

Masarykova univerzita
Filozofická fakulta

POČÍTAČOVÉ VIDĚNÍ
seminární práce

V Brně, dne 29. 11. 2014

Vypracovala
Adriana Babincová

Obsah

1 Historie	3
2 Obor počítačového vidění	3
2.1 Snímání obrazu	3
2.2 Předzpracování	4
2.3 Segmentace obrazu	4
2.4 Popis objektů	4
2.5 Klasifikace (rozpoznávání) objektů	4
3 Vlastní rozpoznávání – metody	4
3.1 Optické rozpoznávání znaků (OCR)	5
3.2 2D čtení kódů	5
3.3 Rozpoznávání tváře (face detection)	5
3.4 Metoda DeepFace	6
3.5 Využití - biometrie	6
4 Metody analýza pohybu	6
4.1 Rozdílová metoda	6
4.2 Detekce pohybu pomocí kontury	6
4.3 Optický tok	6
5 Rekonstrukce 3D scény	7
5.1 Konverze z 2D obrazu na 3D scénu	7
5.2 3D scanner / 3D čtečka	7
6 Robotické rozpoznávání objektů	7
Zdroje	8

POČÍTAČOVÉ VIDĚNÍ *(computer vision)*

Obecně můžeme říci, že počítačové vidění je snaha napodobit lidské vidění snímáním obrazu elektronickými prostředky a porozumění jejich obsahu počítačovým zpracováním. Klíčovou úlohou počítačového vidění je rozpoznání objektu, kde výsledný obraz je výsledkem perspektivního zobrazení části trojrozměrného do dvourozměrného prostředí.

1 Historie

Se zájmem o počítačové vidění se setkáváme od 60. let minulého století. První software schopný rozpoznávání objektů se jmenoval DIDAK. Vytvořil jej roku 1980 Joachim Wiesel, německý vedoucí pracovník v Ústavu fotogrammetrie. Vstupem byly uměle vytvořené matice a malé obrázky, které byly za pomoci algoritmu rozpoznány. Roku 1977 vznikla Konference o počítačovém vidění a rozpoznávání (CVPR), funguje i v současnosti a spolu s konferencí ICCV patří v rámci počítačového vidění ke konferencím nejvyšší úrovně. Každoročně udělují cenný za nejlepší rozpoznání objektů v obraze, např. loňský rok vyhrála cenu skupina vědců z Washingtonu, která vymyslela software, který dokázal detekovat 100 000 objektů na jednom počítači.

2 Obor počítačového vidění

Podle složitosti dělíme počítačové vidění na dvě úrovně. Nižší úroveň počítačového zpracování chápeme jako pouhé zpracování obrazu, kdy se nevyužívá sémantika objektů, tzn. že se obrazy neinterpretují. Využívá se metoda zpracování signálů, např. 2D Fourierova transformace. Cílem nižší úrovně je analyzovat vstupní dvourozměrná data číselného charakteru, odstranit šum z obrazu, rozpoznat jednoduché objekty v obraze a nalézt potřebné informace pro vyšší úroveň. Vyšší úroveň je chápána jako porozumění obsahu. Je o mnoho složitější a jejím jádrem jsou znalostní systémy a techniky umělé inteligence.

Základní kroky pro zpracování obrazu:

- 1) snímání a digitalizace
- 2) předzpracování
- 3) segmentace obrazu na objekty
- 4) popis objektů
- 5) porozumění obsahu obrazu (klasifikace objektů)

2.1 Snímání obrazu

K získání obrazu slouží libovolný přístroj, který obsahuje snímací obrazový prvek neboli snímač (v současnosti převládá CCD snímač). Podle typu snímače získáváme dvojrozměrný, trojrozměrný obraz nebo sekvenci snímků (video).

Nejprve dojde k zachycení fotonů rozptýleného světelného záření z objektivu obrazovým snímačem a následně k jejich převodu podle jejich intenzity (resp. jejich vlnové délky) na elektrický náboj. Intenzita elektrického proudu se na obraze projevuje jasnějšími a

tmavšími místy. Abychom mohli s obrazem dále pracovat, je třeba převést získaný analogový signál na digitální signál a ve formě 0 a 1 jej uložit na paměťové médium.

Snímače jsou zpravidla šeroslepé, proto pro ukládání obrazu užíváme RGB barevný model s 8 bity na barvu. Díky RGB modelu pak snadno získáme velkou škálu barev (256^3 barev).

2.2 Předzpracování

Cílem předzpracování je potlačit šum vzniklý při digitalizaci a přenosu obrazu, odstranit zkreslení, která vznikla při snímání. Mezi základní metody předzpracování obrazu řadíme převedení na stupně šedi, úprava jasů a kontrastu, filtraci, potlačení vlivu osvětlení aj.

2.3 Segmentace obrazu

Při procesu segmentace dochází k rozčlenění obrazu na jednotlivé části, pixely. Každý pixel nese určitou vizuální vlastnost konkrétního objektu z obrazu, jako je jas, barva či textura. Informaci o rozdělení obrazu do jednotlivých segmentů využívají vyšší algoritmy zpracování obrazu. Podle toho, jak k segmentaci přistupujeme, použijeme vhodný segmentační algoritmus (např. detekce hran, statistické metody, hybridní metody aj.)

2.4 Popis objektů

Jedná se o přesný popis nalezených oblastí obrazu tak, aby na základě tohoto popisu bylo možné objekty identifikovat. Objekty popisujeme buď pomocí číselných charakteristik, pracujeme tedy s příznaky objektů (například při popisu velikosti objektu, barva, šířka, aj.) nebo pracujeme s popisy relací mezi objekty.

2.5 Klasifikace (rozpoznávání) objektů

Cílem klasifikace je porozumět sémantice obrazu, a to na základě společných příznaků jednotlivých objektů. Proces spočívá v zařazování objektů do předem stanovených tříd předmětů, přičemž třídu chápeme jako podmnožinu prvků, jejichž atributy mají z hlediska klasifikace společné rysy. Klasifikaci nám umožňuje klasifikátor, který rozhoduje o zařazení objektu do třídy nikoli podle jeho skutečného, nýbrž podle jeho obrazu. Standardně jsou klasifikátoru předkládány objekty, se kterými je předem obeznámen a pokud mu je předložen neznámý objekt, většinou nedokáže předmět rozpoznat. Existují samozřejmě systémy, které se dokážou adaptovat.

3 Vlastní rozpoznávání - metody

V rámci umělé inteligence je nejužívanější metodou vícevrstvá neuronová síť (MLP). Vícevrstvá síť představuje distribuovaný výpočetní systém tvořený vrstvami umělých neuronů. Každá vrstva má přístup k výstupním hodnotám předchozí vrstvy. Takovéto neuronové sítě jsou schopny na základě algoritmu a procesu učení rozpoznávat objekty, tváře, číslice či písmena.

Příkladem může být Hopfieldova síť, která se často využívá k rozpoznávání státní poznávací značky, a v to v čase jeden znak za 0,09 s. Jedná se tedy o vysoce výkonný systém, který lze uplatnit na parkovištích, v garážích nebo pro bezpečnostní kontrolu při vjezdu do střeženého areálu. Funguje tak, že nejprve dojde k zachycení (snímání) poznávací značky a

následně k segmentaci obrazu, kdy se digitální obraz rozloží na jednotlivé regiony. Vygenerované výstupní znaky abecedy (písmena a číslice) se importují a ukládají do databáze. Uložené šablony písmen slouží v databázovém systému pro následné porovnávání znaků s registrační značkou.

3.1 Optické rozpoznávání znaků (OCR)

Jedná se o metodu využívající program, který umožňuje digitalizaci textových dokumentů. Díky čemuž lze s těmito dokumenty pracovat jako s textovými soubory, nikoli jako s obrázkovými soubory (což se standardně děje).

Příkladem může být program Raster Design, který z digitálního rastrového obrázku rozpozná písmo (tištěné či psané ručně). Nejprve dochází k analýze dokumentu, dále k rozpoznání textu na základě specifických charakteristik pro každé písmo, které má již tento program naučený.

Obtížnější je rozpoznání ručně psaného textu, ale i tento text je s relativně dobrou přesností úspěšný. Podmínkou netištěného textu je psaní tiskacími písmeny. Do této skupiny celá škála nejrůznějších programů. Na obdobném principu funguje i mobilní aplikace Scan, která rozlišuje typy fontů z digitálního obrazu. Při klasifikaci detekovaného textu však vybírá z třídy fontů, nikoli z třídy grafému jako u předchozího programu.

3.2 2D čtení kódů

S dvojrozměrnými kódy se setkáváme jako s formou označení zboží, zásilek, součástek, publikací aj. Prostřednictvím čárového kódu lze zakódovat určitou informaci, která se pak snadno přečte pomocí odpovídající čtečky. Nejužívanější formou je tzv. QR kód, který má čtvercovou maticovou formu. Kód získáme pořízením snímku, nejčastěji pomocí mobilní kamery, kdy se zachytí detekční vzory sloužící pro detekci polohy (tři „poziční“ čtverce), dále jeden nebo více „orientačních“ čtverců (stanovení správné orientace obrázku, pokud je skenován pod úhlem), „měřicí“ linie, které jsou umístěny na dvou stranách QR kódu a propojují „poziční“ čtverce (umožňují zkontrolovat, zda je skenovaný kód úplný). Podle těchto rysů čtečka pozná, která verze a formát QR kódu jsou použity a kód následně interpretuje. Příkladem může být seznam.cz QR čtečka.

3.3 Rozpoznávání tváře (face detection)

První fází je detekce (vyhledání) obličeje v obrazu, a to tak, že ignorujeme ostatní objekty. Standardně se vyhledaný obličej graficky označí „čtvercem“. U většiny metod se takto označený obličej následně porovnává s obličejem uloženými v databázi. Pro rozpoznávání lidské tváře existují dvě základní metody – strukturální a holistická.

Volně dostupným programem rozpoznávání lidské tváře fungující na principu strukturální metody je například software Face++ psaný v jazyce C++. Na tomto konkrétním příkladu implementují princip fungování strukturální metody. Face++ stejně jako většina programů pro rozpoznání tváře zahrnuje tři základní kroky:

1. detekci - klasifikuje objekt do konkrétní třídy objektů, což jsou v tomto případě třídy všech obličejů, přičemž třída všech obličejů má své specifické funkce, které chápeme jako vizuální rysy obličeje a jsou detekovány od středového bodu v obraze - barva pleti, nalezení vzdálenosti mezi očima, nalezení nosu, úst aj.
2. rozpoznání – rysy obličeje získané při detekci se porovnávají s rysy obličejů uložených v databázi na základě použití klasifikačního algoritmu (v případě Face++ se v databázi porovnává na základě 83 společných rysů)
3. analýzu – statistické rozhodnutí o relativní podobnosti s takto vybranou množinou prvků (dochází k rozlišení pohlaví, věku, rasy a eventuálně i emoce)

3.4 Metoda DeepFace

Asi nejběžnější a často i nechtěně používanou metodou je DeepFace užívaná na Facebooku. Tato technologie analyzuje 3D strukturu obličeje, vytvoří z ní 2D vrstvu a dalším postupem izoluje charakteristické prvky. Princip automatického rozpoznání tváří a přidávání jmen k fotkám funguje jednoduše tak, že jakmile nahrajete (nebo někdo jiný) na Facebook fotografie, Facebook využije programu pro rozpoznávání obličejů a tváře na fotografiích srovná s existujícími fotografiemi a jmény. Poté uživateli navrhne automatické použití těchto jmen pro označení lidí na nových snímcích (tzv. tagování). Program využívá devítivrstvou neuronovou síť, která má naučených zhruba 100 000 možných parametrů.

3.5 Využití - biometrie

Velkým přínosem je rozpoznávání tváře zejména v biometrii, kdy se uživatelé mohou přihlásit, aniž by museli zadat jakékoli přihlašovací údaje, a to pouze na základě pohledu na kameru či na základě svého hlasu. Příkladem užití v biometrii je BioID Web Service (BWS), což je program, který umožňuje zejména v rámci firmy přístup do budovy, do počítače, či vstup do počítačového firemního programu. Funguje na principu standardní detekce obličeje a následného dohledávání tváře z databáze všech předem daných tváří.

4 Metody analýza pohybu

4.1 Rozdílová metoda

Na obrázku vidíme praktické užití programu založené na rozdílové metodě. Rozdílová metoda je primitivní metoda založená na využívání rozdílů mezi obrazy snímanými v různých časových okamžicích. Využívá se pro aplikace, které probíhají v reálném čase, přičemž rozdílový obraz nabývá hodnot 0 a 1. Hodnota 0 představuje obraz, kde se rozdíl jasové hodnoty bodů výrazně neliší. Hodnota 1 představuje obraz, kde dochází k výrazné změně jasové hodnoty bodů. Výsledný, tedy třetí snímek, zachycuje výsledný rozdíl. Poloha objektu se stanoví podle pozice těžišť jednotlivých objektů v jednotlivých rozdílových obrazech. Užitým zařízením pro vstup dat je pevně upevněná průmyslová (stacionární) kamera. Příkladem je situace na obrázku představující křižovatku. Tato metoda má však nevýhodu, může se stát, že auto, které je na obou snímcích, na výsledném rozdílu chybí, neboť auto stojí, a proto rozdílem odstraněno.

4.2 Detekce pohybu pomocí kontury

Dalším příkladem je aplikace pro detekci pohybu kontury osob z videa. Tato metoda vychází z detekce pohybu a výpočtu kontury. Při detekci pohybu zjistíme souřadnice pohybujícího se objektu, díky kterým dochází k výřezu obrazu ze vstupního snímku. Současně se podle těchto souřadnic vygeneruje počáteční maska pro výpočet kontury. V tomto okamžiku dochází k rozdělení obrazu pomocí masky na vnitřní a vnější oblasti a spočítá se průměr vnitřních a vnějších hodnot. Zprůměrováním těchto hodnot následně vzniká výsledná kontura, která se zakreslí do výstupního snímku a vloží se do výstupního videa. Nevýhodou této metody je vliv překážek, což je ukázáno na obrázku, kdy algoritmus nerozpoznal, že se jedná o stín na skle.

4.3 Optický tok

Tato metoda zachycuje všechny změny obrazu v čase dt. Každému bodu v obrazu odpovídá dvojrozměrný vektor rychlosti, vypovídající o směru a velikosti rychlosti pohybu v daném místě obrazu. Metodu optického toku můžeme použít v situacích, kdy se setkáváme se statickou kamerou a

pohybující se scénou, pohybující se kamerou a statickou scénou, nebo když se současně pohybuje kamera i scéna. Výstupem metody optického toku je rychlostní pole, které zobrazuje vektory rychlosti pro každý bod v tomto poli. Přičemž platí, že hodnoty, které jsou menší než předem stanovená hodnota, jsou prohlášeny za pozadí, a naopak hodnoty, které jsou vyšší, jsou považovány za objekty.

5 Rekonstrukce 3D scény

Trojrozměrný model tváře můžeme získat nejčastěji dvojím způsobem, a to konverzí 2D obrazu na 3D scénu anebo pomocí zařízení, které je schopno zachytit 3D scénu přímo – tzv. 3D scanner. Oba způsoby uplatňují nějakou z metod umělé inteligence, např. použití neuronové sítě či „eigenface“ a dále se zvolí nějaký algoritmus, který slouží jako matematický vzorec uplatněný na lidskou tvář.

5.1 Konverze z 2D obrazu na 3D scénu

V případě rekonstrukce 3D scény z dvojrozměrného obrazu se nejčastěji používá tzv. normálová mapa, kdy dvojrozměrný obraz reprezentuje povrch 3D grafiky a jeho jednotlivé RGB pixely představují tzv. normály. Nejprve dojde k získání přibližného tvaru obličeje za pomoci aproximace. V tomto kroku je s využitím normál vytvořena rekonstrukce obličeje pomocí polygonů a vznikne síť reprezentující tvar obličeje ve 3D. Následně je vytvořena sférická projekce (promítnutí) 3D sítě na 2D plátno. Posledním krokem je převedení normál ze získané sítě na dvourozměrné pole, kde každý obrazový bod představuje normálu pomocí hodnoty barevného spektra RGB.

5.2 3D scanner / 3D čtečka

Principem 3D scanneru je promítání přesně definované mřížky na skenovaný objekt a její následné snímání kamerou. V jednoduchém softwaru je pak na základě deformace mřížky na snímaném objektu vypočítán a rekonstruován jeho povrch. Skenování probíhá v reálném čase a uživatel vidí na monitoru počítače, která část modelu již byla naskenována.

6 Robotické rozpoznávání objektů

Roboty, kteří jsou schopni rozpoznání objektu, řadíme k robotům 2. generace. Jedná se o roboty vyšší úrovně, kteří jsou vybaveni řadou senzorů nebo schopností vidění pomocí kamery. Prvním robotem svého druhu byl robot se systémem „oko-ruka“, který byl vytvořen roku 1970. Byl vybaven kamerou, manipulační paží o šesti stupních volnosti s elektrickým pohonem, mikrofonem pro příjem řeči a počítačem. Jeho původní úlohou bylo sestavení různých velkých kostek, které byly náhodně rozhozeny po stole. Experimenty spočívaly v analýze a napodobení struktur sestavených z kostek, což vedlo k rozvoji různých metod automatické analýzy scén sestávajících z mnohostěnů.

Na tomto principu by tvořen i robot *Freddy* z Edinburské univerzity, který byl vybaven mechanickou rukou obsahující dotyková čidla, dvěma televizními kamerami a počítačem. Jeho úloha spočívala v nasunutí čtyř různých disků na kolík nebo montáž autíčka z šesti dílů. Na počátku experimentu byla na stole hromada různých předmětů, které Freddy postupně vyčleňoval a rozpoznával díky databázi předem naučených předmětů. Rozpoznávání bylo prováděno na základě klasifikace předmětů do tříd. Potom sestavil předměty do určené konfigurace podle určeného postupu. Jeho pozdějším a náročnějším

úkolem byla montáž vodního čerpadla automobilu Ford, kdy již rozpoznávání předmětů vyžadovalo složitější algoritmy.

Zdroje

BERJAK, Ing. Jiří. *Automatická analýza a rozpoznávání objektů v obraze pomocí fázové korelace*. Brno: VUTIUM, 2010. ISBN 978-80-214-4049-4.

LÝSEK, Ing. Jiří. *Rozpoznávání objektů pomocí evolučních metod*. Brno: VUTIUM, 2013. ISBN 978-80-214-4875-9.

ŠŤASTNÝ, Jiří. *Netradiční metody a algoritmy pro rozpoznání objektů technologické scény*. Brno: VUTIUM, 2006. ISBN 80-2143-3117-2.

Internetové zdroje

Computer vision. In: *Wikipedie, otevřená encyklopedie* [online]. [cit. 2014-11-29]. Dostupné z: http://en.wikipedia.org/wiki/Computer_vision

Neuronové sítě – základy rozpoznávání znaků. *Shrimphood* [online]. [cit. 2014-11-29]. Dostupné z: <http://www.shrimphood.net/neuronove-site-zaklady-rozpoznavani-znaku.html>

MORÁVEK, Bc. Petr. *Identifikace obličeje osoby snímané kamerou: Person face identification using digital camera* [online]. Zlín, 2010 [cit. 2014-11-10]. Dostupné z: https://dspace.k.utb.cz/bitstream/handle/10563/11868/mor%C3%A1vek_2010_dp.pdf?sequence=1. Univerzita Tomáše