

ВИКОРИСТАННЯ МОДЕЛІ ПОТОЧНИХ МАСОК ТА КЛАСИФІКАЦІЇ МЕТОДОМ SVM У ЗАДАЧІ РОЗПІЗНАВАННЯ ЕМОЦІЙ НА ОБЛИЧЧІ ЛЮДИНИ

Ю.В.Крак, О.В.Степанов

Інститут кібернетики ім.В.М. Глушкова НАН України,
03680, МСП, Київ, проспект Академіка Глушкова, 40
Yuri.krak@gmail.com, yltastep@gmail.com

Abstract

The article discusses the Active appearance model and Support Vector Machines to solve the problem of recognition of basic emotions on the human face. The proposed combination of techniques allows to solve the problem accurately. Active appearance model allows to automatically extract the main contours of the face on the video image, the SVM method allows to classify emotions of a person based on the location of points on the face contours.

1. Вступ та постановка задачі

Методи та алгоритми автоматичного аналізу емоційного стану обличчя людини є складовою частиною систем і засобів штучного інтелекту, які спрямовані на дослідження, створення та впровадження алгоритмічних та програмно-апаратних систем і комплексів з елементами штучного інтелекту на основі моделювання інтелектуальної діяльності людини. Моделювання та розпізнавання емоцій, як одного з каналів невербальної сигнальної та регулятивної передачі інформації, що відтворює динаміку внутрішніх переживань людини, є актуальним і важливим напрямом досліджень з метою створення систем комп'ютерного розпізнавання і синтезу зорових образів.

В даному дослідженні формальна модель емоцій отримала подальший розвиток для мімічних відображень емоційних станів людини. Запропоновано використовувати модель поточних масок (Active Appearance Model, ААМ) для автоматичного вилучення ознак емоційного стану обличчя [1]. З вилучених контурів обличчя будується вектор особливостей. Цей вектор будується виходячи з принципів системи кодування станів обличчя або FACS системи запропонованої Полом Екманом [2]. Задача визначення емоції розглядається як проблема класифікації за вектором особливостей. Навчальна вибірка складається з векторів вилучених з бази даних зображень емоцій 108 суб'єктів. В якості алгоритму класифікації використовується SVM. Побудований класифікатор в подальшому використовується для класифікації емоції обличчя людини з відео потоку.

Результати цього дослідження можуть бути використанні для моделювання інтелектуальної діяльності людини, застосування в системах штучного інтелекту, як складова при розробці алгоритмів і програмно-апаратних засобів для систем комп'ютерного розпізнавання та відтворення зорових образів а також для створення

високоінтелектуальних мультимедійних технологій і засобів для систем штучного інтелекту, тощо.

2. Модель поточних масок

Найбільш типовою параметризацією деформації візуального представлення об'єкта є моделювання розподілу точок (Point Distribution Model, PDM) [3]. Вона передбачає лінійну модель надходження нежорстких варіацій об'єкта:

$$s = \bar{S} + \Phi q, \quad (1)$$

де s позначає координати орієнтирів контуру у поточному кадрі, \bar{S} позначає стандартні (середні) положення точок контуру, Φ є матриця, стовпці якої складаються з напрямків зміни положень, і q - параметри моделі.

У моделюванні обличчя використовується трьохвимірна модель контуру, тобто:

$$s = [x_1, y_1, z_1, \dots, x_n, y_n, z_n] \quad (2)$$

\bar{S} і Φ знаходять за допомогою моделювання нежорсткої структури контуру при русі по набору 2D форм [4]. Форма контуру об'єкта в кадрі, тобто координати точок, знаходяться шляхом проектування s на зображення з відповідними масштабуванням, поворотом та переносом.

$$x = [x_1, \dots, x_n] = \alpha (I \otimes R) s + 1 \otimes [t_x, t_y], \quad (3)$$

де α позначає глобальний коефіцієнт масштабу, $[t_x, t_y]$

позначає глобальний перенос, R позначає спрощену матрицю повороту (використовується модель слабкої перспективи). Усі параметри даної моделі збираються до одного вектора

$$p = \{\alpha, R, t_x, t_y, q\} \quad (4)$$

Підходи на основі околів зображення до моделі підгонки деформації зазвичай будуються на локальному пошук кращого розташування кожної точки контуру, яка потім піддається параметризації PDM. Це досягається шляхом мінімізації квадратів нев'язки:

$$Q(p) = \sum_{i=1}^n \|x_i - \mu_i\|^2 w_i, \quad (5)$$

де $\{\mu_i\}_{i=1}^n$ позначає координати точок контуру знайдених процедурою локального пошуку, $\{w_i\}_{i=1}^n$ - вагові коефіцієнти, які представляють важливість кожної окремої точки.

Подальша мінімізація функціоналу (5) здійснюється за допомогою локальних процедур пошуку, ЕМ-алгоритму та лінійних апроксимацій складних функцій залежності [1, 3-6].

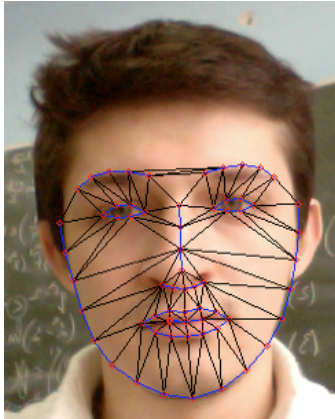


Рис. 1. Приклад роботи методу поточних масок.

3. Вилучення вектору особливостей

Модель поточних масок, описана у розділі 2, дає нам розташування точок, які описують контури обличчя. У розділі 1 було показано, що емоції людини виражаються через зміну напруження м'язів обличчя, та зміну форми контурів, як наслідок. Фактично, зміну контурів можна описати відстанню між визначеними точками, яка відображає напруження відповідного м'язу обличчя.

Використовуючи розроблену П. Екманом модель FACS [2, 7], було запропоноване кодування одиниць руху наведено у Таблиці 1

Таблиця 1.

Одиниці руху	Опис	ID особливості
AU1, AU2, AU4	Внутрішній рух брів між собою, рух між брів та очей	V1, V2, V3, H1
AU5, AU7, AU41, AU42, AU43, AU44, AU45, AU46	Внутрішній рух очей	V6, V7, H2, H3
AU6	Рух щік	H8
AU9	Внутрішній рух носу	V8, H4
AU10	Рух між носом та верхньою губою	V13
AU12, AU15, AU16, AU18, AU20, AU22, AU23, AU24, AU25, AU28	Внутрішній рух рта, зміна форми губ	V9, V10, V11, V12, H5, H6, H7
AU27	Рух між носом та нижньою губою	V5
AU17, AU26	Рух між носом та підборіддям	V4

На рис 2 наведена ілюстрація особливостей. Відмітимо, що H_{η} та V_{η} це нормалізуючі відстані для горизонтальних та вертикальних особливостей. Розділення на горизонтальні та вертикальні особливості дозволяє

унікнити варіації зв'язаних з різноманітністю форм обличчя.

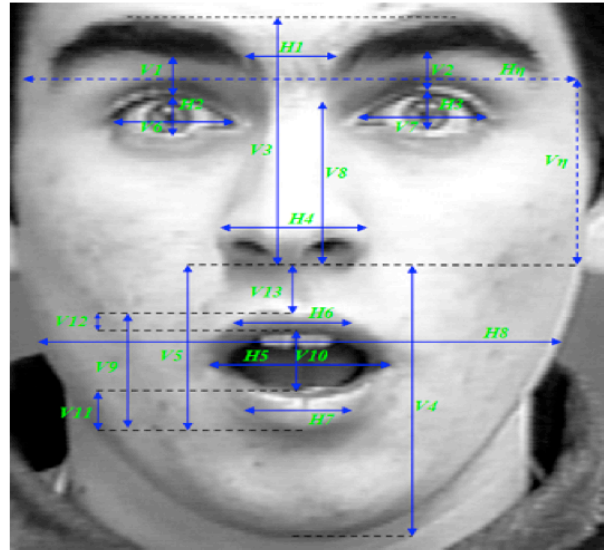


Рис 2. Ілюстрація особливостей, що базуються на одиницях руху FACS, та вилучаються з моделі поточних масок

Таким чином, з моделі поточних масок вилучається вектор розмірності 22, який й буде використовуватися для класифікації.

4. Класифікатор емоційного стану на базі SVM

Метод опорних векторів (SVM), запропонований радянськими вченими В. Н. Вапником та О. Я. Червоненкісом, зарекомендував себе як один з найпотужніших методів в розпізнаванні образів. Класичний SVM це побудова двох-класового класифікатора. Однак дана методика може бути узагальнена на декілька класів. Існує декілька підходів до формалізації метода SVM – C-SVM, та ν -SVM [8].

В нашій роботі використовувалась бібліотека LibSVM [9], що входить до стандартної поставки фреймворка OpenCV. В даній бібліотеці реалізовано розв'язання обох зазначених підходів - ν -SVM та C-SVM. Розв'язання даних задач здійснюється алгоритмом градієнтного спуску.

Також, в даній бібліотеці реалізовані варіації даних методів з заміною скалярного добутку на ядерні функції. У LibSVM реалізовані наступні найбільш популярні функції ядра :

- Радиально-базисна функція (RBF)

$$K(x, y) = \exp\{-\gamma \|x - y\|^2\} \quad (6)$$

- Сигмоїдна функція (SIGMOID)

$$K(x, y) = th(k_0 + k_1 \langle x, y \rangle) \quad (7)$$

Таким чином, маємо 6 різних варіацій методу SVM. Що стосується мульти-класової (більше 2) класифікації методом SVM, то існує декілька підходів. У LibSVM реалізований підхід "один проти одного". Суть його полягає в побудові C_N^2 класифікаторів (N – кількість класів) кожної пари класів. Далі, розв'язуюче правило

підраховується для кожного з класифікаторів, та за відповідь приймається той клас, що набрав більше всіх голосів.

В якості тренувальної виборки було використано базу зображень Bosphorus [10]. Дана база складається з фотографій 105 суб'єктів з багатою різноманітністю напружень м'язів обличчя, позицій (повороту), перекритій обличчя (волосся, окуляри і т.п.). Дана база була створена саме з метою тренування та тестування алгоритмів оцінки емоцій обличчя, вона містить зображення кожної важливої одиниці руху (AU) для кожного суб'єкта, а також окремо зображення 4 базових емоцій - радість, злість, подив та відраза.

Кожна фотографія оброблялася моделлю поточних масок, причому маска підгонялася декілька ітерацій (спочатку детектинг, потім трекінг) для стабілізації контурів.

Специфіка даних зображень в тому, що обличчя іноді обрізані та не мають границь, контурів овалу обличчя. Також, деякі форми вусів та бороди викликають хибне визначення контурів. Отже, було вирішено фільтрувати правильно накладені маски від хибних вручну в режимі реального часу на етапі побудови навчальної вибірки. Таким чином було отримано навчальну вибірку, структура якої наведена у таблиці 2.

Таблиця 2.

Емоція	Кількість вилучених векторів особливостей
Нейтрал	65
Радість	70
Злість	36
Подив	30
Відраза	18
Всього	219

Результати побудови класифікаторів показали, що емоція відрази має найгірші результати, завдяки малому розміру тренувальної виборки та із-за подібності контурів (більш глибокий аналіз показав, що вектори відрази класифікувалися емоціями злості, радості та подиву в порядку зпадання частоти). Зазначимо, що при навчанні параметри моделі підбиралися методом проходження по сітці, тренувальна вибірка розділялась на частини по 80% та 20%, на першій класифікатор навчався, на другій тестувався. Обиралася модель, що показала найкращий результат класифікації у тестуванні.

Також, треба відмітити, що алгоритм v-SVM показав кращі результати ніж C-SVM, звичайний лінійний метод та з радіально-базисною функцією ядра приблизно однакові результати, а ось на базі сигмоїда погану якість класифікації. Для подальших тестів, був обраний класифікатор LINEAR v-SVM.

В силу того, що емоція відрази показала не досить якісний результат, було вирішено побудувати ще один класифікатор для 3-х базових емоцій, вилучивши данні емоції відрази. Треба відмітити неістотне підвищення якості класифікації на тренувальних виборках, скоріше за все завдяки аномальних даних у навчальній вибірці. Це зв'язано з неправильно накладеною маскою при вилученні вектору особливостей на деяких зображеннях.

Однак, даний варіант класифікації 3 емоцій та нейтрального стану, показав досить не поганий результат на реальних даних з веб-камери.

5. Висновки

У даній роботі реалізований алгоритм класифікації базових емоцій на обличчі людини використовуючий модель поточних масок. Запропонований алгоритм вилучення вектору особливостей з маски обличчя що базується на одиницях руху (AU) моделі FACS – загально визнаного стандарту в області оцінки емоційних станів обличчя.

В якості алгоритму класифікації був використаний SVM, який показав достатньо стійкі результати на навчальній вибірці та реальному тестуванні. Було розглянуто два варіанти постановки моделі SVM (C-SVM та v-SVM), та 3 різні функції ядра (лінійна, радіально-базисна та сигмоїд) – всього 6 варіантів. Було проведено ряд експериментів по побудові та тестуванні класифікатора на навчальній вибірці, в яких найкращі результати показали лінійна та радіально-базисна функція. В силу того, що перша є більш обчислювально ефективною, було прийнято рішення використовувати C-SVM з лінійною функцією ядра. Дана модель була протестована на реальних даних онлайн-потоків з веб-камери для декількох суб'єктів, та показала достатній рівень якості класифікації.

6. Література

- [1] J. Saragih, S. Lucey and J. Cohn, "Deformable Model Fitting by Regularized Landmark Mean-Shifts", International Journal of Computer Vision (IJCV), 91(2): 200–215, 2010.
- [2] P. Ekman, W. Friesen, and J. Hager. The Facial action coding system, Research Nexus eBook, Salt Lake City, UT. 2002.
- [3] T. F. Cootes and C. J. Taylor. Active Shape Models - 'Smart Snakes'. In BMVC, pages 266–275, 1992.
- [4] L. Torresani, A. Hertzmann, and C. Bregler. Nonrigid Structure-from-Motion: Estimating Shape and Motion with Hierarchical Priors. PAMI, 2008.
- [5] Y. Wang, S. Lucey and J. Cohn. Enforcing Convexity for Improved Alignment with Constrained Local Models. In CVPR, 2008.
- [6] M. I. Jordan and R. A. Jacobs. Hierarchical Mixtures of Experts and the EM Algorithm. Neural Computation 6, page 181-214, 1994.
- [7] P. Ekman, W. Friesen, and J. Hager. The Facial action coding system, Research Nexus eBook, Salt Lake City, UT. 2002.
- [8] C.-C. Chang and C.-J. Lin. Training v-support vector classifiers: Theory and algorithms. Neural Computation, 13(9):2119–2147, 2001.
- [9] LibSVM Documentation: <http://www.csie.ntu.edu.tw/~cjlin/papers/libsvm.pdf>
- [10] A. Savran, B. Sankur, M. T. Bilge, "Comparative Evaluation of 3D versus 2D Modality for Automatic Detection of Facial Action Units", Pattern Recognition, Vol. 45, Issue 2, p: 767-782, Feb. 2012.