

МЕТОДИ РОЗПІЗНАВАННЯ СТАТІ ТА ВІКУ ЗА ЗОБРАЖЕННЯМ ОСОБИ

UDC 004.93

Ia. O. Myshakivskyi

METHODS OF GENDER AND AGE RECOGNITION FROM THE IMAGE OF A PERSON

Розпізнавання по обличчю є популярним і точним підходом до завдання визначення статі та віку по фото або відео. Причиною цього є присутність на обличчі всіх візуальних особливостей, які дозволяють визначати стать, вік, расу, настрій та багато інших рис людини, візуально видимі в основному (або виключно) на вигляд і подання особи. На відео розпізнавання по особі переважно для визначення статі та віку у випадках, коли на кадрах особа людини видно чітко. У послідовності дій для вирішення описаної задачі можна виділити три етапи: визначення людей на кадрах відеопотоку та побудова детекцій; детектування осіб; розпізнавання статі та віку по особі.

Виявлення особи (face detection) є досить відомим завданням, у якого є множина різних підходів. Для роботи системі необхідно обмежувати область пошуку особи. Це необхідно для можливості однозначного зіставлення виявленої особи з людиною. Так само це рішення дозволить прискорити розробку та загальну швидкість роботи системи, оскільки відсутня необхідність у написанні та тестуванні додаткового модуля в порівнянні обмежуючих вікон детектованих осіб та обмежуючих вікон пішоходів. Нижче розписані підходи до задачі детектування особи на кадрі.

Класичні методи визначення осіб. Підходи, які широко використовувалися на початку 2000х років. Використовувалися, тому числі, класичні статистичні методи, аналіз головних компонентів, лінійні та нелінійні класифікатори. Також використовувалися методи зображення, що відображають на площину і за допомогою знаходження відстані Махаланобіса [1], знаходили ступінь близькості поточного виявленого об'єкта з референс вектор обличчя. Незважаючи на хорошу теоретичну базу для відповідних методів, вони майже не застосовувалися в реальних промислових застосунках. Проблемою була вкрай низька здатність подібних підходів адаптуватись до інваріантності середовища.

Методи, засновані на застосуванні нейронних мереж. Деер CNN добре себе зарекомендували в завданні детектування об'єктів на зображеннях. Вони здатні визначати характерні риси об'єкта безпосередньо з пікселів зображення [2]. До найвідоміших архітектур CNN можна віднести AlexNet, GoogLeNet, VGGNet, ResNet. Для навчання CNN потрібен величезний обсяг даних, тому для застосування їх до вузькоспрямованих завдань, зазвичай, використовують передбачені на відомих датасетах моделі, а потім, застосовуючи метод переносу навчання (transfer learning), розраховують ваги на останніх шарах мережі. Підходи на основі CNN є найбільш точними і в той же час найвибагливішими до обчислювальних потужностей.

Методи, засновані на вейвлетах Хаара та Габора. За допомогою вейвлетів із зображення отримували найбільш сильні сигнали, фактично виділяючи найважливіші ознаки на зображенні, відкидаючи інше. Вейвлети Габора були представлені в дослідженнях алгоритмів розпізнавання осіб. Різні вейвлети застосовувалися з ідеєю виділення візуальних ознак незалежно від завдання. Надалі дані візуальні ознаки використовувалися в авторських алгоритмах, у тому числі і для знаходження особи, як, наприклад, в алгоритмі Віоли-Джонса. Використання ознак, отриманих вейвлетом Габора, є складним та часо затратним. Існуючі спрощені вейвлети та алгоритми на їх основі працюють швидше, проте сильно залежать від освітлення та інших зовнішніх факторів.

Методи, що ґрунтуються на знаходженні дескрипторів особи. Замість використання всього зображення алгоритми розбивають його на підобласті, у кожній з яких розраховують гістограми і знаходять найбільш яскраві та важливі ознаки для вирішуваної на даний момент задачі.

Надалі отримані ознаки можуть бути використані для завдання визначення особи, так і в задачі ідентифікації, якщо є приклад, для якого також можна отримати ознаки подібним способом. Наприклад, автори застосовували підхід на основі локальних дескрипторів, включаючи різні локальні бінарні шаблони, для завдання ідентифікації по обличчю. При цьому знаходилися такі дескриптори, які були стійкі до вікових змін особи. Дані методи компактні та швидко працюють, добре справляються із змінами у висвітленні та різними емоціями на обличчі. Однак ці підходи дуже чутливі до шуму на зображенні.

Методи, що базуються на тривимірних зображеннях та картах глибини. Багато сучасних камер, і навіть смартфонів, оснащені лідарами і мають можливість швидко та точно отримувати карту глибини зображення. Цей метод зйомки дозволяє значно знизити чутливість алгоритмів до освітлення та зовнішніх факторів зйомки. Моделі, що працюють із тривимірними зображеннями, набагато більш точні та інваріантні до змін ніж працюючі з двовимірними. Методи використовують текстури зображення, компенсують варіативне освітлення, а також компенсують різні пози людини за допомогою прагнення до симетричного обличчя особи під час аналізу. Такі системи, що вимагають особливого обладнання, ретельного калібрування та синхронізації даних, що робить їх складно застосовними в системах загального призначення.

Визначення статі та віку. Якщо визначення статі є завданням бінарної класифікації, то завдання визначення віку можна вирішувати як мультикласову класифікацію за віковими кластерам, так і як завдання регресійного передбачення кількості років. Наприклад удля визначення віку автори донавели модель VGG-face методом перенесення навчання на наборі даних із зображеннями людей, для кожного з яких була мітка про вік. На вихідному шарі нейронної мережі було 8 виходів, відповідних до вікових груп. У для завдання визначення віку також використовувалися CNN (Ranking -CNN), але з бінарним виходом, а проставлення мітки про вік людини відбувалося агрегацією виходів «базових» нейромереж. Кожна з базових згорткових мереж була навчена на мітках приналежності людини до групи «старші» або «молодші» певного віку.

Якщо для завдання визначення особи на кадрі методи не на основі нейромереж мають досить хорошу точність і можуть скласти конкуренцію CNN, то для завдання визначення віку по особі згорткові мережі є неперевершеними лідерами. Так, у 2015 році було представлено мережу AgeNet, що демонструє високу точність у завданнях визначення віку по особі та посіла призове місце у змаганні «ChaLearn 2015 apparent age competition». Підхід є глибокою згортковою нейронною мережею, в якій як функція втрат використовуються як Евклідова норма, так і крос-ентропія. Більшість сучасних методів для визначення статі та віку використовують відомі архітектури глибоких згорткових нейромереж, таких як VGG-16 або ImageNet.

Література

1. Perlibakas V. Distance measures for PCA-based face recognition / V. Perlibakas // Pattern Recognition Letters. 2004. Vol. 25(6). Pp. 711–724.
2. Liu W. A survey of deep neural network architectures and their applications / W. Liu, Z. Wang, X. Liu [et al.] // Neurocomputing. 2017. Vol. 234. Pp. 11–26.