



РАСПОЗНАВАНИЕ ЭМОЦИЙ НА ВИДЕО КАК ИНСТРУМЕНТ ЧЕЛОВЕКО-МАШИННОГО ИНТЕРФЕЙСА

Кулишова Н.Е., профессор, кафедра МСТ ХНУРЭ

Существующие методы человеко-машинного взаимодействия зачастую имеют интуитивную природу. Как следствие, необходимо затрачивать какое-то время для адаптации пользователя к программе или компьютеру. Чтобы сделать эту ситуацию более объективной, необходимо включить в интерфейс обратную связь в виде информации о реакции пользователя на действие программы. Это может быть информация, полученная по человеческим коммуникативным каналам – речь, жесты или выражения лица.

В [1] отмечено, что только 7% информации в человеческих коммуникациях несет вербальное общение (речь), 38% – обеспечивается модуляциями голоса, и 55% – выражением лица. Распознавание и классификация эмоций – важное направление для многочисленных исследований в области компьютерного зрения.

Цель данной работы – разработка структуры классификатора, который входит в состав системы автоматического распознавания эмоций человека на видео в реальном времени.

Решение проблемы автоматического распознавания эмоции человека включает несколько основных этапов [2, 3]:

- предварительная обработка изображения;
- выделение области лица на изображении;
- выделение характерных черт на лице, определение их расположения;
- распознавание эмоции на основе значений характерных признаков, описывающих расположение характерных черт лица.

Предварительная обработка изображения необходима, чтобы исключить влияние шумов, особенностей освещения, расположения головы человека в кадре, поворота головы или эффекта движения на результаты распознавания. Следующие процедуры обработки возможны только после выделения области лица в изображении, поскольку связаны с определением размеров этой области, ее конфигурации и т.д.

Для выделения области лица можно использовать сегментацию по цвету, поиск на изображении совокупности геометрических признаков, которые могут быть отнесены только к чертам лица. Количество характерных признаков в области лица может достигать тридцати. К ним относят, например: брови, глаза, нос и рот, их размеры, координаты их центров, координаты областей лба, щек, подбородка, углов рта, глаз, крыльев носа и т.д. [3]. Некоторые подходы к распознаванию используют трехмерную модель лица для объединения набора признаков в общее описание.

Следующий этап – распознавание мимики – означает узнавание привычных объектов среди множества новых. Привычными объектами при



узнавании эмоций служат тестовые образцы с изображением лиц, для которых эмоции однозначно определены. В настоящее время в качестве таких образцов используются многочисленные базы изображений. В системе распознавания текущее изображение лица сравнивается с эталонным (нейтральное эмоциональное состояние), и на основе результатов сравнения формируется решение о характере эмоции (рис. 1). Эту задачу выполняет классификатор, входящий в состав системы.

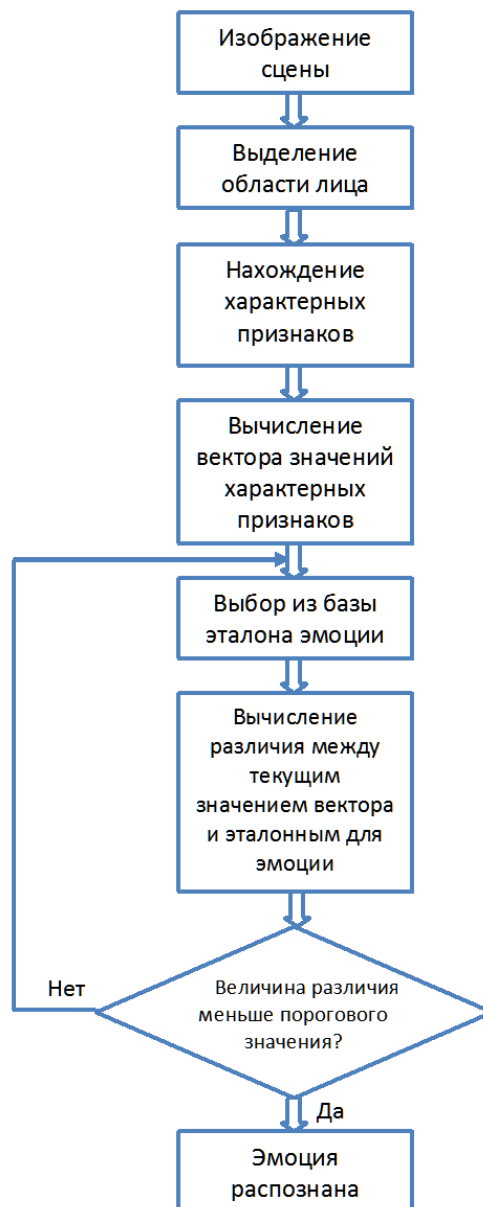


Рисунок 1 – Схема процесса автоматического распознавания эмоции на изображении лица человека

Однако эмоции человека имеют не статический, а динамический характер, они достаточно быстро изменяются, и сама скорость их изменения тоже может служить показателем эмоционального состояния.

Для динамического распознавания эмоций на видео предлагается использовать самоорганизующиеся карты Кохонена. Самоорганизующаяся



карта имеет простую архитектуру с прямой передачей информации и кроме нулевого рецепторного слоя содержит единственный слой нейронов. Каждый нейрон из этого слоя связан с каждым рецептором нулевого слоя прямыми связями и со всеми остальными нейронами поперечными внутрислойными связями. На каждый входной сигнал формируются группы нейронов с максимальным откликом, что позволяет карте Кохонена успешно решать задачи кластеризации при максимальном быстродействии.

Самоорганизующиеся карты Кохонена сохраняют топологию данных, что очень важно именно при обработке, как изображений, так и видео. Однако, они не обеспечивают высокой точности кластеризации. Чтобы решить эту проблему, предлагается на основе самоорганизующихся карт сформировать кластерный ансамбль [4], объединенный в рамках действия алгоритма самообучения:

- генерация случайных значений центров кластеров;
- инициализация функции вычисления расстояний между точками данных;
- отображение данных в пространство самоорганизующейся карты Кохонена;
- вычисление откликов нейронов карты; обнаружение группы нейронов с максимальным откликом;
- генерация кластера данных;
- назначение меток сформированным кластерам;
- осуществление обучения сформированной карты на наборе данных;
- генерация объединенной структуры кластера и соответствующего классификатора в ансамбле с созданием собственной метки;
- вычисление ошибки классификации части ансамбля;
- объединение ансамбля;
- классификация набора данных.

Такая архитектура позволяет с высокой точностью выполнять сегментацию областей на видео в реальном времени.

Список литературы

1. Mehrabian, A. Communication without words / A. Mehrabian // *Psychology Today*. – 1968. – vol. 2. – No. 4. – Pp. 53-56.
2. Khandait, S.P. Automatic Facial Feature Extraction and Expression Recognition based on Neural Network / S.P. Khandait, R.C. Tool, P.D. Khandait // *Int. J. on Advanced Computer Science and Applications*. – 2011. – vol. 2. – No. 1. – Pp. 113-118.
3. Fasel, B. Automatic facial expression analysis: survey / B. Fasel, J. Luettin // *Pattern Recognition*. 2003. – vol. 36. – Pp. 259-275.
4. Rathore, D. Design Hybrid method for intrusion detection using Ensemble cluster classification and SOM network / D. Rathore, A. Jain // *Int. Journal of Advanced Computer Research*. – 2012. – Vol. 2. – Num. 3. – P.181-186.