

УДК 004.932

**РАЗРАБОТКА МУЛЬТИМОДАЛЬНОЙ СИСТЕМЫ КЛАССИФИКАЦИИ ЭМОЦИЙ ЧЕЛОВЕКА
СРЕДСТВАМИ МАШИННОГО ОБУЧЕНИЯ**

Покасова А.И. (Университет ИТМО, Санкт-Петербург)

Научный руководитель – к.т.н, доцент Коржук В.М.

(Университет ИТМО, Санкт-Петербург)

Рынок систем детекции и распознавания эмоций активно развивается. По оценкам ряда экспертов, он продемонстрирует среднегодовой рост в 27,4% и достигнет планки 29,1 млрд долларов в 2022 году.

Предлагаемые на рынке системы распознавания эмоций обладают схожими недостатками: проприетарность решения, что высказывается в высокой стоимости приобретения системы, ориентирование на детекцию и анализ эмоций исключительно по одной модальности: видео- или аудиопоток. Это приводит к заключению о необходимости разработки системы мультимодального анализа эмоций человека.

Введение. Человек выражает свои эмоции и намерения самыми разными методами: изменяя выражение лица, используя жесты, окрашивая свою речь. Из вышеперечисленных способов лицевая передача является одним из ярчайших и быстрейших видов проявления намерений, речевая же занимает второе место по скорости выражения. В обоих типах проявления эмоций выделяют экспрессию базовую и комбинированную.

Одной из областей применения методов автоматического распознавания эмоций является обеспечение безопасности людей с помощью автоматизированных охранных систем. Современные охранные системы часто имеют в своем составе средства регистрации и анализа видеоданных. Однако, как правило, в таких системах решение принимается человеком-оператором. Это может привести к снижению точности и оперативности реагирования на различные ситуации, связанные с поведением людей на охраняемых территориях. Повысить эффективность охранных систем можно за счет повышения степени автоматизации процедур, обеспечивающих анализ изображений.

Среди других важных задач, решение которых необходимо обеспечить в конкретных охранных системах, можно выделить обнаружение и сопровождение людей, проявляющих эмоции, характерные для нарушителей правопорядка, психически больных, террористов и т.д.

Основная часть. В работе проведен обзор существующего на рынке программного обеспечения, предназначенного для распознавания эмоций в видеопотоке и по записи речи. Большинство решений, доступных конечному пользователю, нацелены на распознавание только в видеопотоке и исключительно из фронтального положения лица. Решения, предлагающие мультимодальный анализ, отсутствуют.

Описаны два подхода к категоризации человеческой экспрессии, как лицевой, так и речевой: системы СКЛИД и “Валентность-активация”. В виду особенностей выделения маркеров, отвечающих за конкретную эмоцию, первый подход будет использован при анализе видеоряда, а второй – звукоряда. Рассмотрены четыре больших класса методов распознавания эмоций в видеопотоке: холистические, локальные распознавания, динамические, геометрические. Выделены их преимущества и недостатки. Аналогичным образом рассмотрены методы классификации эмоций в звукоряде.

Предложена система мультимодальной классификации эмоций из видео- и аудиопотока, учитывающая нефронтальное расположение лица и зашумленность звукоряда, использующий сверточные и рекуррентные нейронные сети. Подробно описана архитектура используемых сетей.

Процесс обработки видеопотока поделен на следующие этапы: выделение лица, подавление шумов в изображении, определение угла поворота головы, финальное определение эмоции. Звуковой ряд очищается от шумов и подается на анализ в нейронную сеть.

Предложенный метод обладает повышенной точностью по сравнению с имеющимися решениями в случае классификации при угле поворота головы более 45 градусов.

Выводы. Разработанная система может быть интегрирована в охранные системы, обладающие средствами регистрации и анализа видео- и аудиоданных, предоставляя таким образом автоматизацию процедур обеспечения безопасности.

Покасова А.И. (автор)

Подпись

Коржук В.М. (научный руководитель)

Подпись