

POSUDEK OPONENTA HABILITAČNÍ PRÁCE

Masarykova univerzita

Uchazeč

Habilitační práce

Oponent

**Pracoviště opONENTA,
institute**

RNDr. Jan Sedmidubský, Ph.D.

Content-Based Processing of Human Motion Data

prof. RNDr. Tomáš Skopal, Ph.D.

Katedra softwarového inženýrství MFF UK

Habilitační práce s názvem „Content-Based Processing of Human Motion Data“ autora Dr. Jana Sedmidubského je předkládána jako komentovaná kolekce 10 publikací v oblasti podobnostního vyhledávání v souborech dat zaznamenávajících lidský pohyb. Celá práce je sestavena jako koherentní série na sebe navazujících prací, které dohromady danou problematiku řeší komplexně v celé šíři, tj. od modelování deskriptorů pohybu a podobnostních modelů pro sekvence 3D skeletónů, přes doménově specifické vyhledávání, detekční, temporálně segmentační či rozpoznávací mechanismy, až po databázové indexování. Všechny zmíněné metody jsou implementovány v rámci prototypových softwarových aplikací.

Obsah práce:

První tři práce předkládané kolekce se zabývají reprezentací sekvencí 3D skeletónů. Zde autor (a spoluautoři) aplikovali, resp. adaptovali různé modely z oblastí strojového učení (+ transfer learning) a information retrieval. První práce se zabývá transformací pohybu skeletónu do obrázku (motion image) a následně využitím existujících, či fine-tuned konvolučních hlubokých sítí (DCNN) pro klasifikaci obrázu. Další přístup využívá přímější reprezentaci pro pohyb – reprezentaci pomocí embeddingů ve spojitosti s obousměrnými rekurentními LSTM sítěmi (tato práce není zařazena do habilitačního výběru). Druhá a třetí práce v kolekci se zabývají symbolickou reprezentací pohybu na principu bag of words, tedy konstrukcí slovníku pomocí shlukování synteticky segmentovaných pohybů do menších „slov pohybu“, kvantizací, reprezentací částí pohybu slovy a porovnáváním výsledných „textů pohybu“ pomocí vzdálenostní/podobnostní funkce DTW. V další podkapitole komentáře autor doplňuje problematiku reprezentací výsledky (nezařazenými do kolekce prací) v oblasti detekce chůze (gait recognition), kde popisuje specifické metody segmentace pohybu chůze (walk cycle detection) a její reprezentace pro potřebu identifikace osob (biometrické aplikace).

V další části práce se autor věnuje obecné problematice klasifikace pohybové akce (action recognition), kde představuje dvě metody, resp. publikace 4 a 5 v kolekci prací, využívající již zmíněné reprezentace pohybu 3D skeletónů pomocí DCNN (motion images) a LSTM deep feature vektorů. Samotná klasifikace se realizuje přes kNN klasifikátor. Při použití motion images+DCNN se následně výsledek kNN dotazu přeuspořádá jiným podobnostním modelem (využívajícím jiné deskriptory), čímž lze dosáhnout přesnější klasifikace v případě různých, ale velmi podobných akcí. Při použití LSTM modelu byla navíc navržena metoda zvětšení typicky malého množství trénovacích dat (data augmentation) replikací existujících dat s jejich modifikací specifickými „editacemi“ (temporální střihy, malé náhodné manipulace s klouby skeletónu, detekce klíčových póz).

Další část práce se věnuje problematice rychlého vyhledávání pohybových podsekvencí, což je typická situace v praktických aplikacích (krátký pohyb dotazu je třeba lokalizovat v dlouhých pohybech např. v mnohahodinových záznamech kamer). Autor představuje tři metody (přiložené jako publikace 6, 7, 8 v kolekci), které se liší úrovní granularity segmentace pohybů. První práce je založena na indexaci jednotlivých póz v pohybu, druhá začleňuje také časovou složku a využívá reprezentace krátkých (sekundových) segmentů. Obě vyhledávací metody však potřebují dvě fáze – search and refine – neboť vyhledávání na základě více dotazových podčástí (ať jednotlivých póz nebo podsegmentů) vede k mnoha izolovaným kandidátům, které je třeba verifikovat a filtrovat v refinement fázi. Třetí práce pak tento nedostatek eliminuje tím, že databázové sekvence jsou předem nasegmentovány na různě dlouhé podsekvence do víceúrovňového systému a při vyhledávání je použita taková segmentace, která délkou přímo odpovídá sekvenci aktuálního dotazu a tedy porovnávané segmenty není třeba rozbíjet a následně slučovat v refinement fázi.

Poslední část práce se věnuje detekci akcí (action detection), což lze chápat jako zobecnění úkolu klasifikace akcí, kdy se krátké akce hledají v dlouhých sekvencích (podobně jako předchozí problematika vyhledávání v podsekvencích, která ale nepracuje s klasifikacemi). Pro použití detekce akcí se v aplikacích zpravidla předpokládá streamovaná databáze 3D póz, potažmo video stream z IP kamery v reálném čase, ze kterého jsou pózy extrahovány. Autor k tomuto tématu přikládá do kolekce prací dva výsledky (publikace 9 a 10). První práce problematiku detekce akcí řeší na úrovni víceúrovňové segmentace sekvencí (tedy aplikuje práci zmíněnou výše pro hledání podsekvencí), druhá pak na úrovni jednotlivých póz, která je výrazně rychlejší.

Práce je zakončena shrnutím a diskuzí nad plány do budoucna.

Hodnocení:

Komentář ke kolekci prací (25 stran + 108 referencí) je zpracován srozumitelně a popisuje práce tak, aby na sebe navazovaly. Samotná kolekce 10 prací představuje kvalitní recenzované publikace v časopisech s IF (4) a na prestižních konferencích (6, z toho 1xCORE A, 3xCORE B, 2xCORE C). Kromě toho má za sebou Dr. Sedmidubský řadu dalších publikací v dané oblasti a také tutoriály; jeden na ACM Multimedia (CORE A*). Třetí a osmá práce v kolekci byly vybrány mezi „Best papers“, devátá pak získala ocenění „Best student paper“ (spoluautor student P. Elias). Všechny práce v kolekci mají více autorů, kdy v 6 z nich má Dr. Sedmidubský autorský podíl nad 50% (v ostatních nad 30%). Výzkum v této oblasti vyžaduje týmovou práci a spoluautorství na publikacích je jeho nezbytnou součástí.

Co se týče metodologie, přístupů a kvality výsledků, musím ocenit originalitu přístupů (např. klasifikace motion images přes AlexNet) a zároveň snahu integrovat autorův výzkum se state-of-the-art výsledky z jiných oblastí, zejména z oblasti strojového učení. Veškeré výsledky jsou ověřeny experimentálně, jak je ostatně v mezinárodní výzkumné komunitě standardem. Formální a jazyková úroveň je na velmi dobré úrovni (u přehledového komentáře se neočekává výraznější matematická formalizace). Vysoká kvalita samotných publikací v kolekci již byla prokázána v peer-review procesu u daných konferencí/časopisů.

Daná problematika je velmi aktuální zejména v aplikacích, neboť rozpoznávání událostí na základě lidského pohybu (nejčastěji chůze, ale i dalších, „akrobatičtějších“ pohybů) lze využít v mnoha doménách, které autor zmiňuje, např. v lékařství/rehabilitaci, ve sportu, a zejména v aplikacích dopravních, bezpečnostních a biometrických. Aktuálnost je podpořena dostupností velkého množství dat v podobě databází časových řad 3D skeletonů. Jelikož autor pro své metody zpravidla nepředpokládá specifické senzorické vybavení, ale běžný video záznam z nezávislých indoor/outdoor kamer a jeho zpracování do dostatečně přesné reprezentace

pohybu 3D skeletů, lze do budoucna předpokládat masové a diverzifikované využití navrhovaných metod v praxi. Vedle aplikací nabízí předložená práce mnoho inspirace pro navazující výzkum a rozvoj oboru, jak ostatně autor zmiňuje v poslední části komentáře.

Dotazy oponenta k obhajobě habilitační práce:

1. Ve future work uvádíte složitější podmínky a modely, např. scény se skupinami osob, okluzemi, apod. Zde se nabízí použití záznamu/streamu z pouličních kamer, které produkují nekvalitní obraz (světelné podmínky, rozlišení, motion blur). Lze úspěšně použít vaše metody na takto nekvalitní vstup (tj. jak kvalitní 3D skelet lze dostat)?
2. Jak efektivní jsou motion detection techniky v biometrických aplikacích? Lze je používat i samostatně, nebo pouze v multi-modální kombinaci (zmíněný face deskriptor)?

Závěr:

Habilitační práce RNDr. Jana Sedmidubského, Ph.D., s názvem „Content-Based Processing of Human Motion Data“ **splňuje** požadavky standardně kladené na habilitační práce v oboru Informatika. Doporučuji Dr. Sedmidubskému udělit titul docent v oboru Informatika.

V Berouně,

Dne 20.8.2021

