

ZPŘÍSTUPNĚNÍ OBRÁZKŮ FORMOU DIALOGU PRO UŽIVATELE SE ZRAKOVÝM POSTIŽENÍM

Ivan KOPEČEK, Radek OŠLEJŠEK, Jaromír PLHÁK

Fakulta informatiky, Masarykova univerzita,
Botanická 68a, 602 00 Brno, Česká Republika
{kopecek, oslejsek, xplhak}@fi.muni.cz

***Anotace:** Přístupnost počítačové grafiky lze v současnosti považovat za jedno z nejdůležitějších témat v oblasti asistivních technologií pro osoby se zrakovým postižením. Aplikace v této oblasti se doposud orientovaly zejména na využití hmatových zařízení nebo na převod grafických objektů na neverbální zvuky. V této práci popisujeme nový přístup ke zkoumání grafických dat. Koncept komunikativních obrázků umožňuje poskytnout uživatelům informace o grafických datech prostřednictvím komunikace v přirozeném jazyce. Sémantika vizuálních dat je popsána ontologiemi, které představují formálně definované a dobře strukturované báze dat.*

Prototyp aplikace byl implementován v rámci projektu GATE („Graphics Accessible To Everyone“) a následně testován uživateli se zrakovým postižením. V rámci přednášky budou předvedeny jak základní možnosti dialogového rozhraní, tak i výsledky testování.

Úvod

Současné technologie nám umožňují automaticky připojit k obrázku mnoho zajímavých informací, například datum a čas získání snímku nebo geolokační souřadnice. Tyto informace jsou typicky uloženy přímo ve formátu obrázku a mohou být využity pro jeho klasifikaci či pro získání sémantiky zobrazených objektů, viz např. [1,2,3,4].

Mnohem více zajímavých informací však obvykle zůstává nedostupných. Představme si fotografii z dovolené před deseti lety: žena uprostřed je moje manželka, ale kdo je ten muž stojící za ní? Je zřejmě pořízena někde v Alpách, ale v jakém středisku? A jak se nazývá vrchol v pozadí? Takové informace jsou prakticky nedostupné. Nicméně pomocí geolokačních souřadnic a elektronické mapy lze stanovit, kde byla fotografie pořízena. Rozpoznávání obličejů [5,6,7] může pomoci odhalit, kdo je ten neznámý muž na fotografii. Při znalosti orientace obrazu je možné odhadnout, který vrchol se nachází v pozadí.

Současné algoritmy pro automatické rozpoznávání obrazu bohužel nejsou schopny přesně popsat analyzovaný obrázek. Jinak bychom byli schopni k obrázku připojit spoustu zajímavých informací za použití znalostí získaných z webu či velkých elektronických databází. To by mohlo vést k situaci, kdy by obrázek byl reprezentován komplexní ontologií obsahující nejen samotná grafická data, ale i další relevantní informace. Vzhledem k tomu, že rozsah takových informací by mohl být obrovský, je nutné umožnit uživatelům získávat informace pomocí dialogové komunikace.

Proto jsme si stanovili za cíl umožnit obrázkům komunikovat s uživateli. Aktuálně sice nejsme schopni automaticky získat dostatek informací o obrázku, nicméně komunikace s uživatelem nám může pomoci překonat tuto překážku, jelikož ji můžeme využít i pro získávání informací o obrázku samotném.

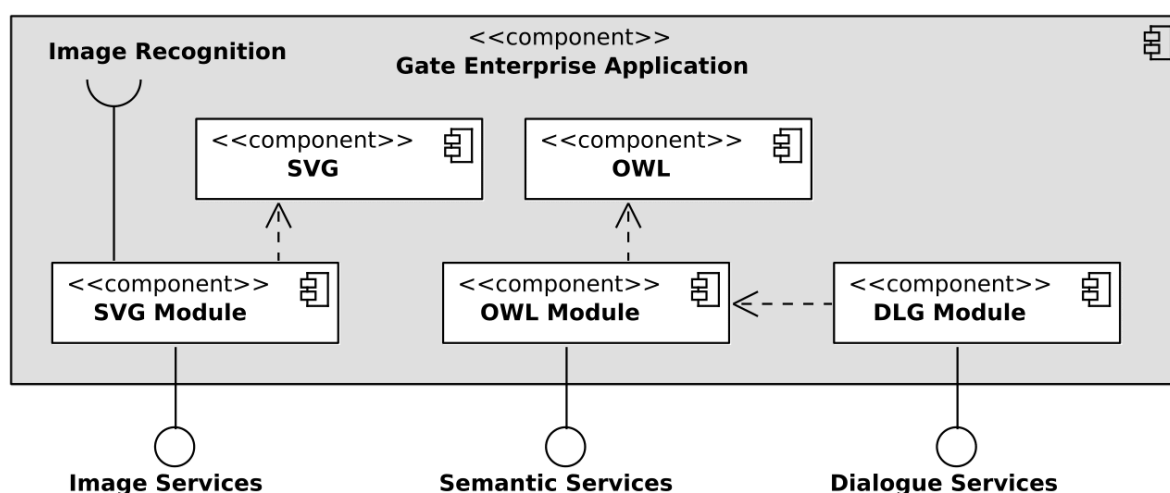
Komunikativní obrázek [8] představuje grafický objekt integrovaný s dialogovým rozhraním a propojený se znalostní databází, jenž uchovává sémantiku vyobrazených objektů. Každý komunikativní obrázek se skládá ze tří datových struktur: (a) grafických údajů, (b) identifikace objektů v obrázku a (c) sémantických dat o objektech. Pro zakódování těchto dat v jediném souboru využíváme SVG formát [9]. Sémantika objektů je uložena pomocí OWL ontologií [10]. Formální ontologie představují klíčový aspekt komunikativních obrázků, protože umožňují definovat a strukturovat slovník, který je sdílen mezi různými obrázky s podobným obsahem. Navíc představují vhodný formalismus pro automatické získávání informací a stroje generované dialogy.

Rozhraní mezi přirozeným jazykem a formalizovanými ontologiemi zajišťuje modul, který transformuje přirozený jazyk do odpovídajícího formálního schématu. Typicky se omezíme jen na malé fragmenty přirozeného jazyka, takže převod může být založen na využití poměrně jednoduchých gramatik v kombinaci s rámci a standardními technikami pro řešení nedorozumění. Například, otázka „*Jak daleko je to od hotelu na nejbližší pláž?*“ je analyzována pomocí rámce „*Jak daleko je to od SLOT1 na SLOT2?*“. Systém očekává, že SLOT1 i SLOT2 jsou vyplněny konkrétními položkami z ontologie. Hlavní principy a podrobnosti o řízení dialogu byly popsány v [11].

Vzhledem k tomu, že uživatelé mohou komunikovat s obrázky uloženými na Internetu nebo lokálně, je nezbytné poskytnout služby pro automatický převod obrázků z běžných formátů do komunikativní formy. Potřebná infrastruktura je založená na cloudovém řešení s tenkými klienty a sdíleným serverem. Úloha klientů je zvládnout interakci uživatele s obrázkem (kliknutí na myši, psaním na klávesnici, výstup z rozpoznávání hlasu) a přeměrovat ji na vzdálený server, který obsahuje jádro aplikace. Server je zodpovědný za sémantickou analýzu, zjišťování nových informací, ukládání, správu a sdílení znalostí, a vedení dialogu.

Implementace prototypu

Celý koncept komunikativních obrázků je implementován v rámci projektu GATE [12]. Server je navržen jako modulární Java EE aplikace založená na komponentách, která poskytuje relačně orientované vzdálené služby dostupné přes SOAP a restful API. Je složena ze tří modulů, které jsou znázorněny pomocí UML diagramu na Obrázku 1.



Obr.1: Architektura GATE serveru

SVG modul umožňuje uživateli nahrát obrázek a zkoumat grafické údaje pomocí procházení SVG DOM stromu. Lze nahrát buď obrázky ve formátu SVG nebo rastrové obrázky. Rastrové obrázky jsou automaticky vloženy do SVG formátu. Navržené rozhraní pro rozpoznávání obrazu se používá k rozšíření schopností modulu o automatické rozpoznávání objektů. To je velmi užitečné zejména v případě obrázků, které neobsahují žádná sémantická data. Algoritmy pro rozpoznávání obrazu jsou poskytovány externí komponentou a propojeny s SVG modulem. Lze navíc využít externí moduly pro rozpoznání obličeje [13,14,15] nebo podobnostní vyhledávání ve velkých kolekcích obrázků [16,17,18].

OWL modul poskytuje služby související se sémantikou dat. Tento modul spolupracuje s SVG modulem na získání anotací, které jsou uloženy přímo v obrázku a jejich propojení se znalostní databází. Anotace i znalostní databáze jsou implementovány pomocí OWL. Dostupné služby zahrnují správu ontologie, procházení OWL DOM stromu a filtrování informací.

DLG modul je zodpovědný za dialogový subsystém, tj. za analýzu otázek v přirozeném jazyce, porozumění dotazu a generování odpovědi. Tento modul úzce spolupracuje s OWL modulem při analýze významu slov. V současnosti je implementována zjednodušená verze tohoto modulu. Tato verze podporuje otázky v jazyce What-Where Language (WWL, viz [12]), které mají formát „kde je co (kdo)“ a „co (kdo) je kde“ případně „co (kdo) daný objekt je“. Kromě toho, modul může být nakonfigurován pro promluvy nad vybranou doménou, které umožňují uživateli zeptat se na otázky týkající se konkrétního obrázku, například otázky týkající se rodinných vztahů.

Komunikativní obrázky v asistivních technologiích

Dialog s obrázkem v přirozeném jazyce zpřístupňuje grafická data zejména uživatelům s postižením zraku. Takoví uživatelé nejsou omezeni jen na jednoduché shrnutí obsahu obrázku, ale mohou vést komplexní dialog o grafických datech. Díky tomu, že data jsou strukturována a souvisejí s různými částmi, objekty a aspekty obrázku může dialog vést k přirozenější a uspokojivější zkušenosti uživatelů se systémem.

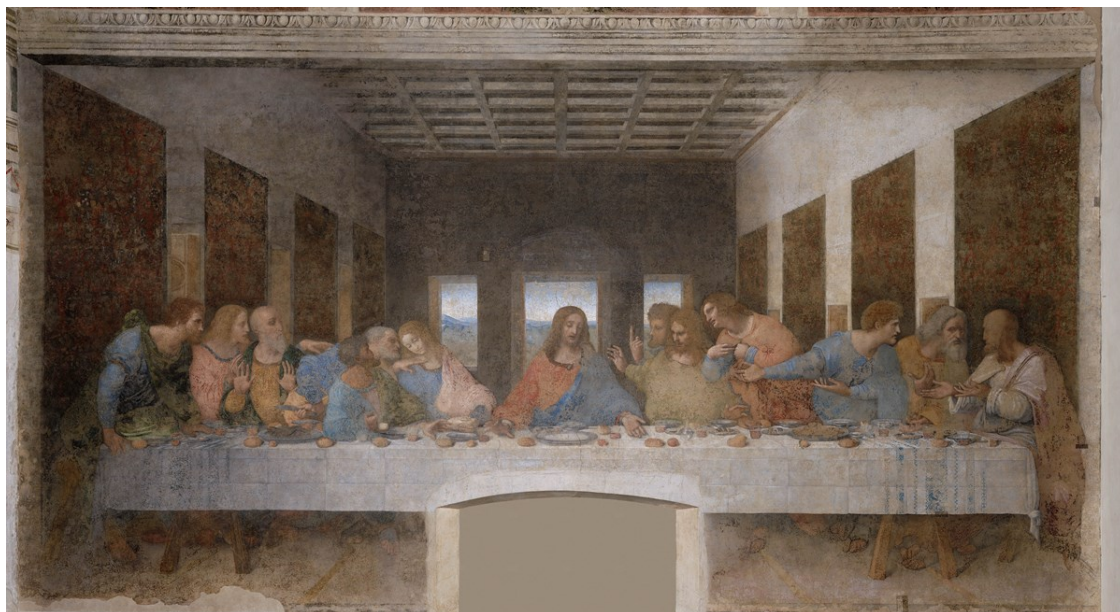
Využití mluvených dialogů je vhodné také pro zlepšení přístupnosti zkoumání grafických dat u ostatních uživatelů se specifickými nároky. Starší lidé a lidé s nižší počítačovou gramotností mohou mít prospěch ze snadného přístupu k informacím o obrázku, které poskytuje dialogový systém. Požadovaná informace může být totiž získána na základě jednoduchého požadavku v přirozeném jazyce. To je užitečné i pro uživatele s poruchou motoriky, osob s dyslexií a s některými dalšími kognitivními poruchami.

Využití cloudových technologií navíc umožňuje integraci komunikativních obrázků do sociálních sítí. Tyto technologie podporují snadné sdílení informací a spolupráci on-line uživatelů, což pomáhá decentralizovaně zpracovávat znalosti. Pokud jeden uživatel publikuje historická fakta o nějaké pamětihodnosti, mohou být tyto informace využity dalšími uživateli. Pro komunikativní obrázky představuje tento druh crowdsourcingu efektivní způsob budování a rozšiřování znalostní báze s dlouhodobou perspektivou učinit grafická data dostupnější pro všechny.

Funkce využívající informace ve formě ontologií nejen naplňují uživatelskou potřebu získat informace, ale také jim pomáhají trénovat vlastní paměť, vnímání a další kognitivní funkce. Uživatelé s neurologickými nebo kognitivními poruchami mohou prohlížet rodinné fotografie, přičemž jim systém může připomínat věk a jména lidí na fotografiích, jejich narozeniny, jména domácích mazlíčků, čas a příležitost, kdy byl snímek pořízen, atd. Z těchto důvodů mohou hrát komunikativní obrázky nezanedbatelnou roli v psychosociálním rozvoji.

Experiment

Pro hodnocení využitelnosti konceptu komunikativních obrázků jsme připravili jednoduchý experiment, v rámci nějž měli uživatelé za cíl zkoumat daný obrázek pomocí dotazů v anglickém jazyce pomocí WWL. Jako referenční obrázek jsme si vybrali obraz Poslední večeře od Leonarda da Vinciho, který je znázorněn na Obrázku 2. Tento obraz byl precizně anotován a následně zkoumán uživateli prostřednictvím GATE systému.



Obr.2: Anotovaný obrázek: Poslední večeře od Leonarda da Vinciho

Da Vinciho Poslední večeře je jedním ze světově nejznámějších obrazů zachycující poslední jídlo Ježíše a dvanácti apoštolů v Jeruzalémě před jeho ukřižováním. Obraz se skládá z několika dominantních objektů: Ježíše, 12 apoštolů a stolu. Tyto objekty byly propojeny skrze sémantická data. Anotace obsahovala historická fakta o Ježíši a jeho apoštolech, jejich pozici na obrázku, vzájemnou polohu postav, dominantní barvy oblečení,

činnosti postav (např. sezení, mluvení, atp.) a výrazy jejich obličejů (např. zvědavý, vzteklý, atd.). Anotace významných osobností (Ježíše, Jidáše a Petra) šla do ještě větších detailů a byla doplněna například o pozici rukou. Kromě anotace obsahu obrázku byla do obrázku vložena fakta o obraze samotném, například kdy byl obraz dokončen, jeho velikost, atd.

Experimentu se zúčastnilo několik bývalých i současných studentů Masarykovy univerzity se zrakovým postižením. Před testováním byli informováni o základních principech konceptu komunikativních obrázků. Účastníci byli požádáni, aby při experimentu mluvili nahlas kvůli snadnějšímu pochopení způsobu jejich interakce s obrázkem. To nám umožnilo zaznamenat jejich záměry, očekávání nebo věci, které je překvapily. Po ukončení testování účastníci vyplnili dotazník s jejich subjektivním hodnocením.

Při experimentech jsme použili dva režimy - s vedením a bez vedení. Účastníci bez vedení neměli tušení, co je na obrázku zobrazeno a libovolně komunikovali s obrázkem. Naopak účastníci s vedením byli pověřeni konkrétními úkoly, například měli zjistit, kdo sedí vedle Ježíše.

Žádný z účastníků neměl potíže s ovládním aplikace pomocí odčítače obrazovky. Celkový dojem účastníků byl velmi pozitivní a většina z nich byla schopna si představit přibližnou polohu objektů na obrázku. Práce s aplikací pro ně byla snadná a velmi příjemná. Obvykle projevíli velký zájem o komunikaci s jinými obrázky, zejména preferovali fotografie svých přátel.

Účastníci částečně kritizovali chybějící podporu otázek, které nepatří do oblasti WWL („Co má daný člověk na sobě?“). Účastníci také nebyli spokojeni s množstvím informací o objektech a nedostatečnou úrovní detailů. Na druhou stranu jednoznačně souhlasili s tezí, že dialogová komunikace je efektivní způsob, jak získat informace o obrázcích a zobrazených objektech.

Závěr a budoucí práce

V této práci jsme nastínili základní principy konceptu komunikativních obrázků a základní strukturu implementovaného systému. Cílem provedeného experimentu bylo ověřit životaschopnost, využitelnost a užitečnost tohoto konceptu. Aktuální implementace je velmi zjednodušená a koncept komunikativních obrázků má stále mnoho otevřených problémů. V současné době musíme například ručně připravit a odladit gramatiky pro dialogový subsystém namísto automatického generování dialogových strategií z vnitřní struktury ontologie. Další výzvy představují neustálé zpřesňování a rozšiřování znalostí báze a jako automatické učení z historie jednotlivých dialogů. Předběžné výsledky ukazují, že tento přístup slibuje zajímavé využití v mnoha aplikačních oblastech, například v asistivních technologiích, e-learningu a efektivní správě velkých sbírek fotografií.

Literatura

- [1] Sandnes, F.: Where was that photo taken? Deriving geographical information from image collections based on temporal exposure attributes. *Multimedia Systems*, pp. 309-318, 2010.
- [2] Boutell, M., Luo, J.: Photo classification by integrating image content and camera metadata. *Proceedings of the 17th International Conference on Pattern Recognition*, pp. 901-904, 2004.
- [3] Yuan, J., Luo, J., Wu, Y.: Mining Compositional Features From GPS and Visual Cues for Event Recognition in Photo Collections. *IEEE Trans. on Multimedia*, pp. 705-716, 2010.
- [4] Ráček, J., Ludík, T.: Development of ontology for support of crisis management processes. *Informační technologie pro praxi*, pp. 106-111, 2008.
- [5] Bartlett, M., Movellan, J., Sejnowski, T.: Face recognition by independent component analysis. *IEEE Transactions on Neural Networks*, pp. 1450-1464, 2002.
- [6] Haddadnia, J., Ahmadi, M.: N-feature neural network human face recognition. *Image and Vision Computing*, pp. 1071-1082, 2004.
- [7] Rowley, H., Baluja, S., Kanade, T.: Neural network-based face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 23-38, 1998.
- [8] Kopeček, I., Oslejšek, R.: Communicative images. *Smart Graphics*, pp. 163-173, 2011.

- [9] Dahlström, E., et al.: Scalable vector graphics 1.1, 2nd edn., 2011.
- [10] Lacy, L.W.: OWL : Representing information using the Web Ontology Language. Trafford Publishing, 2005.
- [11] Kopeček, I., Ošlejšek, R., Plhák, J.: Dialogue management in communicative images. Text, Speech and Dialogue - Students' section, pp. 9-13, 2011.
- [12] Kopeček, I., Ošlejšek, R.: Gate to accessibility of computer graphics. Computers Helping People with Special Needs, pp. 295-302, 2008.
- [13] Bartlett, M., Movellan, J.R., Sejnowski, T.: Face recognition by independent component analysis. Neural Networks, pp. 1450-1464, 2002.
- [14] Rowley, H., Baluja, S., Kanade, T.: Neural network-based face detection. Computer Vision and Pattern Recognition, pp. 203-208, 1996.
- [15] Haddadnia, J., Ahmadi, M.: N-feature neural network human face recognition. Image and Vision Computing, pp. 1071-1082, 2004.
- [16] Jaffè, A., Naaman, M., Tassa, T., Davis, M.: Generating summaries and visualization for large collections of geo-referenced photographs. Proc. of ACM Int. Workshop on Multimedia Information Retrieval, pp. 89-98, 2006.
- [17] Abbasi, R., et al.: Exploiting flickr tags and groups for finding landmark photos. Advances in Information Retrieval, pp. 654-661, 2009.
- [18] Müller, H., Michoux, N., Bandon, D., Geissbuhler, A.: A review of content-based image retrieval systems in medical applicationsclinical benefits and future directions. International Journal of Medical Informatics 73, pp. 1-23, 2004.