

## Image search: kde slova nestačí

Petra Budíková, ÚVT a FI MU

Lidová moudrost pohádek praví, že když chtěl princ požádat o ruku krásnou princeznu, neposlal jí zdvořilou žádost, ale svůj obraz. Na něm princezna uvidí, jak je pohledný a udatný,



a hned se do něj zamiluje. O něco později i tvůrci reklam objevili, že obrázek řekne víc než tisíc slov<sup>1</sup>. Opakovaně tedy zjišťujeme, že nejen textové, ale i obrazové informace jsou cenné – jakožto i jiná multimédia, princové také zpívali pod balkónem...

Textové i multimediální informace jsou dnes všude kolem nás, většinou v digitální podobě a v obrovských množstvích. Aby bylo možné najít to, co je pro nás cenné, potřebujeme nějaké nástroje, které nám umožní velké objemy dat efektivně spravovat a prohledávat. Zatímco dříve byla pozornost zaměřena na vyhledávání v jednoduchých atributových datech a později v rozsáhlých textových kolekcích, v poslední době se intenzivně rozvíjí také metody pro vyhledávání v multimédiích, přičemž nejčastější aplikací je právě vyhledávání obrázků. V tomto článku se proto zkusíme porozhlédnout po existujících systémech a směrech současného vývoje.

### 1 Jak na to

Základním problémem, který je vždy potřeba vyřešit pro práci se složitými daty, je způsob reprezentace jednotlivých objektů. Ta musí odpovídat potřebám uživatele. Člověk se nezajímá o každý pixel obrázku, má spíše nějakou celkovou představu, jak má ten správný obrázek (dokument, písnička) vypadat. Tuto představu může vyjádřit například slovním popisem či pomocí nějakého vzorového obrázku. V každém případě je zadání dotazu jakýmsi nepřesným popisem hledaného

ideálu, ve kterém má počítač rozeznat důležité vlastnosti a pokusit se k němu najít co nejpodobnější objekty.

Jedním z hlavních úkolů současného výzkumu je proto právě snaha o zachycení lidského vnímání podobnosti. Ačkoli nikdo přesně neví, jak funguje lidský mozek při vyhodnocování podobnosti obrázků, určitě vnímáme jak vizuální vlastnosti (barvy, tvary, velikosti), tak obsah obrázku (co je zobrazeno, jakým dojmem to na nás působí). Tyto vlastnosti je proto vhodné při práci s obrazem využívat.

Z hlediska automatického vyhodnocování podobnosti je poměrně dobře zvládnuté rozpoznávání vizuálních vlastností obrazu. Existuje množství algoritmů pro extrakci barev a tvarů, pomocí nichž je možné zachytit jak vlastnosti celého obrazu (získané údaje se pak označují jako globální vizuální deskriptory), tak vlastnosti jednotlivých částí obrazu (lokální deskriptory). Od roku 2002 se postupně vyvíjí ISO standard MPEG-7<sup>2</sup>, který mimo jiné definuje sadu globálních vizuálních deskriptorů pro popis barev, tvarů, textur, umístění a rozpoznávání obličejů. Součástí standardu jsou také algoritmy pro výpočet podobnosti mezi deskriptory. Nejpoužívanějšími lokálními deskriptory jsou algoritmy SIFT<sup>3</sup> a SURF<sup>4</sup>, které popisují vlastnosti významných bodů v obraze.

Mnohem náročnější pro automatické zpracování je rozpoznání obsahu – sémantiky obrazu, kterou nelze jednoduše odvodit z rozpoznávaných barev a tvarů. Vizuálně velmi podobná žlutá koule může být jednou tenisový míč, jindy slunce a jindy pampeliška. Lidé rozpoznávají obsah na základě své zkušenosti s reálným světem, počítači je potřeba tuto zkušenost zprostředkovat a učit je spojovat sémantické koncepty (slova, slovní spojení) s vizuální podobou. Aby tedy vůbec bylo možné sémantiku do vyhledávání zahrnout, je třeba nejprve poskytnout počítači dostatečně velkou kolekci dat s anotacemi. V případě

<sup>2</sup><http://en.wikipedia.org/wiki/MPEG-7>

<sup>3</sup>Scale-invariant feature transform, [http://en.wikipedia.org/wiki/Scale-invariant\\_feature\\_transform](http://en.wikipedia.org/wiki/Scale-invariant_feature_transform)

<sup>4</sup>Speeded Up Robust Features, <http://en.wikipedia.org/wiki/SURF>

<sup>1</sup>Fráze „A picture is worth a thousand words“ je připisována Fred R. Barnardovi, který ji použil při propagaci obrázkových reklam. Viz [http://en.wikipedia.org/wiki/A\\_picture\\_is\\_worth\\_a\\_thousand\\_words](http://en.wikipedia.org/wiki/A_picture_is_worth_a_thousand_words)

specializované medicínské databáze je ještě reálné vytvořit dostatečně reprezentativní vzorovou množinu se systematickými popisy, v případě obecného vyhledávání mezi obrázky na webu to však možné není. V takovém případě tedy nelze využít automatické rozpoznávání sémantiky a nezbyvá než spolehnout se na anotace, které k obrázkům poskytli uživatelé, případně se snažit získat klíčová slova z okolního textu a podobně.

## 2 Implementace

Odhlédněme nyní na chvíli od procesu získávání dat a předpokládejme, že jsme schopni k obrázku získat jak slovní popis obsahu, tak vizuální deskriptory, což jsou typicky vektory s vysokou dimenzí. Abychom získali fungující vyhledávací systém, musíme mít ještě algoritmy, které jsou schopny s touto dvojí reprezentací efektivně pracovat. Zatímco algoritmy pro vyhledávání v textech jsou dobře známé a vyladěné, vysokodimenzionální vektory jsou stále považovány za těžko indexovatelné (mnoho navržených indexových struktur má problémy s přílišným dělením prohledávaného prostoru při vysokém počtu dimenzí). Je tedy logické, že první komerční obrázkové vyhledávače vizuální hledání zanedbaly a pracovaly pouze s textovou informací. V současné době funguje většina komerčních systémů pro obrázkové hledání na základě textů, existují však také i systémy, které využívají skutečné podobnostní hledání na základě vizuálních vlastností obrázků. V dalších odstavcích si představíme dva systémy, které stojí na opačných pólech spektra přístupů, a porovnáme jejich schopnosti.

### 2.1 Google Image Search

Asi nejznámějším systémem pro vyhledávání obrázků na webu je Google Image Search<sup>5</sup>. Do vyhledávacího políčka stačí napsat výraz, který nás zajímá, a Google nám nabídne několik obrázků náhledů, které mají daná slova ve svých metadatech, tedy v názvu, adrese či titulku. Vzhledem k množství dat, která má Google zaindexovaná, tento způsob vyhledávání postačuje k získání dostatečného počtu relevantních výsledků,

<sup>5</sup><http://images.google.com/>

ačkoli jsou z vyhledávání vyřazeny všechny obrázky, které sice mohou mít správný obsah, ale nemají ten správný popis. Naopak také můžeme získat výsledky, které se zamýšleným dotazem vůbec nesoúvisí. Výsledky jsou uspořádány podle běžného PageRanku, který Google používá pro určování důležitosti stránek.

Obrázek 1 ukazuje několik výsledků vyhledávání pomocí Google Image Search. V prvním případě je výsledek velmi dobrý, v dalších je vidět problém s různými významy slov – pokud uživatel hledal boxera-zápasníka, bude se muset pokusit to vyjádřit jiným slovem. Poslední příklad se týká vyhledávání pojmu, který je součástí názvu populárního filmu.

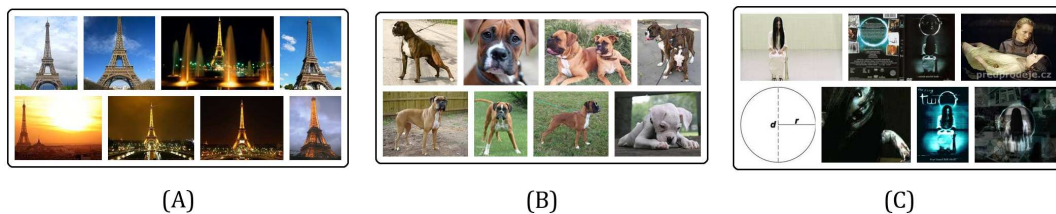
### 2.2 MUFIN Image Search

Vyhledávání pomocí textu může fungovat jen v případě, kdy máme k obrázkům nějaké textové informace. Existuje ale mnoho aplikací, kdy textových informací není dostatek nebo nejsou dostatečně kvalitní. V takovém případě je jedinou možností hledání podle obsahu dat, tedy podle vizuálních deskriptorů v případě hledání obrázků. Deskriptory je možné získat z každého obrázku a jsou tedy použitelné univerzálně. Pro jejich praktické nasazení je ale potřeba mít efektivní implementaci indexu, který umožňuje vyhledávání ve vysokodimenzionálních vektorových prostorech.

Pro takové účely je v posledních letech vyvíjen na Fakultě informatiky týmem prof. Zezuly systém MUFIN – *Multi-Feature Indexing Network*. Tento obecný vyhledávací systém umožňuje podobnostní vyhledávání v téměř libovolných datech – jedinou podmínkou je, že pro každé dva objekty musí existovat funkce, která vyjadřuje jejich vzdálenost neboli nepodobnost a tato funkce musí být metrická (tj. reflexivní, symetrická a splňující trojúhelníkovou nerovnost; detaily viz [1]). Díky využití distribuované vyhledávací sítě a paralelnímu zpracování poddotazů umožňuje MUFIN interaktivní vyhledávání ve velmi rozsáhlých datových kolekcích.

Pro vyhledávání v obrázcích slouží prototypová aplikace MUFIN Image Search<sup>6</sup> [2], která umož-

<sup>6</sup><http://mufin.fi.muni.cz/imgsearch>



Obrázek 1: Výsledky Google Image Search pro dotazy: a) Eiffelova věž, b) boxer, c) kruh.

ňuje hledání v přibližně sto milionech obrázků z webové galerie Flickr<sup>7</sup>. Každý obrázek je popsán pěti globálními deskriptory dle standardu MPEG-7. Dotaz se zadává pomocí vzorového obrázku, který je možno vybrat náhodným procházením kolekce, zadáním klíčových slov či vložením vlastního obrázku. Výsledkem hledání jsou potom obrázky vizuálně podobné vzoru.

Obrázek 2 ukazuje dva možné výsledky hledání pomocí MUFIN Image Search. Ve výsledku je vždy vyznačen vzorový obrázek, k němuž se hledaly podobné objekty. V obou případech je vidět, že výsledky skutečně jsou vzoru podobné, ovšem pouze v prvním případě všechny výsledky zobrazují to, co uživatel pravděpodobně hledal. V druhém případě se potvrzuje, že čistě vizuální informace pro vyhledávání často nestačí, neboť není dostatečná pro zachycení sémantiky obrázku.

### 3 Trendy poslední doby: kombinované vyhledávání

V předchozích odstavcích jsme ukázali, že jak přístup založený na textových popisech, tak vizuální vyhledávání mají své slabé stránky, které neumožňují získání optimálního výsledku. Jednoduchým východiskem je oba přístupy zkombinovat a s využitím dvou různých pohledů na podobnost dosáhnout lepšího výsledku, který bude také odpovídat lidskému vnímání podobnosti. Touto cestou se nyní také ubírají vyhledávací systémy obou představených typů. Opět však narážíme na otázku efektivní implementace – čím více informací je pro vyhledávání využíváno, tím více dat je třeba spravovat a procházet. Proto se často používá dvoufázové vyhodnocování dotazu, kdy se v první fázi prohledá celá databáze

pomocí jednoho kritéria a získá se množina kandidátních objektů. Tyto jsou potom přeuspořádány vzhledem ke druhému kritériu podobnosti a takto seřazený výsledek je zobrazen uživateli.

#### 3.1 Google: Vizuálně podobné

Právě takovouto funkcionalitu nabízí Google pod volbou „Podobné“ (v originálu „Similar“), která je zobrazena u většiny obrázků, najedete-li na ně myší. V případě Google je první fáze hledání samozřejmě založená na textu, pomocí něhož se získá přibližně 1000 nejlepších výsledků. Z nich se extrahují lokální vizuální deskriptory a spočítá se podobnost každé dvojice obrázků. Tyto informace jsou pak využity pro seřazení obrázků podle vizuální podobnosti. Obrázek 3 ukazuje rozdíl mezi výsledkem běžného obrázkového hledání Google a výsledkem hledání vizuálně podobných obrázků k danému vzoru.

Je však třeba podotknout, že kvůli rychlosti Google vizuálně podobné obrázky nepočítá ve chvíli, kdy uživatel spustí dotaz, jak bychom očekávali. Vyhodnocování podobnosti pomocí lokálních deskriptorů je výpočetně náročné, a tak je k velkému množství obrázků množina podobných objektů spočítaná předem. Proto je také možno narazit na příklady, kdy volba „Podobné“ není zobrazena – Google postupně výsledky předpočítává podle popularity daného dotazu, u málo častých výsledky zatím chybí. Tento přístup je také neflexibilní – množina podobných obrázků je fixovaná do doby, dokud Google podobnost znovu nevyhodnotí. Pro časté dotazy s mnoha relevantními objekty jsou však výsledky dobré.

#### 3.2 MUFIN ranking

Také vyhledávač MUFIN jde s dobou a rozšiřuje své obzory o další kritéria pro vyhodnocování

<sup>7</sup><http://www.flickr.com/>



(A)



(B)

Obrázek 2: MUFIN Image Search



(A)



(B)

Obrázek 3: a) Google Image Search, b) Google - „Podobné“

podobnosti. V první řadě se samozřejmě nabízí textová informace, lze však pracovat i s dalšími metadaty, která mohou být k obrázku k dispozici – doba pořízení, GPS souřadnice, popularita objektu mezi uživateli apod.

Podobně jako Google, také MUFIN umí pracovat s více kritérii podobnosti v dvoufázovém vyhledávacím modelu. V prvním kole vyhodnocování se získá  $N$  vizuálně nejpodobnějších objektů a s touto množinou se potom dále pracuje. MUFIN ranking demo<sup>8</sup> umožňuje vyzkoušet si několik způsobů přeuspořádání, které jsou zaměřené zejména na co nejlepší využití textových metadat [3]. Kromě uspořádání výsledné množiny dle klíčových slov hledaného objektu si uživatel sám může vybrat, která slova jsou pro něj relevantní. Další možností je automatické získání nejdůležitějších slov na základě frekvence jejich výskytu v popisech daných  $N$  vizuálně nejpodobnějších objektů. Obrázek 4 ukazuje, jak může zohlednění klíčových slov vylepšit výsledek podobnostního dotazu – z výsledku jsou odfiltrovány objekty, které jsou vizuálně podobné, ale nezobrazují hledaný předmět.

<sup>8</sup><http://mufin.fi.muni.cz/ranking>

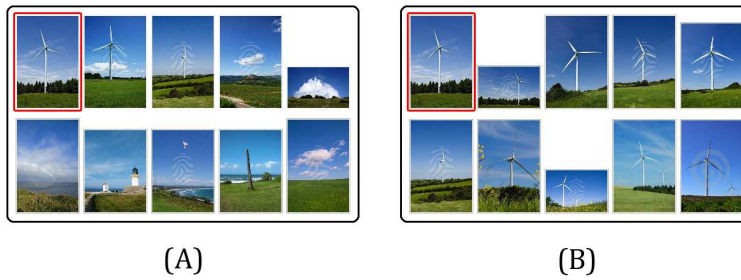
## 4 Závěr

Dva představené vyhledávací systémy samozřejmě nejsou jediné, které lze pro hledání obrázků použít. Mohli bychom mluvit o mnoha dalších<sup>9</sup>, které se nějakým způsobem snaží pracovat s obrázky a využívat text, vizuální deskriptory či obojí k vyhledávání podobných objektů. Různé přístupy jsou vhodné pro různé aplikace, ideální systém pro všechny situace zatím neexistuje a pravděpodobně existovat ani nemůže. Samotný pojem podobnosti, se kterým stále pracujeme, je těžko definovatelný, jeho vnímání je individuální a může se lišit v různých situacích. Můžeme se proto jen snažit o co nejlepší přiblížení a co nejefektivnější vyhodnocování.

Vyhledávače Google a MUFIN ukazují dva zcela odlišné přístupy. Zatímco komerční Google spoléhá především na obrovské množství dat<sup>10</sup>, ve kterých se najde dostatek obrázků s příslušným popisem, MUFIN se snaží vytvářet obecnější řešení, použitelné pro různé situace a různé typy

<sup>9</sup>Neúplný přehled systémů, které nějakým způsobem implementují vyhledávání na základě vizuální podobnosti, lze najít na [http://en.wikipedia.org/wiki/List\\_of\\_CBIR\\_engines](http://en.wikipedia.org/wiki/List_of_CBIR_engines)

<sup>10</sup>Dle tiskové zprávy z 7.2.2005 indexoval Google v té době přes miliardu obrázků; <http://www.google.com/press/annnc/onebox.html>



Obrázek 4: a) MUFIN Image Search, b) totéž s dodatečným přeuspořádáním výsledku podle klíčových slov

dat. V obou případech stejně jako u dalších systémů se současný vývoj zaměřuje především na efektivní implementaci složitějších vyhodnocovacích kritérií. V případě MUFINu se chystáme v blízké době provést experimentální vyhodnocení a porovnání různých přístupů ke kombinování textové a vizuální informace. Dalším z úkolů pro podobnostní vyhledávání je pak umožnit flexibilní vyhledávání, které si uživatel může přizpůsobit svým potřebám.

## Literatura

- [1] P. Zezula, G. Amato, V. Dohnal, M. Batko. *Similarity Search: The Metric Space Approach*, volume 32 of *Advances in Database Systems*. Springer-Verlag, 2006.
- [2] D. Novak, M. Batko, P. Zezula. *Web-scale system for image similarity search: When the dreams are coming true*. In *Proc. of CBMI '08*, page 8. IEEE, 2008.
- [3] P. Budikova, M. Batko, P. Zezula. *Similarity Query Postprocessing by Ranking*. In *Proc. of AMR '10*. Linz, 2010. □