FAKULTA APLIKOVANÝCH VĚD ZÁPADOČESKÉ UNIVERZITY V PLZNI

STUDENTSKÁ VĚDECKÁ KONFERENCE 2016

Rekonstrukce pózy ruky z hloubkového obrazu

Milan Herbig[1]

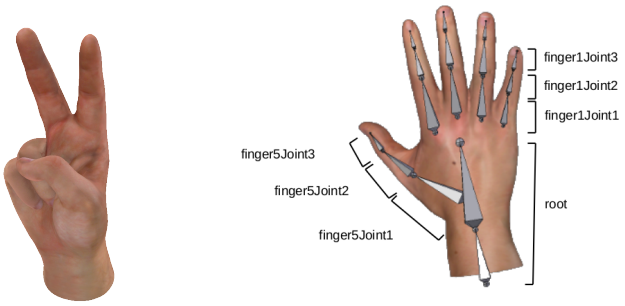
1 Úvod

Úloha přesné a robustní rekonstrukce pózy ruky dodnes představuje výzvu pro skupiny výzkumníků po celém světě. Příkladem dnes již běžně fungujícího prostředku pro odhad polohy a pohybu částí těla je zařízení Microsoft Kinect. Problémovou částí však nadále zůstávají ruce. Ty bývají kvůli zjednodušení nebo nedostatečnému rozlišení snímače modelovány velice zjednodušeně, či vůbec. Skelet ruky totiž oproti skeletu těla skrývá daleko více stupňů volnosti. Zároveň je ruka nepoměrně menší vůči zbytku těla, tedy veškerá informace je zakódovaná v podstatně menším množství pixelů. Jednotlivé části ruky se navíc často překrývají a vzájemně vykazují velkou lokální podobnost.

Schopnost rozpoznat či zrekonstruovat pózu ruky otevírá nové možnosti ovládání nejrůznějších zařízení na dálku, například televize. Dlouholetou motivací je bezpochyby úloha rozpoznávání znakové řeči. Aktuálním trendem je rychle rostoucí perspektivní odvětví rozšířené a virtuální reality. Pro vyřešení úlohy odhadu parametrů ruky používám konvoluční neuronovou síť.

2 Návrh řešení úlohy

Použití konvoluční neuronové sítě pro regresi parametrů (pózy) ruky lze označit za novátorskou myšlenku. Až teprve v průběhu příprav experimentů byla publikována práce Tompson et al. (2014), která jako první potvrdila, že konvoluční neuronovou síť lze pro úlohu regrese parametrů ruky použít. Avšak na rozdíl od ostatních dnes již existujících prací ve své práci provádím odhad rotace jednotlivých částí ruky v kvaternionech namísto jejich polohy. Výstupem estimátoru jsou tedy přímo úhly jednotlivých kostí zjednodušeného modelu ruky, který je uveden na obrázku 1.



Obrázek 1: Ilustrace použitého modelu ruky a popis jednotlivých článků.

[1] student navazujícího studijního programu Inženýrská informatika, obor Řídící a rozhodovací systémy, e-mail: herbig@students.zcu.cz

Rekonstrukci odhadnuté pózy do 3D modelu provádím pomocí programu Blender[1]. Trénování konvoluční neuronové sítě probíhá na uměle vygenerovaných datech z 3D modelovacího programu. Získání reálných anotovaných dat je totiž nejen komplikované vzhledem k počtu stupňů volnosti, ale i časově náročné vhledem k variabilitě ruky (více herců).

3 Návrh konvoluční neuronové sítě

Začal jsem se sériovou architekturou tvořenou dvěma až třemi konvolučními vrstvami a dvěma až čtyřmi fully-connected vrstvami. Problémem této architektury byl kompromis mezi zachycením lokálních detailů (přesnost) a globálních vazeb (hierarchie ruky). Nakonec jsem stejně jako autoři ostatních prací použil paralelní tzv. multi-scale architekturu. Tedy vstupní snímek je transformován na další dva snímky vždy o poloviční velikosti (96, 48 a 24 pixelů). Díky tomu lze stejnou velikostí konvolučního okénka zachytit jinak velikou oblast, a tak pochytit včetně detailních lokálních vazeb i vazby globální. Každá z větví je tvořena dvěma konvolučními vrstvami. Všechny tři větve jsou posléze spojeny do jedné čtyřvrstvé fully-connected sítě. Oproti monolitické architektuře se multi-scale architektura liší jak v dosažené přesnosti, tak v době potřebné pro natrénování (paralelizace).

4 Nejlepší dosažené výsledky

finger1joint1	**finger1joint2**	**finger1joint3**	**finger2joint1**	**finger2joint2**	**finger2joint3**
6.70°	5.56	6.55°	5.18°	5.48°	7.45°

finger3joint1	**finger3joint2**	**finger3joint3**	**finger4joint1**	**finger4joint2**	**finger4joint3**
5.46°	5.50°	7.00°	6.30°	5.15°	4.96°

finger5joint1	**finger5joint2**	**finger5joint3**	**root**
8.93°	5.83°	6.74°	4.66°

průměr
6.09°

Tabulka 1: Tabulka s nejlepšími dosaženými výsledky.

V tabulce 1 jsou uvedeny nejlepší dosažené výsledky na umělých datech s multi-scale architekturou konvoluční neuronové sítě. Chyba uvedená v metrice popsané vztahem 1 vyjadřuje rozdíl mezi predikovanou a ground-truth hodnotou ve stupních. Jedná se o rotaci, kterou je potřeba vynaložit k přetočení jednoho z normalizovaných rotačních vektorů (kvaternionů) $\overrightarrow{q_1}$ nebo $\overrightarrow{q_2}$ tak, aby byly oba vektory shodné.

$$\theta = cos^{-1}(2(\overrightarrow{q_1} \cdot \overrightarrow{q_2})^2 - 1) \tag{1}$$

Pro trénování neuronové sítě (GPU) i pro generování trénovacích dat jsem využil kapacit výpočetního prostředí Metacentra. Bez těchto výpočetních kapacit by výstupní diplomová práce nikdy nemohla vzniknout.

Literatura

Tompson, J., Stein, M., Lecun, Y., and Perlin, K., 2014. Real-time continuous pose recovery of human hands using convolutional networks. *ACM Transactions on Graphics (TOG), 33(5), 169.*

[1]Blender je open-source 3D modelovací program - http://www.blender.org/