

Разпознаване на човешки активности на база 3D реконструкция на скелета

<http://kinect-human-action-recognition.googlecode.com/>

Автор: Любомир Янчев;
11 клас, ЧНГ „Ерих Кестнер“;
София;
yanchev.lyubomir@gmail.com

Ръководител: Светослав Колев;
University of Washington;
swetko@cs.washington.edu

Абстракт

Наскоро нашумялата тема за разпознаване на човешки действия се радва на все повече популярност. Целта на настоящата разработка е да представи качествено ново решение на посочения проблем.

В предложеното решение са използвани многоизмерни статистически алгоритми, алгоритми за филтриране на данни, алгоритми за оценка и други. Триизмерната информация за скелета се извлича от Kinect сензор. Чрез извършването на малка модификация, предложените алгоритми биха могли да се използват за разпознаване и на друг вид зависимости.

В процес на развитие е сравнение между разработения алгоритъм, традиционно използваните и някои нестандартни подходи, описани в секцията „Допълнителни имплементации и тестове“

Въведение

Разпознаването на човешки действия остава една от най-значимите сфери на изследване за компютърно зрение. Разработването на алгоритми за целта е стимулирано от многото потенциални приложения на подобна технология – от автоматизирано видео наблюдение, през базирано на съдържанието видео, управление на компютри до вграждането му в „умни домове“. Подходът може да се използва и за наскоро нашумялото предопределяне на болести на база промени в движенията на пациента – често преди инфаркт се забелязва леко завлачване на левия крак.

След щателно проучване на достъпните технологии установихме, че в Kinect алгоритмът за позициониране на ставите в триизмерното пространство е най-ефективно имплементиран. Поради тази причина и качеството на разработката, взехме решението да използваме този сензор.

В раздел „Допълнителни имплементации и тестове“ са приложени тестове с други алгоритми, освен основния за тази разработка.

Премишна работа

Множество решения на проблема, използващи различни източници на данни, са предложени през изминалите години. Информация за позата е най-често извличана от системи с маркери (1), монокулярни камери (2) (3) (4) (5), силуетни образи (6) и алгоритми, използващи дълбочинна картина (7). След цялостно проучване, не успяхме да намерим публикувана разработка, използваща Kinect сензор.

На разположение са също и алгоритми, способни да заменят Kinect сензора. Такива са например базираните на TOF камери (8), монокулярни (9) (10) (11) (12), дълбочинни (13) (14) и други (15), които бяха отхвърлени заради качеството на Kinect сензора.

Методи

Тази секция цели да изясни подробно методите използвани за реализация на алгоритъма.

Ъгли и ротация на ставите

След анализ на данните, които бихме могли да извлечем от използвания сензор, взехме решение, че най-добрият подход към проблема би бил да използваме ротацията на ставите в пространството, тъй като тя еднозначно дава информация за движението. (16)

Друг плюс на използването на ротациите на ставите за първостепенна информация за разпознаване на дейностите е, че не се обвързваме с точни координати, а от там – и с точни дължини на костите на скелета.

Представяне на ротацията на ставите чрез геометрични кватерниони

Чрез прилагането на този метод бихме могли да представим ротацията и позицията на ставата чрез геометричен кватернион, който ни служи, както да сравняваме приликата между ротациите, така и да реконструираме позицията и ротацията на ставата.

Под предвид е взето, че системата на кватерниона дава както положителен, така и отрицателен резултат.

Селекция на „най-информативни“ стави

Минали проучвания показват, че за разпознаване на движения на база ротацията на ставите, не е нужно да се разглеждат всички стави на скелета (16) (17). Достатъчно е да следим единствено ставите, които имат най-голяма промяна в ротацията си по време на движението. Прост пример би било махането с ръка. Няма смисъл да взимаме предвид ъгъла на левия глезен, за да разпознаем това движение, тъй като ставите, които определят движението са рамото и лакътят. Подкрепящи това твърдение доказателства могат да бъдат намерени на (17) и на фигура 1.

Фигура 1:

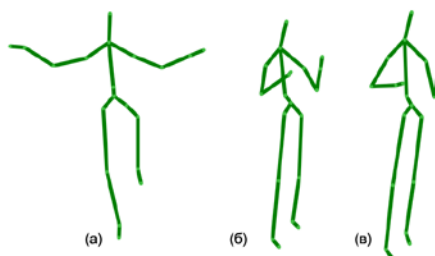


Най-активните стави се маркират при предварително записаните движения. На записите се прави анализ, който определя степента на активност по формулата.

Сравняване на движения

За да решим глобалния проблем за разпознаване на активности в реално време, първо трябва да решим подпроблема за сравняване на два скелета. Прост пример виждаме във фигура 3а, 3б и 3в. Лесно можем да определим, че фигура 3б прилича много повече на фигура 3в, отколкото на 3а.

Фигура 3:



Еднозначен метод за сравнение на ротацията на ставите чрез кватерниони (R, S) е да използваме следната формула:

$$\text{Difference} = (R_x - S_x)^2 + (R_y - S_y)^2 + (R_z - S_z)^2 + (R_w - S_w)^2$$

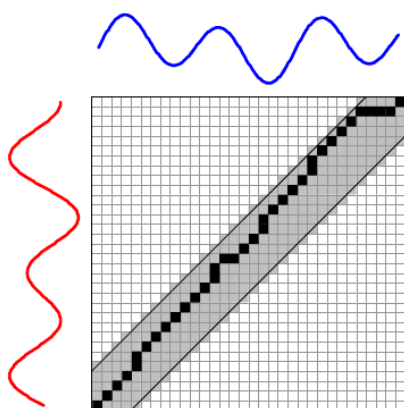
Dynamic time warping (DWT)

Поради липсата на български превод на понятието, в настоящата разработка ще го наричаме „Динамично изкривяване на времето“ (ДИВ). Този метод служи за измерване на приликата между две редици, които могат да варират спрямо време и скорост. (18) (19) (20). Чрез прилагането на ДИВ бихме могли да намерим приликата между две активности, дори и ако те са извършени с различни скорости.

Аналогичен на дистанцията на Левенщайн, ДИВ е динамично-оптимизационен алгоритъм, който цели да намери каква е минималната цена, за която от едната редица може да се достигне другата. Това се достига чрез създаването на таблица, в която във всяка клетка записваме сбора от разстоянието между A_i и B_j и минималното намерено разстояние до сега.

Поради скоростта $O(n*m)$ на този алгоритъм, бяха имплементирани два начина за увеличаване на скоростта и успеваемостта му.

Sakoe-Chiba band



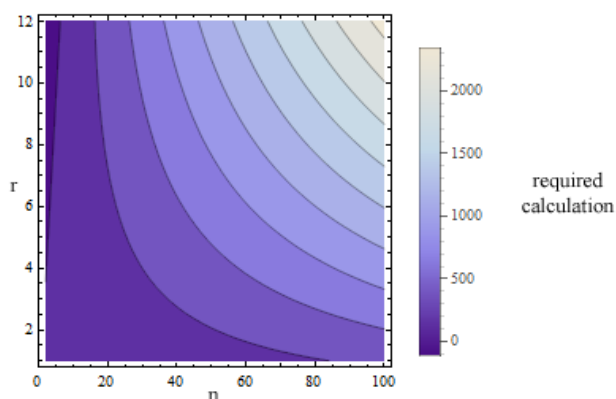
Лентата на Сакое Киб (оцветена в сиво) представлява дефиниран регион, за който се предполага, че съдържа пътя в таблицата с най-малък сбор от стойностите на клетките, дефиниран още като *warping path*.

Направен бе анализ (фигури 8, 9) на броя на извършените сметки при използването на Sakoe-Chiba band, който показва ефективността на подхода. На фигура 10 е представено в табличен вид подобрението при използването на подхода.

Използването на този подход ни позволява да увеличим размера на базата данни от предефинирани движения, които се използват за понататъшно сравнение с извършеното от

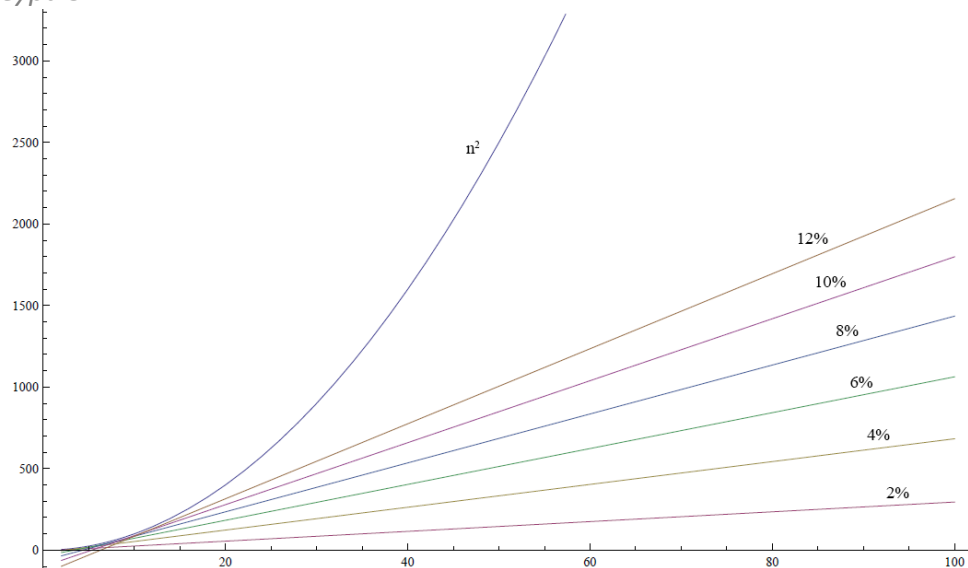
потребителя такова.

Фигура 8:



Контурна графика даваща информация за отношението между широчината на лентата, големината на записите и извършените сметки.

Фигура 9:



Графика, сравняваща нужните сметки за дадените в проценти от общата дължина широчини на Sakoe-Chiba band.

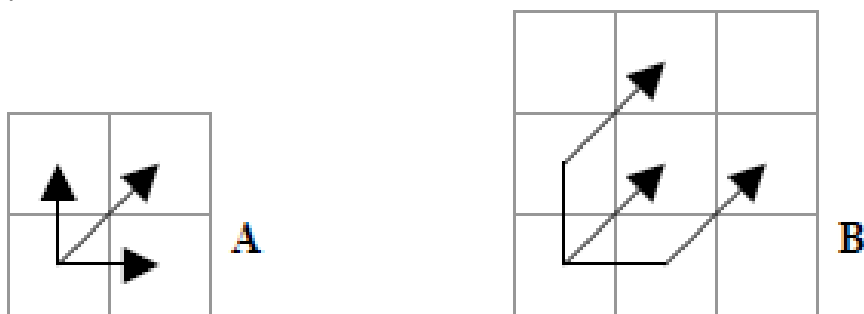
Фигура 9:

| N | Width in % | Calculations | Optimization |
|------------|------------|--------------|---------------|
| 100 | 2 | 295 | 97.05% |
| 100 | 4 | 683 | 93.17% |
| 100 | 6 | 1063 | 89.37% |
| 100 | 8 | 1435 | 85.65% |
| 100 | 10 | 1799 | 82.01% |
| 100 | 12 | 2155 | 78.45% |

Алтернативни стъпкови шаблони

Чрез използването на различни стъпкови шаблони би могло да се постигне по-голям шанс за правилно пресмятане на пътя между две редици. В контекста на разработката алтернативният вариант на стъпков шаблон (В) позволява прескачането на максимум един фрейм, което често се оказва полезен подход заради породилите се смущения във входните данни от Kinect сензора.

Фигура 11:



Заклучение

Разпознаването на активности в реално време е сложен проблем с много решения. В настоящата работа са предложени два подхода, чийто недостатък би бил единствено изискването на повече хардуерна мощ от среднестатистическата за домашен компютър.

В момента авторът работи върху сравняването на ДИВ с други алгоритми с цел намиране на най-подходящия. В процес на имплементация е и решение на проблема с използването на скритите модели на Марков. (22)

Възможни приложения

Прилагането на алгоритъма в реални условия би спомогнало за подобрието на качеството на услугите в широк спектър от сфери. От видеонаблюдение в магазини, където се следи за „съмнителни“ активности. Друго приложение би било интелигентна класификация на филми на база съдържанието им. Особен акцент трябва да се сложи върху системите за наблюдение на пациенти, които могат да следят за най-различни активности.

Друго приложение е изграждането на цялостен Natural User Interface интерфейс. Чрез разработения алгоритъм лесно би могло да се разработи приложение, което бива управлявано единствено чрез човешки действия.

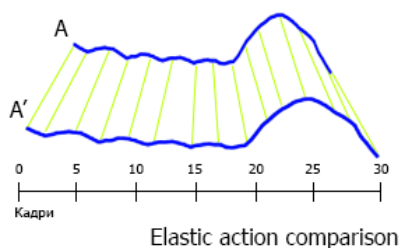
Системите за „умни домове“ все повече навлизат в нашето ежедневие. Чрез използването на предложения алгоритъм лесно може да се разширят възможностите им, като по този начин улеснят още повече ползвателите им.

Допълнителни имплементации и тестове

Еластично сравняване на движения

Еластичното сравняване на движения е подход, предложен от автора, който предстои да бъде тестван. В основата му седи хипотезата, че двете движения, които трябва да бъдат сравнени, започват с началото на движението и свършват точно с края му, поради което ще можем да сравним движения, извършени с различна скорост.

Фигура 5:



Нека означим двете движения с **A** и **B**, които са дълги съответно **m** и **n** кадъра. Ако тази хипотеза е изпълнена, изчисляваме **p**, където **frameNumber** е текущият кадър **A**.

$$p = \frac{\text{frameNumber}}{m} * n$$

Ако означим функцията за намиране на прилика между две пози с **comp(X, Y)**, то получаваме следната формула за намиране на прилика между два записа:

$$Difference = \sum_{frameNumber=0}^n comp(A_{frameNumber}, B_p)$$

Предстои да бъдат извършени тестове с алгоритъма.

Извършени тестове

След имплементирането на горепосочените алгоритми, бяха извършени 15 теста на всяко движение с всеки алгоритъм по отделно (фигура 5). Тестовите бяха проведени от автора поради невъзможността за намиране на доброволци. С увеличение на тренировъчните записи, възможността на системата да разпознава потребители, за които не е обучена, ще се увеличи.

Фигура 6:

| Algorithm Action | Squat | Swipe with left hand | Swipe with right hand | Waving with left hand | Waving with right hand | Taekwondo position |
|---|--------|-------------------------|--------------------------|--------------------------|---------------------------|-----------------------|
| Dynamic time warping | 86.60% | 100% | 100% | 91.66% | 71.43% | 90.91% |
| Elastic action comparison with freedom degree | 80.00% | 75.00% | 81.82% | 100.00% | 75.57% | 91.67% |

От горната таблица можем да постановим, че алгоритъмът, който дава най-добри резултати, е динамичното изкривяване на времето (ДИВ). Въпреки своите недостатъци, еластичното сравняване на движения не отстъпва и след някои подобрения би могло да се справя дори по-добре от ДИВ.

Бъдещо развитие

Главната цел в бъдещето развитие на проекта е имплементирането на скрити модели на Марков и комбинирането им с ДИВ с цел по-абстрактно разпознаване на активности – времетраенето на дадена активност би могло да се увеличи значително, а разработката ще може да се използва още по-ефективно в изброените ѝ сфери на приложение, както и в много други.

Заклучение

Преложените алгоритми решават поставения проблем чрез използването на ефективен подход за подбор на данни, при който те придобиват интуитивно физическо значение. За разпознаването на действия в реално време са използвани няколко алгоритъма, които предоставят релативно ефективен начин за разпознаване на еднакви движения, извършени с различни скорости. С цел по-високо ниво на качество, за набавяне на данни за триизмерния скелет на потребителя, се използва Kinect сензор. По-голяма точност на разпознаване на движенията се постига чрез филтриране на входните данни за скелета.

Благодарности

Авторът желае да изкаже благодарност на Румен Данговски, Валерия Станева, Калина Петрова и ръководителя на проекта, Светослав Колев, за ценните съвети и изказани мнения. Също така желае да изкаже благодарност и към Ученическия институт по математика и информатика (УЧИМИ) към БАН, който организира „Лятна изследователска школа“, където бе поставено началото на проекта.