

УДК 519.673

О.А. Кривцов, А.М. Корилов

Отслеживание положения головы человека в пространстве на основе анализа видеокадров

Предлагается система отслеживания пространственной ориентации головы пользователя на основе анализа видеокадров, работающая по схеме: распознавание фронтального изображения лица – полуавтоматическая инициализация модели головы – отслеживание межкадрового перемещения головы с помощью градиентного метода оптимизации. В качестве модели головы используется трехмерная цилиндрическая поверхность, на которую наложена двумерная текстура.

Ключевые слова: отслеживание головы человека, модель головы, прямой композиционный метод, сопоставление изображений, веб-камера, интерфейс человек-компьютер, управление мышью, фильтр Калмана.

Задачи отслеживания пространственного положения головы человека путем анализа видеокадров, получаемых с бытовой камеры в реальном масштабе времени, возникают при анализе мимики лица человека – пользователя компьютера, создании компьютерной анимации, разработке новых способов человеко-машинного взаимодействия. Известные методы решения подобных задач можно разделить на три группы.

Во-первых, выделим группу методов [1, 2], основанную на построении модели головы и отслеживании ее перемещения путем сопоставления изображения головы и проекции ее модели. В качестве модели применяются трехмерные поверхности, на которые накладывается текстурное изображение. Если предполагается, что форма головы может деформироваться, то используют так называемые активные модели внешнего вида [3], которые позволяют отслеживать и детальные изменения мимики лица. Эти методы имеют следующие достоинства: приемлемая для практики точность определения ориентации головы, невысокие требования к качеству видео, незначительная вычислительная сложность.

Во-вторых, выделим методы [4], основанные на отслеживании «особых» точек лица либо фрагментов лица. Несмотря на хорошую точность, эти методы накладывают ограничения на качество видео.

В-третьих, существуют попытки [5] использовать иерархические классификаторы образов при решении задачи классификации изображения (лицо/не лицо) и определения пространственной ориентации лица. Для этих методов характерны высокие вычислительные затраты и ограничения по точности определения ориентации головы в пространстве.

Назовем два примера реализации перечисленных выше методов: библиотека WATSON (<http://groups.csail.mit.edu/vision/vip/watson/>) и программа института робототехники Бостонского университета (<http://www.cs.bu.edu/groups/ivc/HeadTracking/>).

Из краткого анализа существующих методов отслеживания головы пользователя компьютера следует, что многие задачи практики могут успешно решаться методами, основанными на построении модели головы. В качестве модели предлагаем использовать жесткий трехмерный цилиндр. Его преимущество – сравнительно невысокая чувствительность к ошибкам инициализации модели. В процессе отслеживания объекта выделим 3 этапа: 1) распознавание объекта на первом видеокадре; 2) инициализация модели в пространстве; 3) определение перемещения модели для каждого из последующих кадров. Этапы 1 и 2 в нашей системе могут выполняться полуавтоматически: пользователю необходимо расположить голову фронтально к камере и запустить процесс отслеживания. При этом производится детекция головы 2D детектором лица человека, и модель инициализируется с помощью полученной от детектора информации. Этап 3 будем выполнять с использованием прямого композиционного метода (ПКМ) сопоставления изображений [1, 7]. Отметим преимущества выбранного метода: невысокая вычислительная сложность (способность работать в реальном времени); приемлемая точность в оценке положения головы; нечувствительность к качеству видео, т.е. возможность обработки видео низкого качества с USB веб-камер.

1. Модель головы и построение ее проекции

ПКМ сопоставления изображений использует двумерные растровые изображения, поэтому построим проекцию трехмерной модели головы и преобразуем примитивы, составляющие модель, из векторного формата в растровый формат. Далее используется терми-

Доклады ТУСУРа, № 2 (18), часть 2, декабрь 2008

нология, применяемая в спецификации компьютерной графики OpenGL (<http://opengl.org>).

Преобразование координат

Трехмерная модель объекта (рис. 1, а) представляет собой набор примитивов: треугольников или плоских выпуклых многоугольников. Позиционирование объекта относительно наблюдателя определяется матрицей модель-вид **M** размера 4×4:

$$\mathbf{M} = \begin{pmatrix} \mathbf{R} & \mathbf{T} \\ \mathbf{0}_{1 \times 3} & 0 \end{pmatrix}. \tag{1}$$

Матрица **M** содержит в себе матрицу вращения **R** размера 3×3 и вектор параллельного переноса $\mathbf{T} = (t_x, t_y, t_z)^T$, определяющий смещение системы координат объекта относительно точки отсчета системы координат, связанной с наблюдателем. Матрица **R** задает поворот объекта относительно определенной оси вращения. Ее можно представить как произведение матриц **R_x(α)**, **R_y(β)**, и **R_z(γ)** (pitch-yaw-roll представление):

$$\mathbf{R} = \mathbf{R}_x(\alpha)\mathbf{R}_y(\beta)\mathbf{R}_z(\gamma), \tag{2}$$

где α, β, γ – углы Эйлера, определяющие последовательный поворот системы координат объекта вокруг оси X_e, затем Y_e, и, наконец, Z_e.

Перспективная модель камеры определяется матрицей проекции

$$\mathbf{P} = \begin{pmatrix} \frac{2n}{r-l} & 0 & \frac{r+l}{r-l} & 0 \\ 0 & \frac{2n}{t-b} & \frac{t+b}{t-b} & 0 \\ 0 & 0 & \frac{-(f+n)}{f-n} & \frac{-2fn}{f-n} \\ 0 & 0 & -1 & 0 \end{pmatrix}, \tag{3}$$

где l, r, b, t, n, f – координаты соответственно левой, правой, нижней, верхней, ближней и дальней плоскостей отсечения.

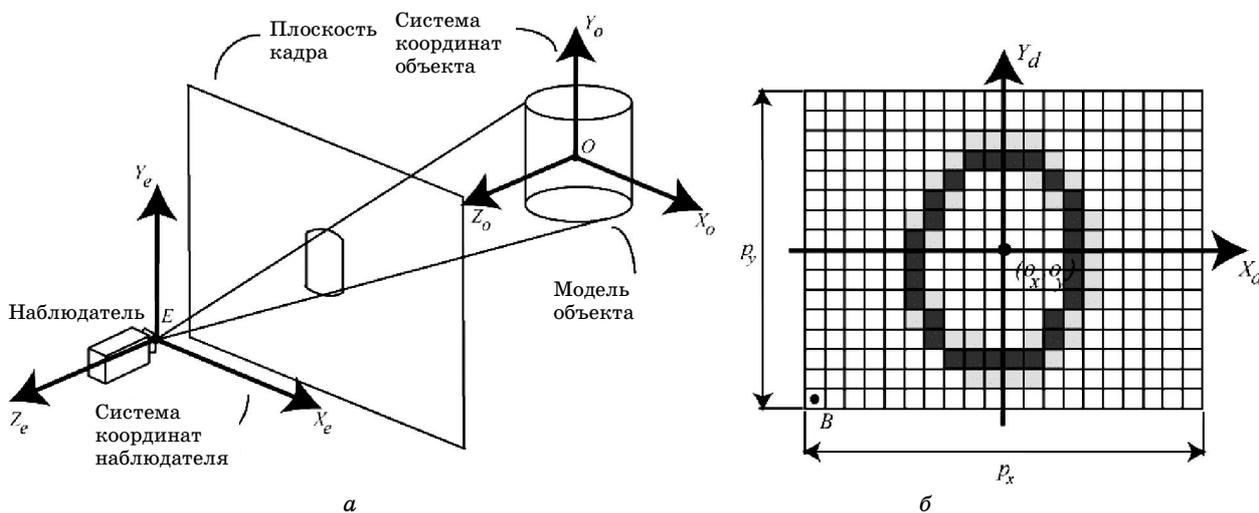


Рис. 1. Проецирование модели объекта на экран

Координаты объекта задаются в виде вектора $\mathbf{X}_o = (x_o; y_o; z_o; w_o)^T$. После умножения координат объекта **X_o** на матрицу модель-вид **M** они преобразуются в координаты наблюдателя $\mathbf{X}_e = (x_e; y_e; z_e; w_e)^T$:

$$\mathbf{X}_e = \mathbf{M}\mathbf{X}_o. \tag{4}$$

Умножая матрицу проецирования **P** на вектор **X_e**, получим вектор координат отсечения $\mathbf{X}_c = (x_c; y_c; z_c; w_c)^T$:

$$\mathbf{X}_c = \mathbf{P}\mathbf{X}_e. \quad (5)$$

И, наконец, разделив компоненты вектора \mathbf{X}_c на w_c , получим вектор нормализованных экранных координат $\mathbf{X}_d = (x_d; y_d; z_d)^T$:

$$\mathbf{X}_d = (x_c/w_c; y_c/w_c; z_c/w_c)^T. \quad (6)$$

Нормализованные экранные координаты (6) могут принимать значения в интервале $[-1; 1]$. В [6] предлагается использовать иной способ нормализации экранных координат, который аргументируется упрощением вычислений. Для нас же стандартность принятых в компьютерной графике преобразований имеет приоритетное значение.

Каждый примитив, проецируемый на плоскость кадра, проходит процедуру растеризации (перевода из векторного формата в интенсивность дискретных пикселей, выводимых на экран). Область экрана, или окно (см. рис. 1, б), в котором производится рисование объекта, представляет собой растровое изображение $I(\mathbf{x})$, где $\mathbf{x} = (u; v)^T$ – дискретные координаты пикселя изображения. Координаты пикселей раstra могут принимать целочисленные значения: $u \in [0, 1, 2, \dots, p_x - 1]$; $v \in [0, 1, 2, \dots, p_y - 1]$. Значение пикселя раstra называется его интенсивностью.

Преобразование нормализованных экранных координат (6) в координаты окна имеет вид

$$\begin{pmatrix} x_w \\ y_w \end{pmatrix} = \begin{pmatrix} (p_x/2)x_d + o_x \\ (p_y/2)y_d + o_y \end{pmatrix}, \quad (7)$$

где p_x, p_y – соответственно ширина и высота окна; (o_x, o_y) – координаты центра окна.

Цилиндрическая модель головы

При незначительных движениях рта и небольших изменениях выражения лица можно упрощенно считать, что голова представляет собой твердое недеформируемое тело, которое можно описать цилиндрической моделью (рис. 2, а). В качестве примитивов модели возьмем треугольники. Нас интересует только боковая поверхность цилиндра, повернутая к наблюдателю. Часть поверхности, обозначенная на рис. 2, а серым цветом, считается невидимой и не используется. Крышка и дно у цилиндра отсутствуют.

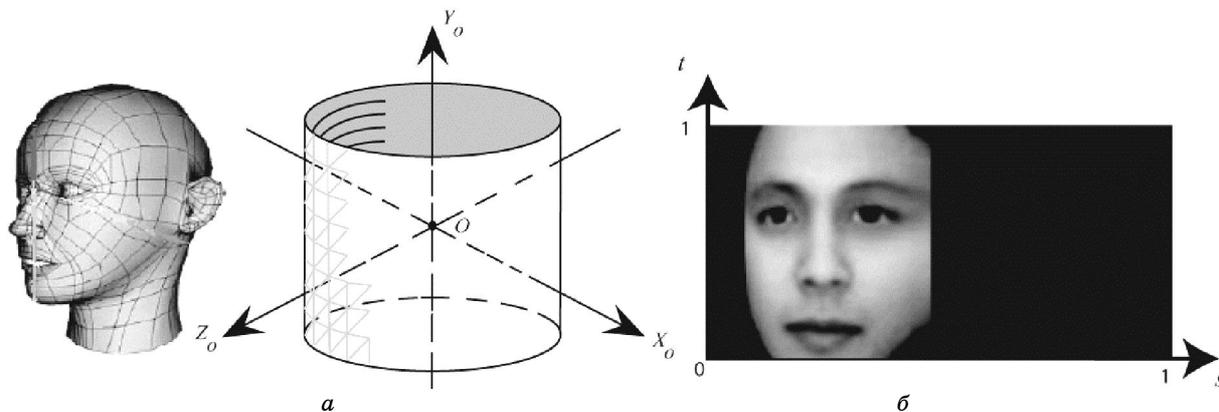


Рис. 2. Цилиндрическая модель головы

На каждом i -м кадре цилиндр имеет некоторое положение в пространстве, определяемое вектором конфигурации \mathbf{c}_i :

$$\mathbf{c}_i = (x_o; y_o; z_o; \alpha; \beta; \gamma)^T, \quad (8)$$

где x_o, y_o, z_o – координаты центра цилиндра (точка O на рис. 2, а) в системе координат наблюдателя; α, β, γ – три угла поворота системы координат модели вокруг осей системы координат наблюдателя X_e, Y_e, Z_e соответственно.

На боковую поверхность цилиндра накладывается двумерная текстура (см. рис. 2, б). Пусть точка текстуры задается нормализованными координатами $(s; t)^T$, где $s \in [0; 1]$; $t \in [0; 1]$. Тогда объектные координаты точки поверхности цилиндра $\mathbf{X}_o = (x_o; y_o; z_o; w_o)^T$, соответствующей данной точке текстуры, определяются следующей формулой:

$$\mathbf{X}_o = (-R \cos 2\pi s; H(t-0,5); R \sin 2\pi s; 1)^T.$$

Имея изображение головы пользователя, являющейся проекцией трехмерного тела, построим текстуру для боковой поверхности цилиндра. Пусть модель состоит из треугольников, причем каждой вершине треугольника соответствует пара текстурных координат. Отобразим модель в режиме отклика OpenGL (feedback mode). При этом выдается информация обо всех отображаемых примитивах. Получив информацию о том, в какие координаты окна проецируется каждая вершина треугольника и соответствующая ей точка текстуры, мы можем построить текстурное изображение по его проекции.

Построение проекции модели

Операцию получения проекции модели разделим на четыре этапа: 1) построение шаблона $T(\mathbf{x})$; 2) определение прямоугольной «области интереса» Ω ; 3) построение карты уверенности $C(\mathbf{x})$; 4) определение 3D координат для каждого пикселя шаблона $T(\mathbf{x})$ с ненулевым весом $w(\mathbf{x})$.

Произведем рисование модели (рис. 3, а) в буфер кадра и получим шаблон $T(\mathbf{x})$. OpenGL производит проверку на видимость для каждого примитива, поэтому невидимая часть цилиндра отбрасывается.

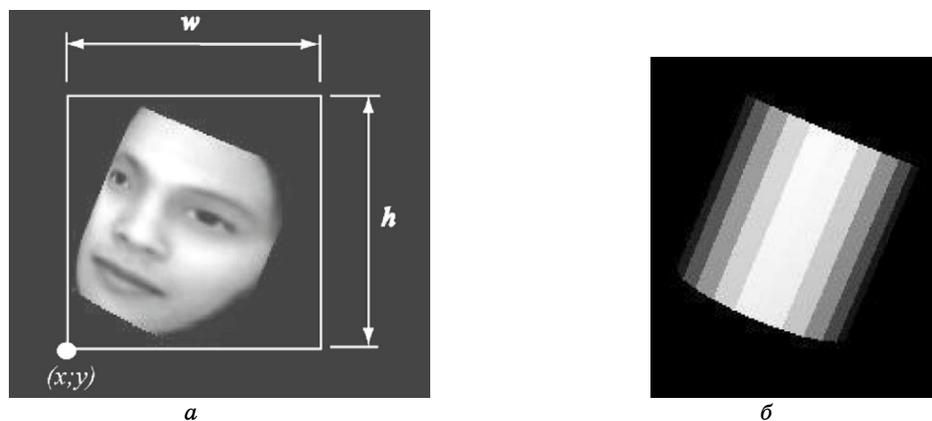


Рис. 3. Построение изображения лица пользователя

Для каждого пикселя проекции модели сохраним соответствующие ему трехмерные объектные координаты, т.е. 3D координаты точки поверхности цилиндра $\mathbf{X}_o = (x_o; y_o; z_o; w_o)^T$, которая проецируется в соответствующую точку окна $\mathbf{X}_w = (x_w; y_w)^T$. Эти 3D координаты используются в процедуре минимизации ПКМ.

В изображении $T(\mathbf{x})$ выделим прямоугольную область интереса Ω (обозначена прямоугольником на рис. 3, а). Координаты области Ω вычисляются путем отображения модели в режиме отклика и нахождения минимальных и максимальных координат, проецируемых в область вывода вершин.

Изображение на рис. 3, б, называемое картой уверенности $C(\mathbf{x})$, отражает величину вклада каждого пикселя шаблона $T(\mathbf{x})$ в процесс оптимизации. Чем выше интенсивность пикселя – тем больше его вес $w(\mathbf{x})$. Если интенсивность пикселя равна нулю, он не принадлежит проекции модели.

2. Отслеживание головы на видеокдрах

Пусть на каждом i -м кадре последовательности отображена проекция головы пользователя, имеющей некую пространственную конфигурацию \mathbf{c}_i . Необходимо: 1) позиционировать модель для первого кадра последовательности, т. е. определить вектор \mathbf{c}_1 , задать высоту H , радиус R цилиндра, а также матрицу проецирования \mathbf{P} ; 2) для каждого i -го кадра, начиная со второго, определить параметры межкадрового перемещения головы и оценить конфигурацию \mathbf{c}_i .

Инициализация модели на первом кадре

Воспользуемся алгоритмом распознавания образов, предложенным П. Виолой и М. Джонсом [8]. Пусть пользователь расположил голову фронтально к камере (рис. 4, а). Получим с помощью детектора координаты прямоугольной области $F = \{x; y; w; h\}$, в кото-

рой предположительно находится лицо. Параметры камеры и размер трехмерного объекта считаем неизвестными. Выберем их таким образом, чтобы область проекции модели на кадре приближенно соответствовала области F .

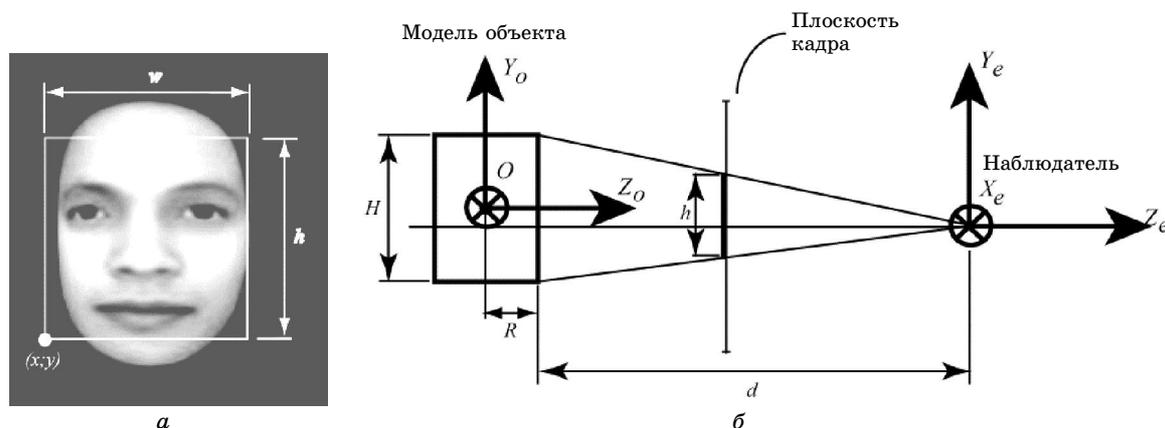


Рис. 4. Проецирование объекта в плоскость кадра

Пусть высота H цилиндра примерно соответствует вертикальному размеру области лица, заключенной между линией подбородка и линией волос. Размер лица каждого человека индивидуален, но для расчета используем субъективную оценку $H = 0,15$ м. Расстояние d от ближайшей к камере точки головы до наблюдателя (рис. 4, б) примем равным $d = 0,6$ м, что равно расстоянию от монитора, на который пользователь, сидя за компьютером, интуитивно ориентирует свою голову. Считая, что соотношение ширины и высоты области F примерно соответствует соотношению диаметра и высоты цилиндра, можно вычислить $R = \frac{Hw}{2h}$ (м).

Зададим координаты плоскостей отсечения l, r, b, t, n, f и компоненты матрицы \mathbf{P} (3). Координаты ближней и дальней плоскостей отсечения установим соответственно $n = 0,1$ м и $f = 10$ м. Используем симметричную перспективную проекцию ($l = -r; b = -t$). Пусть направления оси Z_o системы координат модели и оси Z_e системы координат наблюдателя совпадают. Обозначим через $f_L = \frac{hd}{H}$, тогда координаты плоскостей отсечения r и t соответственно определяются формулами:

$$r = \frac{np_x}{2f_L}, \quad t = \frac{np_y}{2f_L},$$

где p_x и p_y – соответственно ширина и высота окна.

Подставив l, r, b, t, n, f в (3), получим матрицу проецирования \mathbf{P} .

Определим координаты центра O цилиндра (x_O, y_O и z_O). Так как мы приняли расстояние от камеры до головы равным $d = 0,6$ м, то $z_O = -0,6$, а x_O и y_O соответственно определим по следующим формулам:

$$x_O = \left(\frac{w - p_x}{2} + x \right) \frac{d}{f_L}, \quad y_O = \left(\frac{h - p_y}{2} + y \right) \frac{d}{f_L},$$

где x, y, w, h – соответственно координаты нижнего левого угла, ширина и высота области F ; p_x, p_y – соответственно ширина и высота окна.

Так как направления координатных осей Z_o и Z_e совпадают, то углы поворота цилиндра α, β и γ вокруг осей координат равны нулю.

Итак, определены все компоненты вектора конфигурации \mathbf{c}_i . Подставив их в (1) и (2), получим матрицу модель-вид \mathbf{M} .

Отметим, что изложенный способ инициализации модели не является полностью автоматическим, что обусловлено недостаточной точностью детектирования фронтального положения лица. Необходим критерий, по которому можно было бы судить, не «поте-

рял» ли трекер голову на видеокадре, чтобы можно было перейти на этап детектирования лица, однако это – предмет дальнейших исследований.

Отслеживание межкадрового перемещения модели

Инициализировав модель для первого кадра, будем искать параметры ее движения между первым и вторым кадрами, вторым и третьим кадрами, ..., i -м и $(i+1)$ -м кадрами.

Пусть имеются два последовательных видеокадра $I_i(\mathbf{x})$ и $I_{i+1}(\mathbf{x})$. Предположим, что условия освещения постоянны, и разница между этими изображениями обусловлена лишь перемещением головы пользователя. Конфигурация головы \mathbf{c}_i на i -м кадре известна. Необходимо найти оценку конфигурации головы \mathbf{c}_{i+1} на кадре $(i+1)$.

Сформулируем постановку решаемой задачи, заменив обозначения $I_i(\mathbf{x})$ и $I_{i+1}(\mathbf{x})$ соответственно на $T(\mathbf{x})$ и $I(\mathbf{x})$.

Постановка задачи. Пусть имеются два растровых полутоновых изображения: шаблон $T(\mathbf{x})$ и видеокадр $I(\mathbf{x})$. Предположим, что условия освещения неизменны, и разница между изображениями $T(\mathbf{x})$ и $I(\mathbf{x})$ обусловлена лишь перемещением головы пользователя. Конфигурация \mathbf{c}_i головы для изображения $T(\mathbf{x})$ известна. Необходимо найти оценку конфигурации головы \mathbf{c}_{i+1} для изображения $I(\mathbf{x})$.

Для определения параметров движения модели между двумя последовательными кадрами применим ПКМ [1, 7]. Обработаем ПКМ два растровых полутоновых изображения (шаблон $T(\mathbf{x})$ и изображение $I(\mathbf{x})$) и в результате получим матрицу деформации $\mathbf{W}(\mathbf{p})$, с помощью которой $T(\mathbf{x})$ деформируется так, чтобы оптимально соответствовать изображению $I(\mathbf{x})$, т.е. решается следующая экстремальная задача:

$$\min_{\Delta \mathbf{p}} \sum_{\mathbf{x} \in \Omega} [I(W(\mathbf{x}; \Delta \mathbf{p}); \mathbf{p}) - T(\mathbf{x})]^2,$$

где $W(\mathbf{x}; \mathbf{p})$ – функция деформации; \mathbf{p} – вектор параметров.

Найдя вектор параметров деформации \mathbf{p} , мы определим параметры межкадрового перемещения головы. Пошаговая схема метода приведена в [7], далее кратко описывается его применение к нашему случаю.

Пусть имеем пиксель шаблона $T(\mathbf{x})$ с координатами $\mathbf{x} = (u; v)^T$, которому соответствует точка поверхности цилиндра с объектными координатами $\mathbf{X}_o = (x_o; y_o; z_o; w_o)^T$. Пусть положение модели для $T(\mathbf{x})$ (матрица модель-вид \mathbf{M}) известно, а на изображении $I(\mathbf{x})$ положение модели стало $\mathbf{M}' = \mathbf{W}(\mathbf{p})\mathbf{M}$. Матрицу $\mathbf{W}(\mathbf{p})$ будем называть матрицей деформации.

Следуя [1], используем твист-представление матрицы деформации $\mathbf{W}(\mathbf{p})$, которое записывается в виде

$$\mathbf{W}(\mathbf{p}) = \begin{pmatrix} 1 & -w_z & w_y & t_x \\ w_z & 1 & -w_x & t_y \\ -w_y & w_x & 1 & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{9}$$

где $\mathbf{p} = (w_x; w_y; w_z; t_x; t_y; t_z)^T$ – вектор параметров деформации.

$\mathbf{W}(\mathbf{p})$ входит в состав функции деформации $W(\mathbf{x}; \mathbf{p})$, которая переводит координаты пикселя \mathbf{x} шаблона $T(\mathbf{x})$ в координаты соответствующего ему пикселя \mathbf{x}' изображения $I(\mathbf{x})$. Определим вид функции деформации $W(\mathbf{x}; \mathbf{p})$.

Учитывая (4) и (5), координаты отсечения трансформированной точки имеют вид

$$\mathbf{X}'_c = \mathbf{P}\mathbf{W}(\mathbf{p})\mathbf{X}_e. \tag{10}$$

С учетом (9) распишем (10) более детально:

$$\begin{pmatrix} x'_c \\ y'_c \\ z'_c \\ w'_c \end{pmatrix} = \begin{pmatrix} p_{11}(x_e - w_z y_e + w_y z_e + t_x) + p_{13}(-w_y x_e + w_x y_e + z_e + t_z) \\ p_{22}(w_z x_e + y_e - w_x z_e + t_y) + p_{23}(-w_y x_e + w_x y_e + z_e + t_z) \\ p_{33}(-w_y x_e + w_x y_e + z_e + t_z) + p_{34} \\ -1 \end{pmatrix}, \tag{11}$$

где через p_{ij} обозначены соответствующие элементы матрицы проецирования \mathbf{P} .

Разделив (11) на компоненту w'_c и применяя оконное преобразование (6), получим

$$\begin{pmatrix} u' \\ v' \end{pmatrix} = \begin{pmatrix} (x'_c/w'_c)p_x/2 + o_x \\ (y'_c/w'_c)p_y/2 + o_y \end{pmatrix}. \quad (12)$$

Выражения (11) и (12) задают функцию деформации $W(\mathbf{x}; \mathbf{p})$. Тогда матрица Якоби $\frac{\partial W(\mathbf{x}; \mathbf{p})}{d\mathbf{p}}$ запишется в следующем виде:

$$\frac{\partial W(\mathbf{x}; \mathbf{p})}{d\mathbf{p}} = \begin{pmatrix} \frac{P_{11}P_x}{2z^2} & 0 \\ 0 & \frac{P_{22}P_y}{2z^2} \end{pmatrix} \begin{pmatrix} xy & -x^2 - z^2 & yz & -z & 0 & x \\ y^2 + z^2 & -xy & -xz & 0 & -z & y \end{pmatrix}. \quad (13)$$

Отметим, что матрица Якоби (13) имеет более сложный вид, чем предлагаемая в [1], что обусловлено отличием используемой нами матрицы проецирования \mathbf{P} .

Критерий останова цикла минимизации

В качестве критерия останова итеративного цикла оптимизации используется следующее правило: искать минимум, пока не выполнено ни одно из условий: 1) $\|\Delta\mathbf{p}\| \leq \varepsilon$, 2) $k_{\text{текущий}} > k_{\text{макс}}$, где $\Delta\mathbf{p}$ – приращение вектора параметров движения на текущей итерации; ε – скаляр, значение которого определяет, осуществляется ли выход из цикла; $k_{\text{текущий}}$ и $k_{\text{макс}}$ – соответственно текущий и максимально возможный номер итерации.

Обнаружилось, что в описанной итеративной процедуре возможны осцилляции значения средней ошибки $\bar{E}_k = \frac{1}{N} \sum_{\mathbf{x}} |T(\mathbf{x}) - I(W(\mathbf{x}; \mathbf{p}))|$, где суммирование производится по всем N пикселям шаблона с ненулевым весом. Это может быть обусловлено, например, несоответствием реальной формы головы и ее цилиндрической модели и т.п. Чтобы избавиться от осцилляций, дополним критерий выхода из цикла следующим условием: пусть E_k – значение средней ошибки на k -й итерации, \bar{E}_{\min} – минимальное значение средней ошибки на всех пройденных итерациях; выход из цикла производится, если в течение K итераций подряд (в наших экспериментах $K = 5$) выполняется условие $E_k \geq \bar{E}_{\min}$. В качестве результата принимается матрица $\mathbf{W}(\mathbf{p})$, при которой средняя ошибка была наименьшей.

3. Управление экраным курсором мыши

Традиционные средства взаимодействия пользователя с компьютером (клавиатура и мышь) приводятся в действие движением рук пользователя. При этом используется лишь один способ коммуникации. Однако возможно использовать и дополнительную информацию, например, анализируя видеосъемку пользователя.

Среди работ, посвященных развитию новых методов человеко-машинного взаимодействия с помощью анализа изображения пользователя на видеокдрах, отметим работу [9], в которой делается попытка скомбинировать детектор лица с 2D трекером, основанным на цветовой информации, для определения ориентации лица в пространстве. Отслеживание ориентации лица пользователя используется для управления курсором мыши.

В нашей работе делается попытка управления курсором мыши путем анализа видео, полученного с веб-камеры. Для обзора всего монитора пользователю обычно бывает недостаточно лишь движения глаз, а требуется вращение головой, поэтому предлагается использовать наклон головы для управления курсором мыши. Для каждого видеокдра трекер головы определяет матрицу модель-вид \mathbf{M} для модели головы. Из \mathbf{M} можно получить три угла вращения объекта (вокруг осей X_e , Y_e , и Z_e), которые используются для позиционирования курсора. Для компенсации дрожания курсора применяется фильтр Калмана.

При исследовании разработанной системы использованы синтетическая последовательность видеокдрах с движущимся на них лицом и «живое» видеоизображение пользователя, получаемое с веб-камеры. Эксперименты подтвердили вполне удовлетворительное качество управления курсором мыши. Однако отметим, что система пока не способна работать полностью автоматически, так как для инициализации модели головы требуется вмешательство человека. Созданная система доступна по адресу <http://headtracking.googlepages.com>.

Заключение

Задача отслеживания ориентации головы человека – пользователя компьютера на видеокадрах актуальна при создании новых методов человеко-машинного взаимодействия, исследовании мимики лица человека, анализе визуальной речи, компьютерной анимации. В данной работе предложена реализация системы отслеживания положения головы человека, базирующаяся на анализе видеокладов, поступающих с неоткалиброванной USB веб-камеры. Отслеживание головы производится путем сопоставления видеоклада с проекцией цилиндрической текстурной модели головы и оценки перемещения модели в пространстве с помощью прямого композиционного метода сопоставления изображений. Созданная система применена для управления экраным курсором мыши путем отслеживания наклона головы человека – пользователя компьютера и способна работать в реальном масштабе времени (около 20–25 кадров в секунду на процессоре Pentium IV 3 GHz).

Литература

1. Robust Full-Motion Recovery of Head by Dynamic Templates and Re-registration Techniques / J. Xiao, T. Moriyama, T. Kanade, J. Cohn // Intl. Journal of Imaging Systems and Technology. – N.Y., USA: Wiley, – 2003. – Vol. 13. – P. 85–94.
2. Cascia M. Fast, Reliable Head Tracking under Varying Illumination: An Approach Based on Registration of Texture-Mapped 3D Models / M. La Cascia, S. Sclaroff, V. Athitsos // In IEEE Tran. on Pattern Analysis and Machine Intelligence. – Washington, DC, USA: IEEE Computer Society, 2000. – Vol. 22, № 4. – P. 322–336.
3. Matthews I. Active Appearance Models Revisited / I. Matthews, S. Baker // Intl. Journal of Computer Vision. – Hingham, MA, USA: Kluwer Academic Publishers, 2004. – Vol. 60, № 2. – P. 135–164.
4. Face Recognition by Elastic Bunch Graph Matching / L. Wiskott, J.-M. Fellous, N. Krüger, C. Malsburg // IEEE Trans. Pattern Anal. Mach. Intell. – Washington, DC, USA: IEEE Computer Society, 1997. – Vol. 19, № 7. – P. 775–779.
5. Robust Head Tracking with Particles Based on Multiple Cues Fusion / Y. Li, H. Ai, C. Huang, S. Lao // In Proc. of CVHIC06. – Berlin Heidelberg: Springer-Verlag, 2006. – P. 29–39.
6. Szeliski R. Image alignment and stitching: A tutorial // Foundations and Trends in Computer Graphics and Computer Vision. – Hanover, MA, USA: Now Publishers Inc., 2006. – Vol. 2, № 1. – P. 1–104.
7. Baker S. Lucas-Kanade 20 Years On: A Unifying Framework: Part 1 / S. Baker, I. Matthews // Tech. report CMU-RI-TR-02-16, Robotics Institute, Carnegie Mellon University. – Pittsburg, PA, USA, 2002. – 47 p.
8. Viola P. Robust Real-Time Face Detection / P. Viola, M. J. Jones // International Journal of Computer Vision. – Kluwer Academic Publishers: Netherlands, 2004. – Vol. 57, № 2. – P. 137–154.
9. Вежневцев В. Система управления курсором «мыши» при помощи движений головы пользователя / В. Вежневцев, Р. Шоргин, А. Вежневцев // Труды конф. по комп. графике и машинному зрению Graphicon-2006. – Новосибирск, 2006. – С. 455–458.

Кривцов Олег Александрович

Аспирант кафедры автоматизированных систем управления ТУСУРа

Тел.: 8 (382-2) 41-42-79

Эл. почта: olegkrivtsov@mail.ru

Корилов Анатолий Михайлович

Д-р техн. наук, проф., зав. кафедрой автоматизированных систем управления ТУСУРа

Тел.: 8 (382-2) 41-42-79

Эл. почта: korikov@asu.tusur.ru

О.А. Krivtsov, А.М. Korikov

Spatial tracking of human head position based on the video frames analysis

A spatial tracking system of PC user head position is suggested. The system operates on frames captured by video camera, and includes the following operation stages: detection of user's front face view, semi-automatic initialization of head model, tracking of inter-frame head movement with the use of gradient-based optimization method. As a head model, 3D cylinder surface with applied 2D texture has been used.

Key words: head tracking, head model, forwards compositional method, image alignment, web-camera, human-computer interface, mouse control, Kalman filter.