

УДК 004.8

DOI: 10.18413/2518-1092-2022-7-2-0-8

Медведев А.А.
Лаптев А.А.

**АЛГОРИТМ ВЫЯВЛЕНИЯ НЕВЕРБАЛЬНЫХ МАРКЕРОВ
ПОВЕДЕНИЯ ЧЕЛОВЕКА НА ВИДЕО**

Федеральное государственное автономное образовательное учреждение высшего образования «Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики», Кронверкский пр., д. 49, г. Санкт-Петербург, 197101, Россия
e-mail: anatoimdvdv@gmail.com, nickname.avast@gmail.com

Аннотация

Микровыражения – бессознательные, кратковременные невербальные сигналы, которые позволяют определить эмоциональное состояние человека. Они возникают, когда человек блокирует свои эмоции или скрывает истинные намерения. Определение невербальных сигналов становится актуальной задачей в ситуациях, где ложь или сокрытие информации приводят к ресурсным или финансовым потерям, влияют на безопасность и здоровье других людей. Возросшее количество онлайн-конференций открывает возможности программной обработки видеоканала выступления человека для анализа его эмоций и поведения с целью выявления конгруэнтности или противоречивости высказываний человека. В статье рассматриваются методы компьютерного зрения и машинного обучения, которые позволяют извлекать и анализировать лицо человека по видеоканалу для определения его невербальных маркеров и эмоционального состояния. Подробно рассмотрены метод лицевых ориентиров, особых точек лица, классификация эмоций человека по лицевым ориентирам, детекция морганий и отворачиваний человека во время выступления.

Ключевые слова: невербальные сигналы; детекция лица; лицевые ориентиры; отслеживание взгляда; классификация эмоций; машинное обучение.

Для цитирования: Медведев А.А., Лаптев А.А. Алгоритм выявления невербальных маркеров поведения человека на видео // Научный результат. Информационные технологии. – Т.7, №2, 2022. – С. 58-64. DOI: 10.18413/2518-1092-2022-7-2-0-8

Medvedev A.A.
Lapteva A.A.

**AN ALGORITHM FOR DETECTING NON-VERBAL MARKERS
OF HUMAN BEHAVIOR ON VIDEO**

Saint Petersburg National Research University of Information Technologies, Mechanics and Optics,
49 Kronverkskiy prospekt, St. Petersburg, 197101, Russia
e-mail: anatoimdvdv@gmail.com, nickname.avast@gmail.com

Abstract

Microexpressions are unconscious, short-term non-verbal signals that allow to determine the emotional state of a person. Microexpressions occur when a person blocks emotions or hides true intentions. Determining non-verbal signals becomes an urgent task in situations where lying or hiding information leads to resource or financial losses, affects the safety and health of other people. The spread of online conferences opens up the possibility of programmatic processing of a human speech video channel to analyze emotions and behavior in order to identify the congruence or inconsistency of person's statements. The article discusses computer vision and machine learning methods that allow extracting and analyzing person's face from a video channel to determine its non-verbal markers and emotional state. The method of facial landmarks, key points of the face, classification of human emotions by facial landmarks, detection of blinking and turning of a person during speech are considered in detail.

Keywords: non-verbal signals; face detection; facial landmarks; eye tracking; emotion classification; machine learning

For citation: Medvedev A.A., Laptev A.A. An algorithm for detecting non-verbal markers of human behavior on video // Research result. Information technologies. – Т.7, №2, 2022. – P. 58-64. DOI: 10.18413/2518-1092-2022-7-2-0-8

ВВЕДЕНИЕ

Общение людей состоит из вербальных и невербальных признаков. Невербальными являются сознательные и бессознательные сигналы, которые раскрывают поведение человека, дополняют или заменяют классические средства общения [1-5]. Распространение онлайн конференций открывает доступ к инструментам обработки и анализа материалов выступлений людей по видео, аудио и текстовым каналам. Анализ невербальных признаков позволяет найти взаимосвязь между речью человека и проявлениями бессознательных движений тела и мимики.

Ложь, сокрытие фактов провоцируют людей совершать противоречивые действия, блокировать свои эмоции [6]. Микровыражения — это невербальные сигналы, которые возникают, когда человек скрывает истинные чувства или намерения. Сравнительный анализ вербальных и невербальных компонент коммуникации позволяет оценить конгруэнтность поведения человека, выделить взаимосвязь или расхождение между высказываниями и поведением человека [7].

Одним из методов определения микровыражений человека является отслеживание движений мимики лица, взаимосвязи между мимической активностью и словами человека. Для распознавания мимики человека используется метод лицевых ориентиров, множества ключевых точек лица для определения формы головы, областей глаз, бровей, носа и губ [8].

Детектирование области глаз и зрачков позволяет выделить зоны интереса человека, к которым он возвращается повторно [9]. Взгляд человека — это невербальный показатель внимания человека. Частое моргание, расширенные зрачки, нарушение контакта или резкие перемещения взгляда сигнализируют о внутреннем конфликте человека, что является фактором неискренности высказываний [10]. Системы регистрации взгляда способны воссоздать траекторию движений глаз пользователя с помощью методов окулографии [11-12].

Кроме направления взгляда, при фокусном отслеживании лицевых ориентиров используется область рта человека [13]. Ориентиры губ применяются в задаче классификации эмоций человека. Улыбка, открытость или напряженность губ релевантны для оценивания счастливых или депрессивных выражений, возникающих у человека. Губы служат индикатором яркого выражения эмоций человека, либо дополнительным критерием оценки его поведения. В случае анализа поведения улыбка может подтвердить или опровергнуть предсказание эмоции, например, при одновременном распознавании ориентиров рта и бровей человека.

Модели распознавания невербальных сигналов по видеопотоку могут быть применены для анализа стратегии поведения человека во время выступления и определения комбинаций маркеров перспективных для изучения эмоций. Результаты исследования [14] показали, что сочетание движений рук и мимики релевантно для оценки эффективности коммуникации. Количество и частота производимых человеком жестов и сигналов могут быть применены для подготовки к публичным выступлениям и интервью.

ОСНОВНАЯ ЧАСТЬ

Методы машинного обучения позволяют извлекать изображение лица человека из видеопотока независимо от формы, размера и положения головы. Увеличение репрезентативности обучающего набора данных повышает устойчивость моделей к фону и качеству изображений. Открытые наборы данных объединяют видео и изображения людей при различных внешних условиях, факторах вращения головой и направления взгляда, полагаясь на разнообразие данных в выборке, а не на их количество. Тем не менее, собрать единый репрезентативный набор данных невозможно, так как ситуаций и условий, в которых может оказаться человек, множество, в связи с чем возникает проблема сбора данных в естественных условиях.

Примерами наборов данных, основанных на методе лицевых ориентиров, являются наборы MPIGaze [15] и Helen [16]. Labeled Faces in the Wild Home позволяет решить проблему статичных

фонов и узкого диапазона условий [17]. Набор данных состоит из фотографий, сделанных респондентами в естественных условиях.

Для детекции лица человека на видео была использована сверточная нейронная сеть BlazeFace [18] из фреймворка MediaPipe. Дополнительный модуль Face Mesh [19] позволяет распознавать лицевые ориентиры на полученном изображении. MediaPipe Face Mesh принимает на вход цветное изображение размером 128×128 пикселей и возвращает 468 пространственных ориентиров распознанного лица человека. Информация о взаимном расположении координат областей губ, бровей, глаз пользователя открывает возможности анализа мимики человека. Каждый лицевой ориентир содержит X , Y и Z координаты, где значения X и Y нормированы относительно ширины и высоты входного изображения. Координата Z отражает нормализованное расстояние между человеком и камерой. Анализ Z компоненты лица человека позволяет выявить глубину изображения и положение человека относительно камеры.

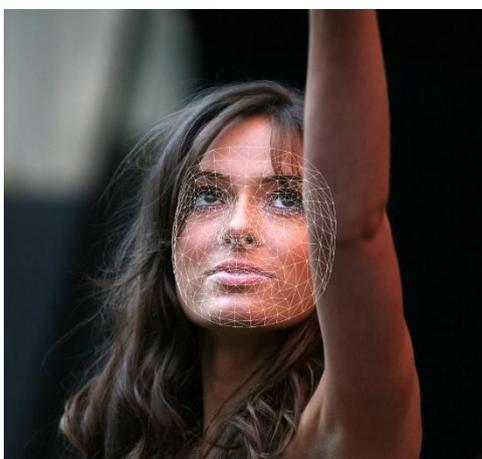


Рис. 1. Детекция лицевых ориентиров с помощью фреймворка MediaPipe [19] на примере набора данных Helen [16]

Fig. 1. Facial landmarks detection using MediaPipe [19] framework on a Helen [16] dataset sample

Изучение отдельных областей лица человека позволяет определить моменты мимической активности по движениям губ, бровей, перемещениям взгляда. С этой целью было рассчитано евклидово расстояние между лицевыми ориентирами и их центром масс. Изменение значений удаленности ориентиров от центра отражает смену микровыражений, возникающих на лице человека. Видеоаналитика выражений лица позволяет определить временные промежутки, в которые человек разговаривал, активно проявлял или скрывал свои эмоции, а также выявить мимические аномалии, которые могут быть отнесены к невербальным сигналам. Тем не менее, модели распознавания лицевых ориентиров возвращают предсказание координат, основанное на аппроксимации обучающего набора данных, а не их реальное положение. Отсутствие репрезентативных примеров выражения эмоций в исходных данных способно привести к неточной интерпретации мимики и паттернов поведения человека.

Отдельным преимуществом модуля Face Mesh является отслеживание положения зрачка человека. Модуль отслеживания зрачков человека был применен для построения карт плотностей распределения положений, в которых находился центр радужной оболочки относительно поверхности глаза.

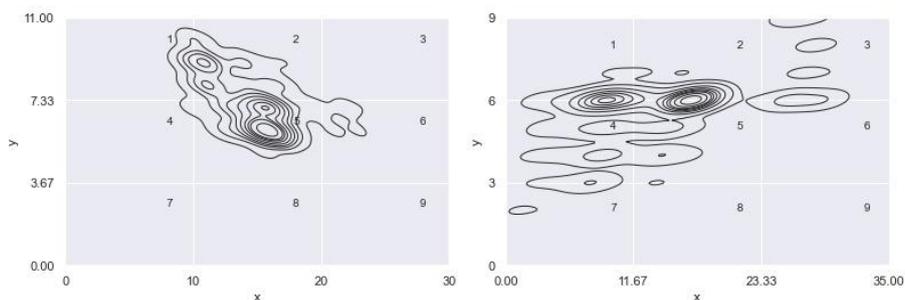


Рис. 2. Построение карт плотности распределения положений зрачков человека во время видеовыступлений

Fig. 2. Building density maps of human pupil positions distribution during videoconferencing

Статистический анализ распределения положений центра зрачка позволяет рассчитать количество отклонений зрачка от центра глаза, частоту и длительность отклонений, выделить области, в которых зрачок человека находился длительную часть времени или, наоборот, не направлялся. Согласно теории утечки эмоций, факт лжи вызывает дискомфорт у человека, что приводит к высокой глазодвигательной активности [1]. Сопоставление активности зрачков человека с видеофрагментами выступления позволяет выявить невербальные маркеры отвода или напряженности взгляда, что говорит о неискренности выступающего или сокрытии информации.

Другим примером невербальных сигналов, связанных с глазами человека, является количество и частота морганий человека. При эмоциональном подъеме движения глаз становятся произвольными, а количество актов морганий увеличивается. Сочетание показателей частоты морганий и изменения положений зрачка позволяет связать невербальную активность с эмоциями страха, гнева или восторга, которые испытывает человек [3]. Для распознавания акта моргания были использованы лицевые ориентиры глаз человека и рассчитано отношение горизонтальной протяженности разреза глаз к вертикальной. В момент моргания глаз человека закрывается, а значит вертикальная протяженность разреза глаз стремится к нулю. Как следствие, отношение горизонтали глаза к вертикали резко возрастает. Подсчет скачков взаимного отношения компонент разреза глаз человека позволяет определить количество и частоту морганий человека во время выступления.

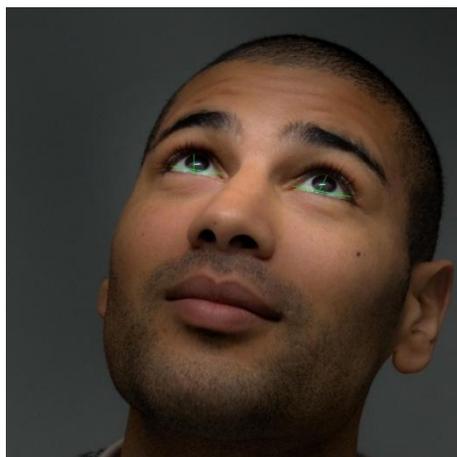


Рис. 3. Детекция разреза глаз для определения момента моргания на примере набора данных Helen [16]

Fig. 3. Eye shape detection to determine blinking moment on a Helen [16] dataset sample

Во время экспрессивной речи голова человека не занимает статичное положение относительно камеры. Частые вращения головой во время выступления или ответа на вопрос являются невербальным сигналом неискренности, неуверенности человека или блокирования

эмоций [1]. Для регистрации момента отворачивания человека были использованы пространственные координаты лицевых ориентиров. С этой целью лицевые ориентиры были нормализованы относительно высоты и ширины лица человека и центрированы в координатном пространстве. Таким образом, положение инвариантных координат лица описывается тремя ортогональными матрицами поворота вокруг осей декартовой системы координат. Преобразование координатного пространства позволило рассчитывать углы и направления вращений головы человека в каждый момент выступления. Частота отворачиваний, количество опусканий головы являются вспомогательными маркерами, которые отражают конгруэнтность поведения человека относительно его высказываний.

ЗАКЛЮЧЕНИЕ

Использование видео модальности для детектирования невербальных сигналов позволяет выявлять маркеры поведения человека, которые сложно определить непрофессионалу. Метод лицевых ориентиров открывает доступ к инструментам исследования положения и мимической активности лица человека. Тем не менее, методы компьютерного зрения и машинного обучения требуют наличия крупного и репрезентативного источника данных. Разработчикам и исследователям доступны открытые наборы данных, которые содержат снимки, сделанные в естественных условиях, фронтальные и нефронтальные, с разными уровнями освещенности и разрешениями камеры. Однако, не существует решения, которое объединило бы данные о всех информативных признаках. Комплексность задачи, уникальные области применения, такие как психология и неврология, медицинская статистика, профайлинг и ораторское искусство требуют индивидуальный подход к каждой из сфер. Системы психоэмоциональной оценки человека по видео прогрессируют в обнаружении и нормализации ориентиров для определения микровыражений, чему способствует распространение программных решений с открытым исходным кодом. В статье представлены результаты применения метода лицевых ориентиров и разработки модулей анализа и предобработки полученных признаков. Извлеченные невербальные характеристики позволяют программно расширить описание действий человека, сформировать паттерны поведения выступающего, а также оценить достоверность его высказываний.

Список литературы

1. Ekman, P., Friesen, W.V. Non-Verbal Leakage and Clues to Deception // *Social Encounters: Contributions to Social Interaction*. – 2017. – P. 132-148.
2. Ekman, P., Friesen W.V. *Hand Movements* // *Communication Theory: Second Edition*. – 2017. – P. 273-292.
3. Matsumoto, D., Hwang H.C., Ekman P. Subjective Experience and the Expression of Emotion in Humans // *Encyclopedia of Behavioral Neuroscience: Second Edition*. – 2021. – № 3-3. – P. 671-677.
4. Ekman, P., Friesen W.V. *The Repertoire of Nonverbal Behavior: Categories, Origins, Usage, and Coding* // *Mouton Classics: From Syntax to Cognition. From Phonology to Text*. – 2013. – P. 819-868.
5. Frank, M.G., Yarbrough, J.D., Ekman, P. Investigative Interviewing and the Detection of Deception // *Investigative Interviewing: Rights, Research and Regulation*. – 2013. – P. 229-255.
6. Shen, X., Fan, G., Niu, C., Chen, Z. Catching a Liar through Facial Expression of Fear // *Frontiers in Psychology*. – 2021. – № 12.
7. Goodman, G.S., Luten T.L., Edelstein R.S., Ekman, P. Detecting Lies in Children and Adults // *Law and Human Behavior*. – 2006. – № 30-1. – P. 1-10.
8. Zhang, K., Zhang, Z., Li, Z., Qiao, Y. Joint Face Detection and Alignment using Multitask Cascaded Convolutional Networks // *IEEE Signal Processing Letters*. – 2016. – № 23-10. – P. 1499-1503.
9. Zhang, X., Sugano, Y., Fritz, M., Bulling, A. It's Written all Over Your Face: Full-Face Appearance-Based Gaze Estimation // *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*. – 2017. – P. 2299-2308.
10. Frank, M. G., Yarbrough, J.D., Ekman, P. Detecting Lies in Children and Adults // *Investigative Interviewing and the Detection of Deception*. – 2013. – P. 229-255.

11. Krafska, K., Khosla, A., Kellnhofer, P., Kannan, H., Bhandarkar, S., Matusik, W., Torralba A. Eye Tracking for Everyone // Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. – 2016. – P. 2176-2184.
12. Tonsen, M., Zhang, X., Sugano, Y., Bulling, A. Labelled Pupils in the Wild: A Dataset for Studying Pupil Detection in Unconstrained Environments // Eye Tracking Research and Applications Symposium. – 2016. – № 14. – P. 139-142.
13. Kathi, M.G., Shaik, J.H. Estimating the Smile by Evaluating the Spread of Lips // Revue d'Intelligence Artificielle. – 2021. – № 35-2. – P. 153-158.
14. Pereira, M., Meng, H., Hone, K. Prediction of Communication Effectiveness during Media Skills Training using Commercial Automatic Non-Verbal Recognition Systems // Frontiers in Psychology. – 2021. – №12.
15. Zhang, X., Sugano, Y., Fritz, M., Bulling, A. Appearance-Based Gaze Estimation in the Wild // IEEE Conference on Computer Vision and Pattern Recognition. – 2015. – № 7.
16. Le, V., Brandt, J., Lin, Z., Bourdev, L., Huang, T.S. Interactive Facial Feature Localization // Lecture Notes in Computer Science. – 2012.
17. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. – 2007.
18. Bazarevsky, V., Kartynnik, Yu., Vakunov, A., Raveendran, K., Grundmann, M. BlazeFace: Sub-millisecond Neural Face Detection on Mobile GPUs // CVPR Workshop on Computer Vision for Augmented and Virtual Reality. – 2019.
19. Kartynnik, Yu., Ablavatski, A., Grishchenko, I., Grundmann, M. Real-time Facial Surface Geometry from Monocular Video on Mobile GPUs // CVPR Workshop on Computer Vision for Augmented and Virtual Reality. – 2019.

References

1. Ekman, P., Friesen, W.V. Non-Verbal Leakage and Clues to Deception // Social Encounters: Contributions to Social Interaction. – 2017. – P. 132-148.
2. Ekman, P., Friesen W.V. Hand Movements // Communication Theory: Second Edition. – 2017. – P. 273-292.
3. Matsumoto, D., Hwang H.C., Ekman P. Subjective Experience and the Expression of Emotion in Humans // Encyclopedia of Behavioral Neuroscience: Second Edition. – 2021. – № 3-3. – P. 671-677.
4. Ekman, P., Friesen W.V. The Repertoire of Nonverbal Behavior: Categories, Origins, Usage, and Coding // Mouton Classics: From Syntax to Cognition. From Phonology to Text. – 2013. – P. 819-868.
5. Frank, M.G., Yarbrough, J.D., Ekman, P. Investigative Interviewing and the Detection of Deception // Invstigative Interviewing: Rights, Research and Regulation. – 2013. – P. 229-255.
6. Shen, X., Fan, G., Niu, C., Chen, Z. Catching a Liar through Facial Expression of Fear // Frontiers in Psychology. – 2021. – № 12.
7. Goodman, G.S., Luten T.L., Edelstein R.S., Ekman, P. Detecting Lies in Children and Adults // Law and Human Behavior. – 2006. – № 30-1. – P. 1-10.
8. Zhang, K., Zhang, Z., Li, Z., Qiao, Y. Joint Face Detection and Alignment using Multitask Cascaded Convolutional Networks // IEEE Signal Processing Letters. – 2016. – № 23-10. – P. 1499-1503.
9. Zhang, X., Sugano, Y., Fritz, M., Bulling, A. It's Written all Over Your Face: Full-Face Appearance-Based Gaze Estimation // IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. – 2017. – P. 2299-2308.
10. Frank, M. G., Yarbrough, J.D., Ekman, P. Detecting Lies in Children and Adults // Investigative Interviewing and the Detection of Deception. – 2013. – P. 229-255.
11. Krafska, K., Khosla, A., Kellnhofer, P., Kannan, H., Bhandarkar, S., Matusik, W., Torralba A. Eye Tracking for Everyone // Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. – 2016. – P. 2176-2184.
12. Tonsen, M., Zhang, X., Sugano, Y., Bulling, A. Labelled Pupils in the Wild: A Dataset for Studying Pupil Detection in Unconstrained Environments // Eye Tracking Research and Applications Symposium. – 2016. – № 14. – P. 139-142.
13. Kathi, M.G., Shaik, J.H. Estimating the Smile by Evaluating the Spread of Lips // Revue d'Intelligence Artificielle. – 2021. – № 35-2. – P. 153-158.

14. Pereira, M., Meng, H., Hone, K. Prediction of Communication Effectiveness during Media Skills Training using Commercial Automatic Non-Verbal Recognition Systems // *Frontiers in Psychology*. – 2021. – №12.
15. Zhang, X., Sugano, Y., Fritz, M., Bulling, A. Appearance-Based Gaze Estimation in the Wild // *IEEE Conference on Computer Vision and Pattern Recognition*. – 2015. – № 7.
16. Le, V., Brandt, J., Lin, Z., Bourdev, L., Huang, T.S. Interactive Facial Feature Localization // *Lecture Notes in Computer Science*. – 2012.
17. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. – 2007.
18. Bazarevsky, V., Kartynnik, Yu., Vakunov, A., Raveendran, K., Grundmann, M. BlazeFace: Sub-millisecond Neural Face Detection on Mobile GPUs // *CVPR Workshop on Computer Vision for Augmented and Virtual Reality*. – 2019.
19. Kartynnik, Yu., Ablavatski, A., Grishchenko, I., Grundmann, M. Real-time Facial Surface Geometry from Monocular Video on Mobile GPUs // *CVPR Workshop on Computer Vision for Augmented and Virtual Reality*. – 2019.

Медведев Анатолий Андреевич, студент 1-го курса магистратуры, инженер Национального центра когнитивных разработок

Лаптев Андрей Александрович, аспирант, инженер Национального центра когнитивных разработок

Medvedev Anatoly Andreevich, 1st year Master's student, engineer, National Center for Cognitive Development

Laptev Andrey Aleksandrovich, postgraduate student, engineer, National Center for Cognitive Development