

УДК 004.93

ПАССИВНЫЕ ТЕХНОЛОГИИ МОДЕЛИРОВАНИЯ ГОЛОВЫ ЧЕЛОВЕКА**Бацева Д.А., Белов Ю.С.**

*Калужский филиал ФГБОУ ВО «Московский государственный технический университет им. Н.Э. Баумана (национальный исследовательский университет)», Калуга,
e-mail: wwdusya@gmail.com*

В стремлении к увеличению надежности и скоростей распознавания лиц создавались системы, которые использовали распознавание на основе 3D. Благодаря этому появился новый параметр – информация о глубине, с помощью которого возможно минимизировать влияние изменений позы и освещенности для достижения большей точности распознавания. В статье рассматриваются основные пассивные методы моделирования 3D-головы человека, которые используются на первом этапе в любой системе распознавания лиц с поддержкой 3D. Методы пассивного моделирования используют несколько изображений, которые были захвачены без датчиков. Существует два хорошо известных метода пассивной реконструкции – стереозрение, морфинг, определение формы по перемещению. Описываются сами алгоритмы, точность, преимущества и недостатки этих методов, а также их эффективность для распознавания трехмерного лица.

Ключевые слова: распознавание лиц, 3D-модель, 3D-реконструкция лица, стереозрение, морфинг, реконструкция по видео, shape-from-motion

PASSIVE TECHNOLOGIES OF HUMAN HEAD MODELING**Batseva D.A., Belov Y.S.**

*Branch of Moscow state technical University n.a. Bauman (national research University), Kaluga,
e-mail: wwdusya@gmail.com*

In its quest for more reliability and higher recognition rates the face recognition community has been focusing more and more on 3D based recognition. Depth information adds another dimension to facial features and provides ways to minimize the effects of pose and illumination variations for achieving greater recognition accuracy. This chapter reviews, therefore, the major techniques for 3D face modeling, the first step in any 3D assisted face recognition system. Passive modeling techniques use several images that were captured without sensors. The reviewed techniques are stereo vision, morphing, shape from motion. Concepts, accuracy, feasibility, and limitations of these techniques and their effectiveness for 3D face recognition are discussed.

Keywords: 3D face reconstruction, face recognition, stereo vision, morphing, shape from motion

Распознавание лица на неподвижном изображении – важная задача и часто она является первым шагом в алгоритмах приложений для реализации взаимодействия человек–компьютер. Целью таких алгоритмов является определение отсутствия или присутствия лица на произвольном изображении и, если оно есть, то необходимо выявить его положение и размер [3].

Технологии 3D-моделирования лица можно разделить на два класса – активные и пассивные – в зависимости от методов обработки изображений и методов дальнейшей реконструкции 3D-модели.

Существующие на данный момент активные системы 3D-распознавания лиц используют специальное оборудование для реконструкции трехмерной модели лица (сенсорные системы) [1]. Существует зависимость между ценой и точностью системы.

Большинство высокоточных систем стоят десятки тысяч долларов и требуют высокой квалификации персонала, как для настройки, так и для использования. Но высокие темпы развития компьютерных технологий и технологий обработки видео привели к созданию относительно недо-

рогих систем, которые можно применять в большинстве областей [2].

Методы пассивного моделирования используют несколько изображений, которые были захвачены без датчиков. Для пассивной реконструкции необходимо одно изображение, либо совокупность нескольких изображений, захваченных одновременно, а также может использоваться последовательность изображений, то есть видео. Одни из самых распространенных методов пассивного моделирования включают в себя стереозрение, морфинг и определение формы по перемещению.

Стереозрение. Для метода стереозрения характерны две камеры, которые работают синхронно. С их помощью можно восстановить изгиб и местоположение наблюдаемого объекта в системе координат сцены. Главная задача данной технологии – получить две необходимые карты. Одна из них называется карта глубины – это такое изображение, где для каждого пиксела хранится не только информация о цвете, но и также содержится информация о том, как далеко объект расположен от камеры. Другая карта называется карта глубины, она создает-

ся с помощью стереопары изображений. Принцип ее создания прост: для всех точек на одной фотографии находятся парные им точки на другой фотографии. Точно также для всех пикселей первого изображения находятся парные пиксели на втором изображении. Предположим, что координаты пикселя на первом изображении равны $(x_0 - d, y_0)$, где d – параметр, который называется несоответствием или смещением. Чем больше несоответствие, тем меньше глубина. Если соответствие между точками сцены стереоизображений установлено максимально правильно, то стереосистемы смогут восстановить высокоточную глубину. Рис. 1(a) иллюстрирует бинокулярную стереоустановку с двумя идентичными камерами, сфокусированными на одной и той же сцене.

Точка $P_1(X, Y, Z)$ находится на расстояниях x_l и x_r от точки фокусировки соответственно левого и правого изображений. Несопответствие $|x_l - x_r|$ для P_1 меньше, чем для точки P_2 , расположенной относительно ближе к камерам. Для установки связи между 2D-координатами точек изображения (x, y) и 3D-координатами (X, Y, Z) сцены требуется ряд параметров. В процессе калибровки оцениваются два типа параметров: (1) внешние, связанные с относительным положением камер и (2) внутренние, связанные с самой камерой. Внутренними параметрами являются характеристики камеры, такие как фокусное расстояние f , длина пикселя s_x , ширина s_y и координаты главной точки (O_x, O_y) . Внешними параметрами являются три поворота и три перевода камеры относительно мировой системы координат.

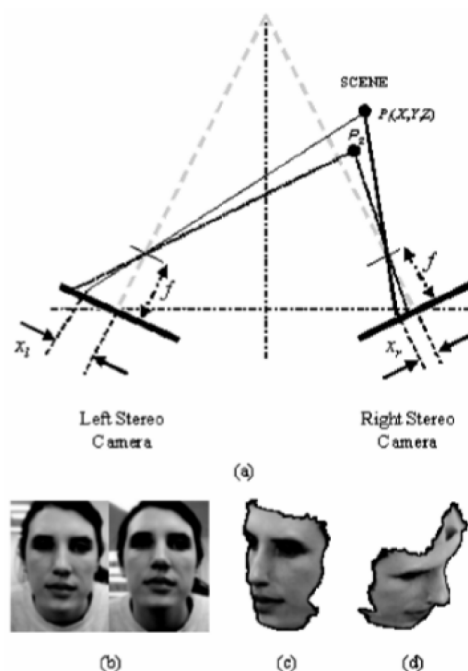


Рис. 1. 3D-реконструкция головы человека методом стереозрения:
 а – бинокулярная система стереозрения,
 б – стереопара, с-д – реконструированная часть трехмерной модели

Пренебрегая радиальным искажением объектива, внешние и внутренние параметры могут быть представлены двумя матрицами преобразования M_{ext} и M_{int} . Предполагая, что $O_x = O_y = 0$ и $s_x = s_y = 1$, единственная матрица преобразования M может быть записана как [8]

$$M = M_{int} M_{ext} = \begin{bmatrix} -f/s_x & 0 & O_x \\ 0 & -f/s_y & O_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & -R_1^T T \\ r_{21} & r_{22} & r_{23} & -R_2^T T \\ r_{31} & r_{32} & r_{33} & -R_3^T T \end{bmatrix} = \begin{bmatrix} m_a & m_b & m_c & m_d \\ m_e & m_f & m_g & m_h \\ m_i & m_j & m_k & m_l \end{bmatrix}. \quad (1)$$

Здесь r_{ij} – элементы R – вращения мировых координат относительно координат камеры, R_i – i -й вектор столбца R ; T – вектор сдвига из начала мировой системы координат.

В гомогенной системе координат получаем линейное матричное уравнение, описывающее перспективную проекцию как:

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = M \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}. \quad (2)$$

Координаты изображения определяются как $x = u/w$ и $y = v/w$. Мы можем разделить обе части (2) на m_1 и уменьшить число параметров до 11. Если (x_1, y_1) – точка на изображении, снятом левой камерой, можно записать

$$x_i = \frac{u}{v} = \frac{m_1 X + m_2 Y + m_3 Z + m_4}{m_9 X + m_{10} Y + m_{11} Z + 1}; \quad (3)$$

$$y_i = \frac{v}{w} = \frac{m_5 X + m_6 Y + m_7 Z + m_8}{m_9 X + m_{10} Y + m_{11} Z + 1}; \quad (4)$$

В наши дни мощности процессоров и видеокарт постоянно увеличиваются, однако существующие алгоритмы стереозрения значительно медлительны.

Морфинг с использованием обобщенной модели. Морфинг требует, чтобы обобщенная форма объекта была априори известна. Общая форма лица человека может быть получена с помощью лазерных сканеров или структурированных световых систем путем захвата нескольких моделей, а затем их усреднения. Общая модель 3D-головы может быть изменена (т.е. деформирована) для того, чтобы получить конкретную трехмерную модель определенного человека.

Процесс морфинга (т.е. деформации общей 3D-модели) состоит в том, что 2D-проекции 3D-модели должны максимально приближаться к одному или нескольким оригинальным изображениям лица (рис. 2) [7].

Алгоритм морфинга не вызывает сложностей. Если нам известна величина смещений и весовых коэффициентов базовых точек, необходимо изменить значения местоположений в пространстве этих точек. Каждое значение точек необходимо увеличить на число, которое вычисляется, как произведение ее смещения и веса. Алгоритм можно объяснить следующим уравнением:

$$T(i) = T(i) + L(j) * W(L(j)), \quad (5)$$

здесь в T хранится информация о местоположении точек в пространстве, $i = 0, 1315$; $j = 0$; $L(j)$ – смещение для j -й базовой точки, а $W(L(j))$ – значение веса i -й точки, связанной с j -й базовой точкой. Уравнение для алгоритма морфинга, записанное для всех точек изображения в анфас:

$$T(i).x = T(i).x + L(j).x * W(Lf(j)), \quad (6)$$

$$T(i).y = T(i).y + (L(j).y / 2) * W(Lf(j)), \quad (7)$$

Уравнение для алгоритма морфинга, записанное для всех точек изображения в профиль:

$$T(i).z = T(i).z + Ls(j).x * W(Ls(j)), \quad (8)$$

$$T(i).y = T(i).y + (Ls(j).y / 2) * W(Ls(j)), \quad (9)$$

Морфинг является довольно дорогостоящим из-за его итеративного характера. В некоторых случаях требуется ручная подгонка при выравнивании общей 3D-модели. Эти ограничения делают невозможным использование морфинга в приложениях, работающих в режиме реального времени. Тем не менее, морфинг способен восстанавливать трехмерную часть головы человека, используя всего одно 2D-изображение [5].

Определение формы по перемещению. Shape-from-motion, или 3D-реконструкция по видео, создает 3D-модель объекта из видеопоследовательности. Видео должно иметь по крайней мере два кадра и небольшие различия между точками в последовательных кадрах. В видеороликах с большими различиями между кадрами объект может быть восстановлен с использованием стереозрения, когда известны внутренние параметры камеры [8].



Рис. 2. Иллюстрация лицевой реконструкции методом морфинга: *a* – лицевая часть 3D-шаблона; *b* – трехмерный шаблон, наложенный на двумерное изображение лица; *c* – промежуточная стадия морфинга 3D-шаблона; *d* – реконструированная трехмерная модель лица

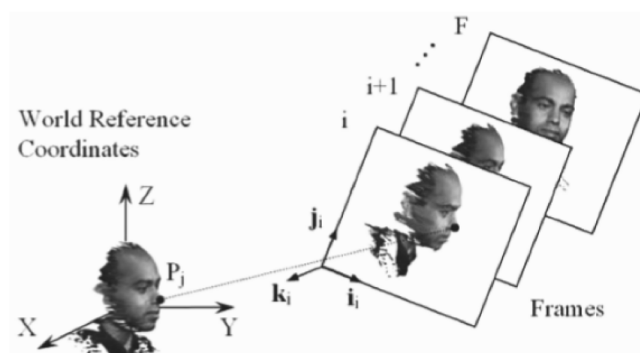


Рис. 3. Иллюстрация трехмерной реконструкции лица из видео. N – число характерных точек, которые отслеживаются по F числу кадров, где P_j ($j = 1, \dots, N$)

Глубина по видео может быть восстановлена с использованием одного из двух основных подходов – либо путем отслеживания характерных особенностей, либо вычисления оптического потока. Подход по отслеживанию характеристик требует одновременного согласования токенов в нескольких выбранных точках. Оптический поток наоборот может быть рассчитан либо абсолютно точно в каждой точке, либо ограниченно в нескольких точках объекта из пространственных и временных производных яркости изображения.

3D-реконструкция по видео широко используется в ряде приложений, таких как surveillance [3]. Тем не менее, данный метод имеет свой недостаток – такое моделирование довольно сильно зависит от наличия шумов на видео, и это может привести к дальнейшей некачественной реконструкции [4, 6].

В настоящее время активные методы более распространены, чем пассивные, однако они имеют существенный недостаток – с помощью активных методов невозможно восстановить трехмерную структуру из 2D-изображений головы человека. Поэтому, в случае, когда 2D-изображения являются единственными доступными данными, пассивные методы реконструкции 3D-модели могут сыграть ключевую роль.

Список литературы

1. Кузнецов Г.С., Белов Ю.С. Обзор метода 3D распознавания лиц без преобразования лицевой поверхности // Наука, техника и образование. [Электронный ресурс]. – URL: <http://nto-journal.ru/catalog/informacionnye-texnologii/160/> (дата обращения: 7.06.2017).
2. Тихонова Т.С., Белов Ю.С. Введение в системы отслеживания взгляда // Наука, техника и образование. [Электронный ресурс]. – URL: <http://nto-journal.ru/catalog/informacionnye-texnologii/108/> (дата обращения: 6.06.2017).
3. Тихонова Т.С., Белов Ю.С. Основные подходы к отслеживанию и распознаванию лица // Наука, техника и образование [Электронный ресурс]. – URL: <http://nto-journal.ru/catalog/informacionnye-texnologii/178/> (дата обращения: 6.06.2017).
4. Abidi B., Abidi M. Face Biometrics for Personal Identification: Multi-Sensory Multi-Modal Systems. Springer Science & Business Media: 228, 2011.
5. Bebis G., Boyle R., Parvin B.. Advances in Visual Computing. Springer Science & Business Media, 269, 2010.
6. Blanz V., Grother P., Phillips J., and Vetter T. Face recognition based on frontal views generated from non-frontal images. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR), 446–453, 2005.
7. Chellappa R., Sankaranarayanan A., Veeraraghavan A. Statistical Methods and Models for Video-based Tracking, Modeling, and Recognition. Now Publishers Inc: 89, 2010.
8. Koschan A., Pollefeys M., Abidi M.. 3D Imaging for Safety and Security. Springer Science & Business Media: 4–5, 2007.
9. Medioni G. and Pesenti B. Generation of a 3–D face model from one camera. Proc. of the 16th Int'l Conf. on Pattern Recognition (ICPR), 3: 667–671, 2002.