Descriptor, EHD, as a feature vector which represents the spatial distribution of five types of edges. We also select geometric moment invariant as another shape feature vector which is extensively used to extort global features, can improve the previously recognition rate to a significant measure. Due to Sobel operator's smoothing effect, it is less sensitive to noise present in Images. On the other hand, smoothing affects the accuracy of edge detection. So next we select Sobel edge detection as a third feature vector that extract the shape features. Finally, we combined the all three shape features that form the 3 dimensions feature vectors of the entire image. Experiment results conclude that the proposed scheme has a very good performance in respect of the precision recall.

Keyword: Edge Histogram Descriptors, Geometric Moment, Sobel Edge Detector.

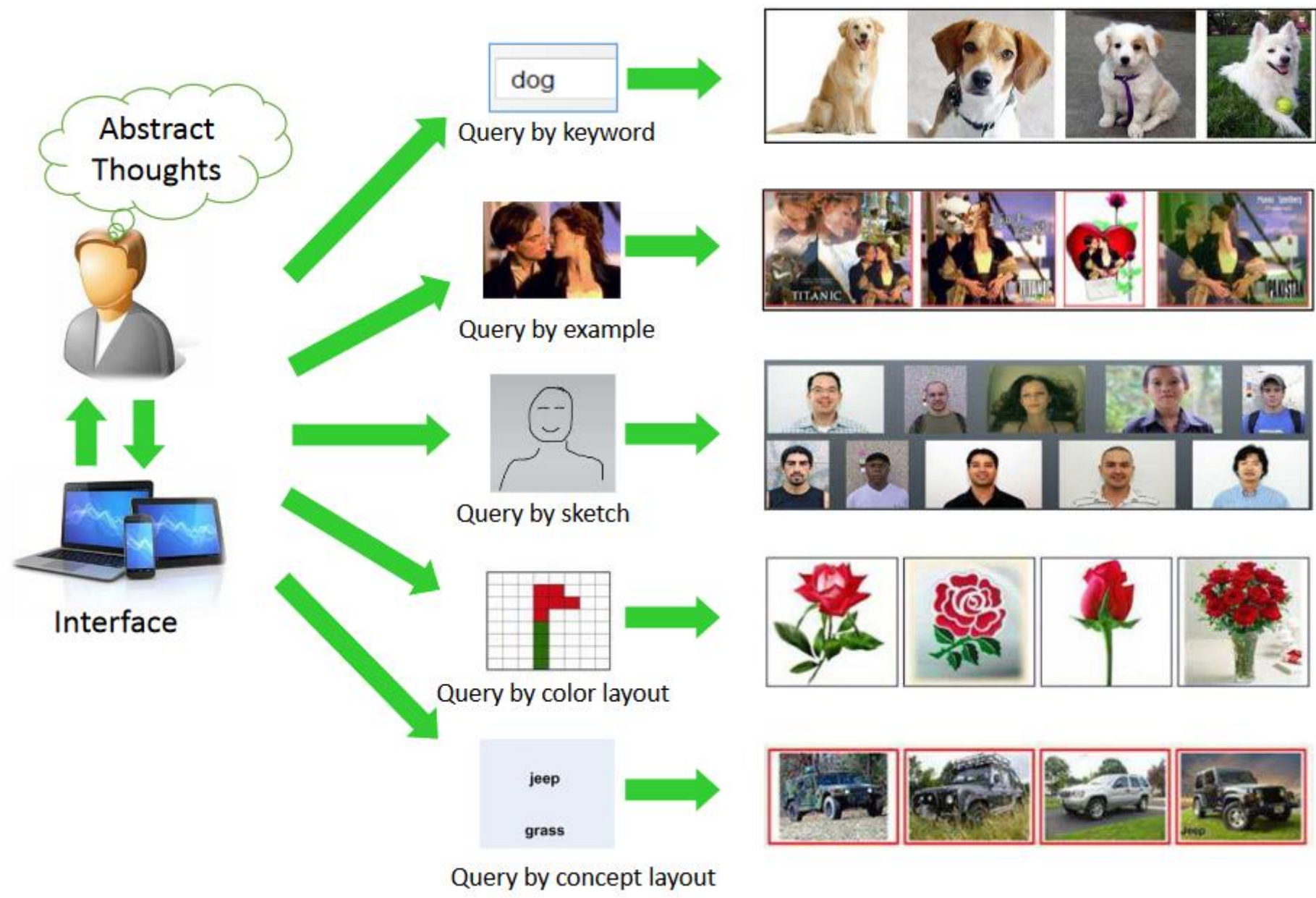
I. INTRODUCTION

For image retrieval. Object shape features can be used to

order to meet certain qualitative standards [3] as well as their retrieval computational cost. While studies have been extended to content-based three dimensional (3D) shape retrieval methods [4], still pattern recognition by 2D shape descriptors can be used in many practical tasks, for example in image matching, multi temporal image sequence analysis, shape classification and character recognition. Furthermore, their quantitative characteristics which still remain superior make them widely used and effective [5].

Various shape descriptors exist in the literature, mainly categorized into two groups: contour-based shape descriptors and region-based shape descriptors. Contour-based methods need an extraction of boundary information which in some cases may not available. Such methods completely ignore the important features inside the boundaries. Region-based methods, however, do not rely on shape boundary information, but they take into account all the pixels within the shape region. Region-based image retrieval methods, firstly apply segmentation to divide an image into different regions/segments, by setting threshold values according to the desirable results. Whereas the boundary of an image can be obtained by applying any edge detection method to an image [6]. Therefore, for generic purposes, both types of shape representations are necessary.

In this study, a computer vision system recognizing objects



Query-by-color

- implementováno v různých systémech
- využívá se často u digitalizovaných sbírek
- používá např. **Europeana**

- [TinEye Multicolor Engine](#)
- [MultiColr](#)
- [Google ArtPalette](#)

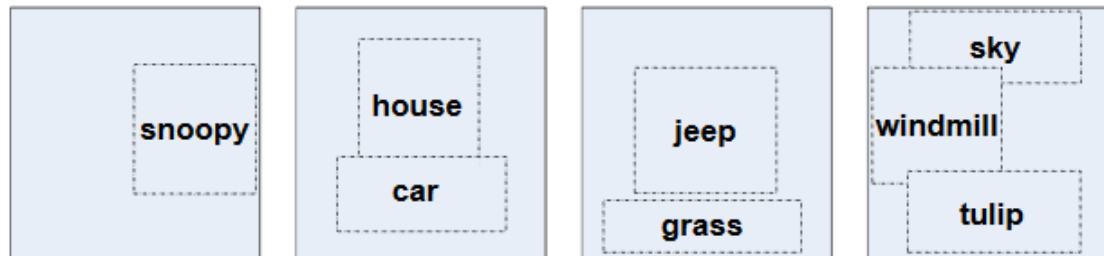
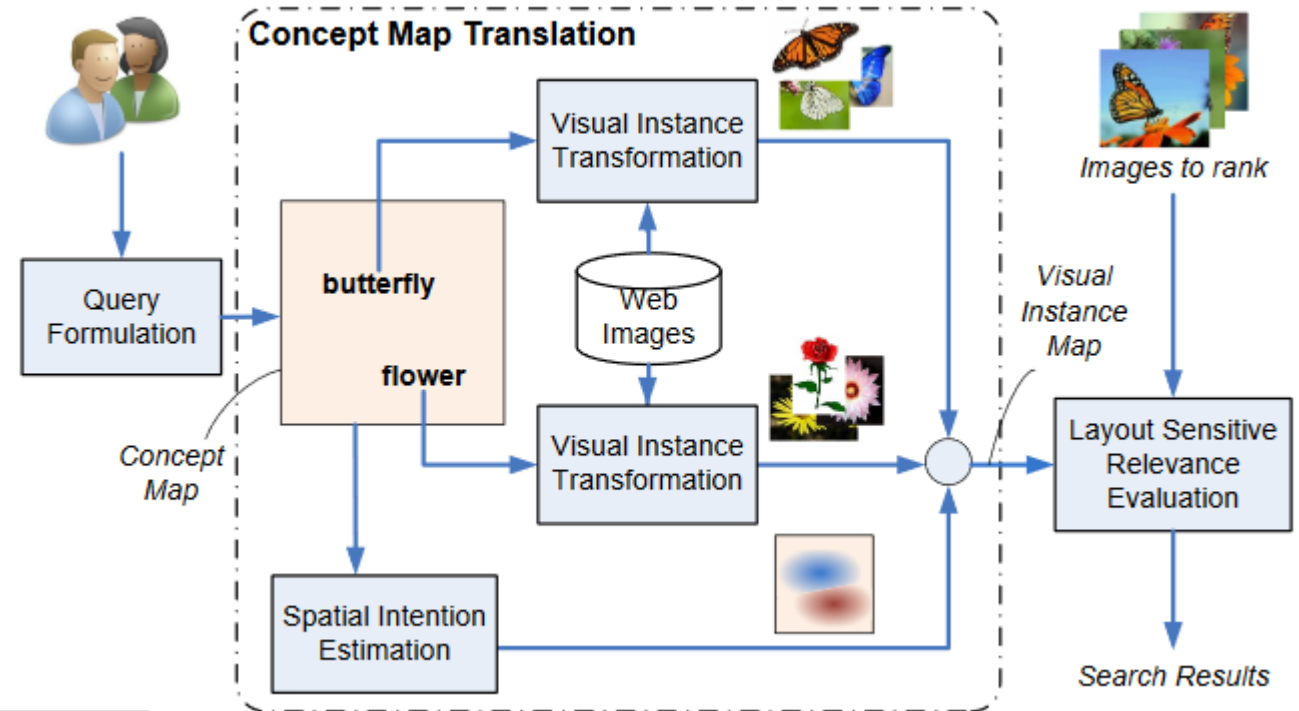


QBE | Query-by-example

- dotaz příkladem, reverzní vyhledávání
- Google Image Search, TinEye
- [Google Lens](#)

Query-by-concept-map

- [Microsoft](#)
- *concept layout*
- záleží i na rozložení

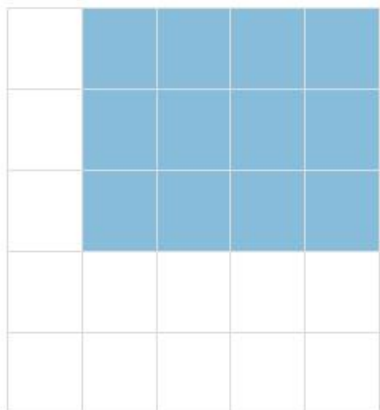




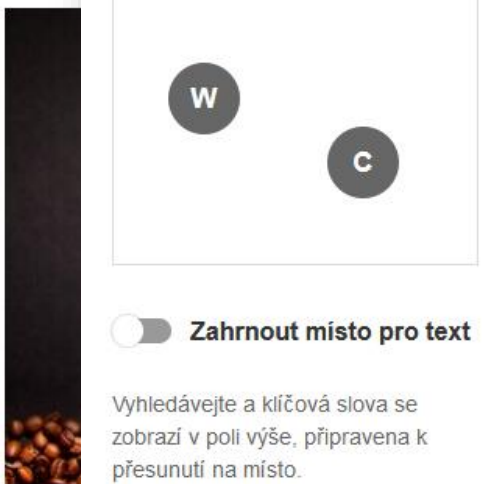
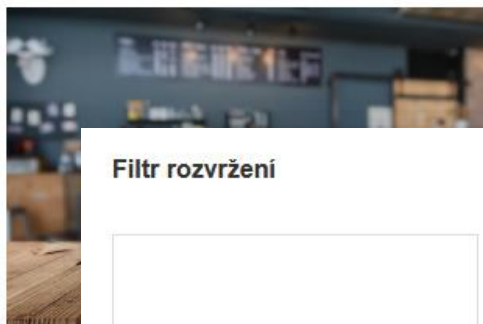
Layout filter

Reset

Copy Space



Click and drag within the grid to select where within the image copy space is needed.



Filtr rozvržení

Vyhledávání kompozice



Zahrnout místo pro text

Vyhledávejte a klíčová slova se zobrazí v poli výše, připravena k přesunutí na místo.

Proč je tato stránka částečně v angličtině?

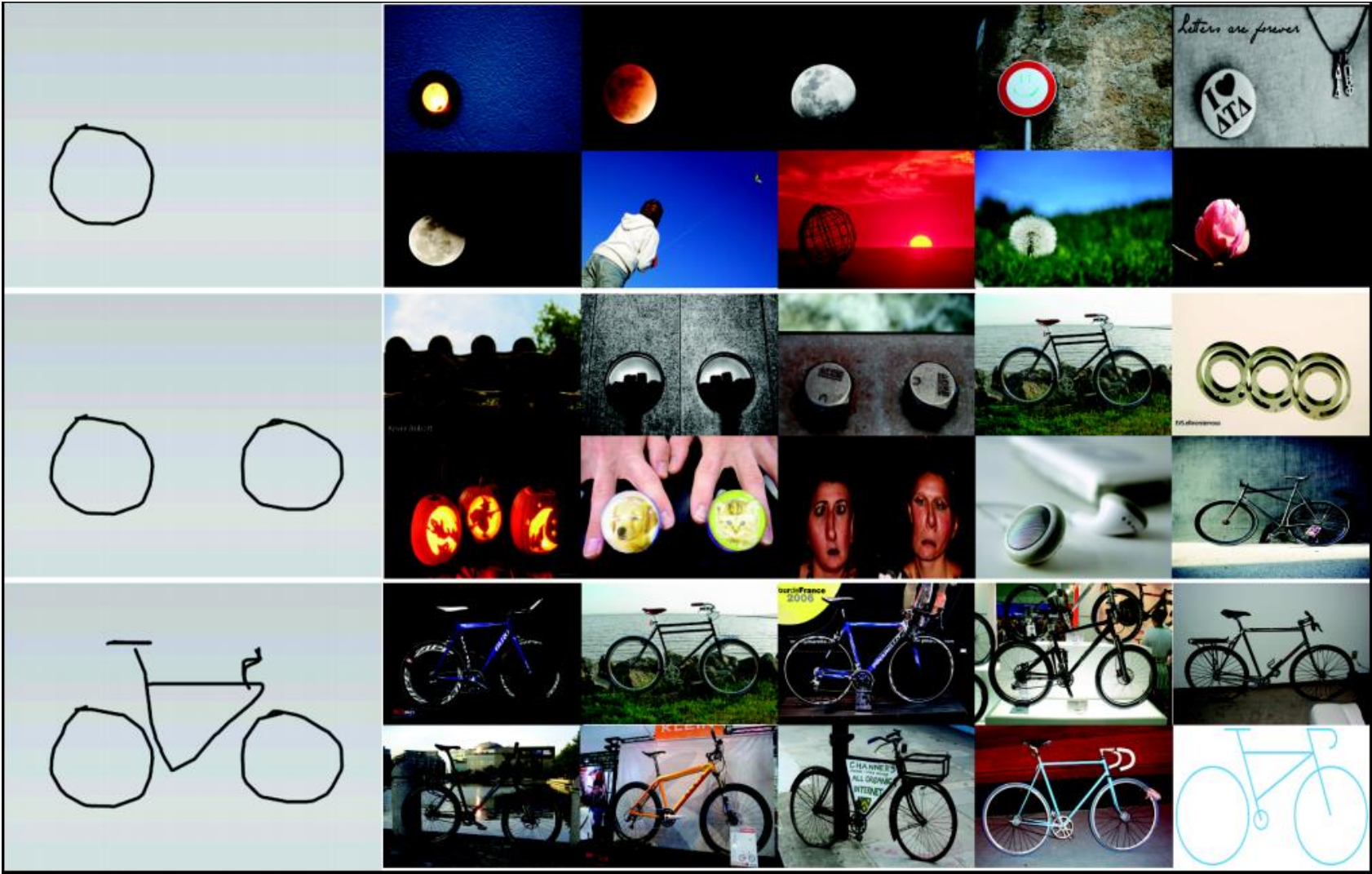


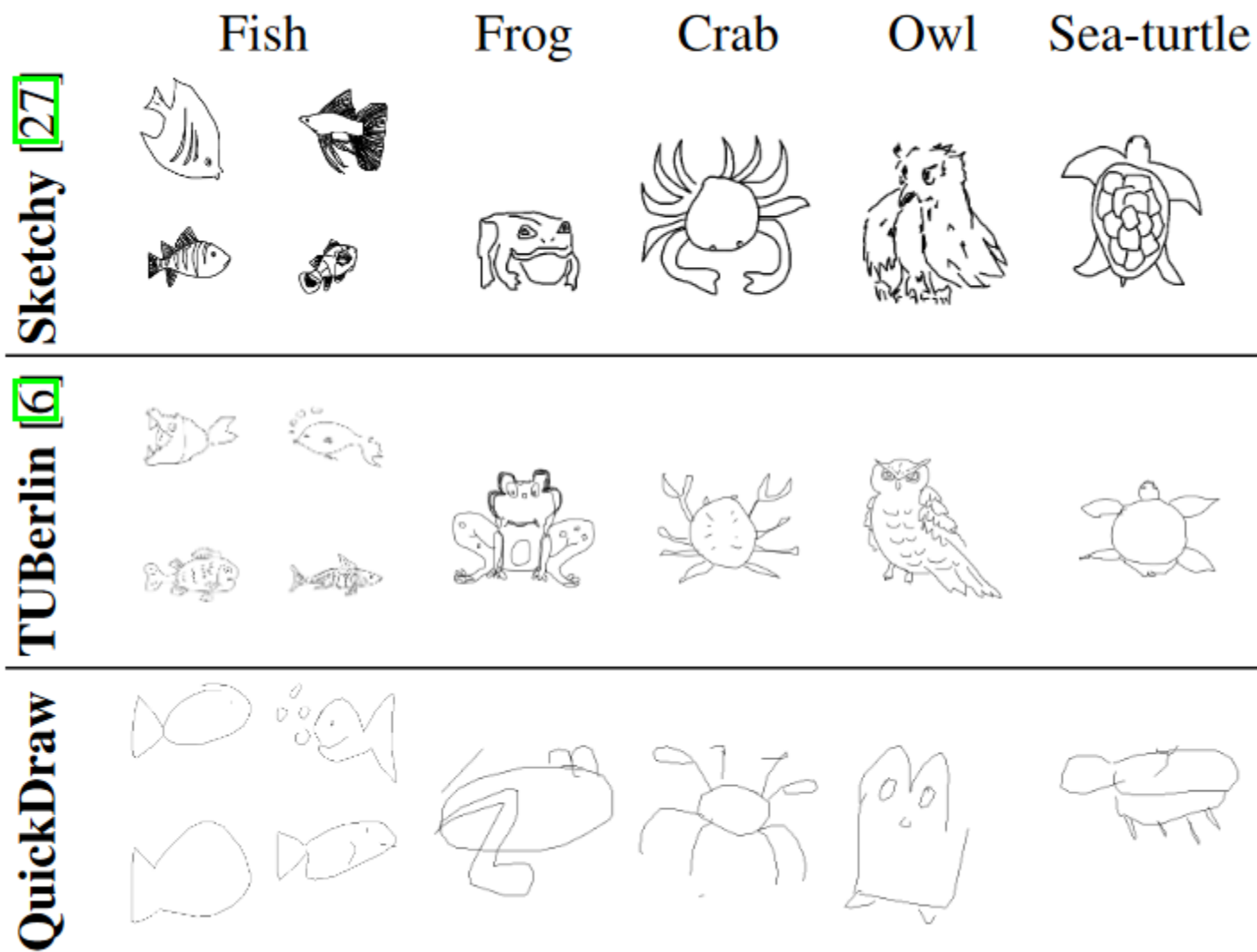
SBIR | Sketch-based ImR

Google QuickDraw



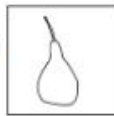






















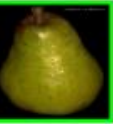










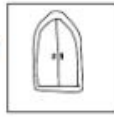

































MS Mind Finder

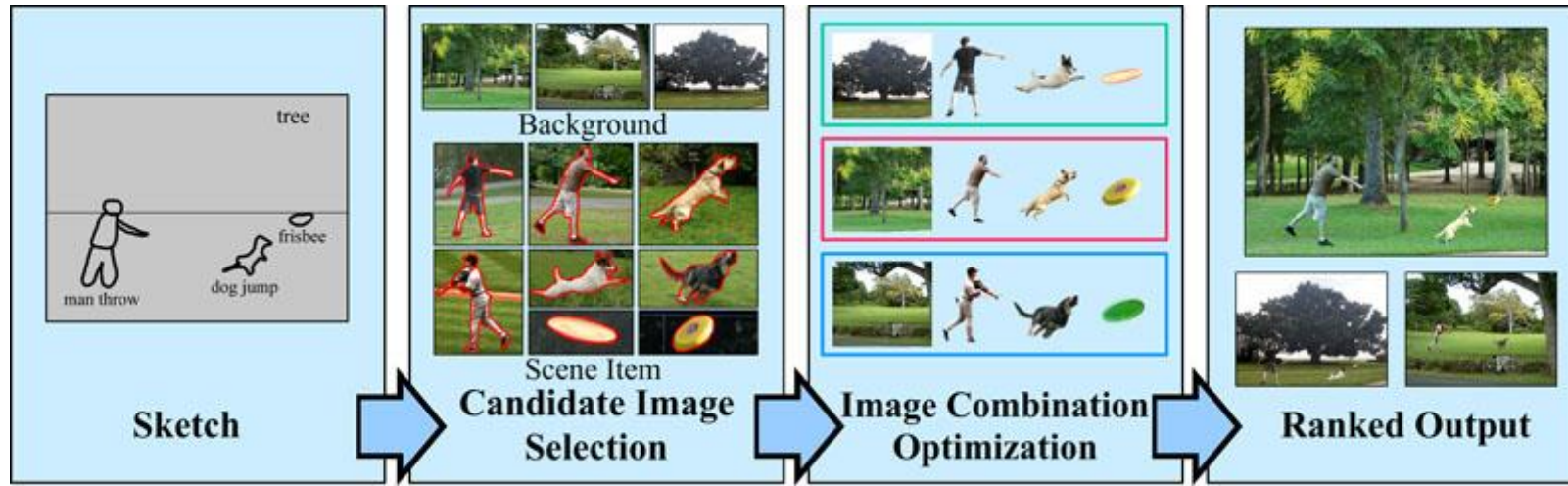




Sketchy [27]

QuickDraw

	Query	Top-8 retrieved candidates									Query	Top-8 retrieved candidates							
CVAE [36]																			
Ours	pear									skyscraper									
CVAE [36]																			
Ours	door									helicopter									



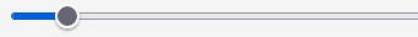
Fill/brush color:



Brush shape:



Brush size: 24



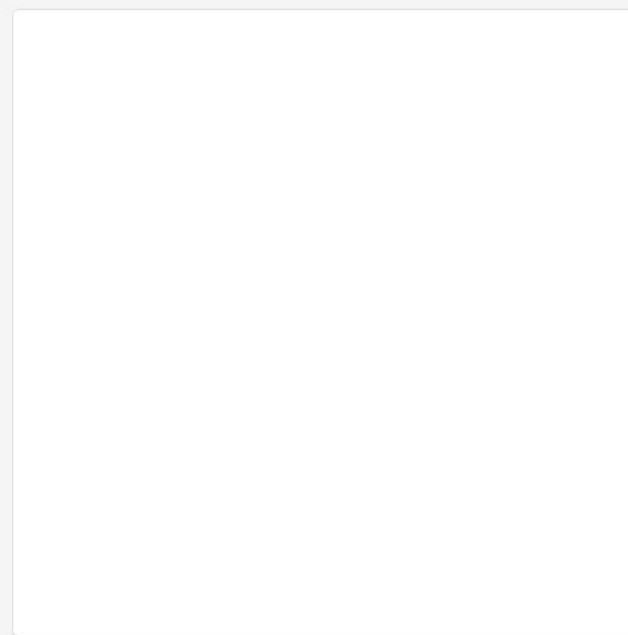
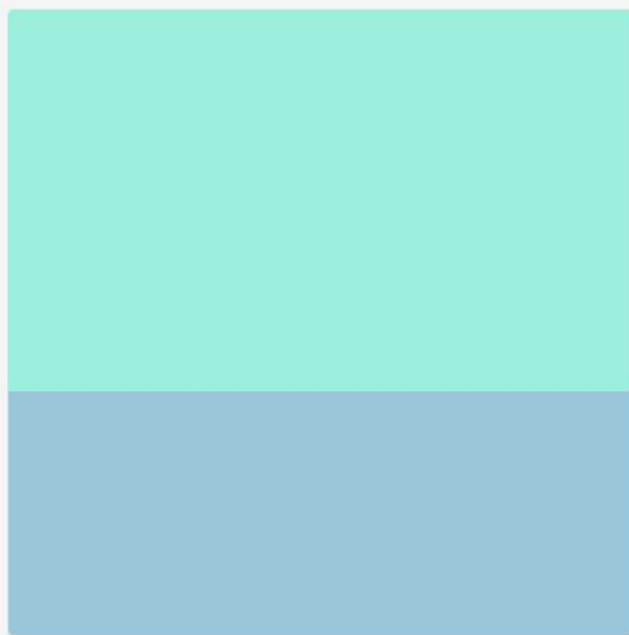
GauGAN Beta

Building

Ground

Landscape

Plant



Upload Segmentation Map ⓘ

Browse...

No file selected.

Upload

Upload Landscape Image ⓘ

Browse...

No file selected.

Upload

Upload Custom Style Filter ⓘ

Browse...

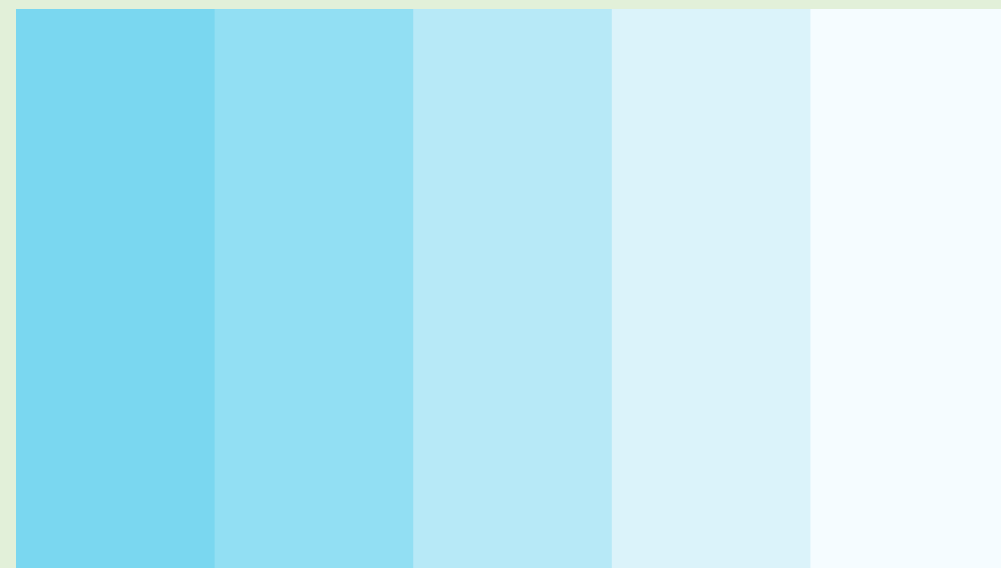
No file selected.

Upload

Vyhledávačka



Najděte obrázky z digitalizovaných sbírek švédských paměťových institucí jejichž dominantní barvou je modrá.



Vyhledávačka



Najděte obrázky z digitalizovaných sbírek švédských paměťových institucí jejichž dominantní barvou je modrá.

