

УДК 519.673

О.А. Кривцов

## МОДЕЛИРОВАНИЕ ГОЛОВЫ ЧЕЛОВЕКА ДЛЯ ЗАДАЧИ ОТСЛЕЖИВАНИЯ ЕЕ ПОЛОЖЕНИЯ НА ВИДЕОКАДРАХ

Проведен сравнительный анализ трех простых жестких моделей головы (трехмерный цилиндр, трехмерный эллипсоид и плоский прямоугольник) применительно к задаче отслеживания пространственного положения головы человека на видеокадрах.

**Ключевые слова:** цилиндрическая модель головы, эллипсоидальная модель головы, плоская модель лица, отслеживание головы, сопоставление изображений.

Одним из подходов к решению задачи отслеживания головы человека-пользователя компьютера является метод [1], основанный на построении модели головы, сопоставлении ее проекции с изображением, полученным с камеры, и определении оптимальных параметров перемещения модели между последовательными видеокадрами. В этом подходе сохраняется актуальность проблема моделирования головы человека: построение трехмерной поверхности, аппроксимирующей форму головы, и наложение текстуры, задающей светоотражательные свойства поверхности головы.

В качестве одного из классов моделей головы можно выделить жесткие двумерные или трехмерные поверхности: плоский прямоугольник [2], трехмерный цилиндр [3, 4], трехмерный эллипсоид [5]. Отметим, что плоская двумерная модель фактически аппроксимирует не форму головы, а форму лица. Далее кратко рассмотрим, как происходит построение трех вышеупомянутых моделей и отслеживание головы с их использованием. Также приведем экспериментальные данные, на основе которых выполним сравнительный анализ данных моделей.

### 1. Сопоставление проекции модели головы и текущего видеокадра

Воспользуемся методом отслеживания головы, реализованным нами в [1]. Данный метод содержит три этапа. На этапе 1 определяется область первого видеокадра, в которой содержится лицо пользователя. Этап 2 представляет собой задание положения и ориентации модели в 3D пространстве. На этапе 3 производится сопоставление двух изображений – проекции модели головы и текущего видеокадра. Рассмотрим этап 3 подробнее (рис. 1).

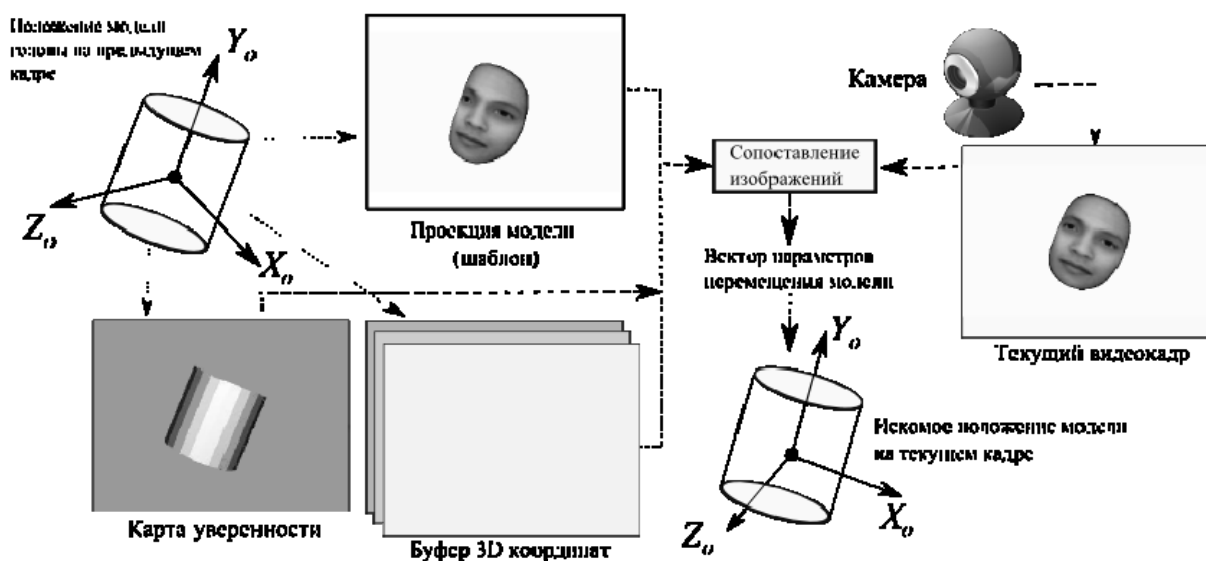


Рис. 1. Определение перемещения головы

Построив изображение – проекцию модели головы в том ее положении, которое она имела на предыдущем кадре, и имея второе изображение – текущий видеокادر, мы можем произвести процедуру сопоставления этих изображений, представляющую собой итеративный цикл минимизации некоторого критерия. В результате определим вектор параметров межкадрового перемещения модели головы, который позволяет определить искомое положение модели для текущего кадра.

Процедуру подготовки данных для сопоставления изображений разделим на четыре этапа: 1) построение проекции модели (шаблона) и перевод его из векторного в растровый формат; 2) определение прямоугольной «области интереса» на шаблоне, т.е. наименьшего прямоугольника, в который целиком входит проекция модели; 3) построение карты уверенности (определение весов пикселей шаблона); 4) нахождение 3D координат для каждого пикселя шаблона с ненулевым весом.

Воспользуемся стандартной библиотекой компьютерной графики OpenGL (<http://opengl.org>). Она позволяет избавиться от необходимости самостоятельной реализации алгоритмов построения проекции объекта, проверки его фрагментов на видимость, построения растра для каждого видимого фрагмента, интерполяции, а также позволяет использовать возможности аппаратного ускорения.

Поверхность модели будем аппроксимировать простыми геометрическими фигурами: треугольниками и плоскими выпуклыми многоугольниками. Так как считаем, что модель головы не может деформироваться, будем хранить информацию о вершинах модели в списке отображения (OpenGL display list). Это позволяет ускорить производительность рисования. Произведя с помощью OpenGL рисование модели головы в буфер кадра и скопировав полученное изображение из памяти видеокарты в клиентскую память, получим растровое изображение, которое назовем шаблоном, так как именно с ним итеративно сравнивается текущий видеокادر путем минимизации критерия для поиска параметров движения модели головы. Процедура копирования – сравнительно медленная операция на некоторых видеокартах, и это является узким местом системы.

В изображении шаблона выделим прямоугольную «область интереса». Координаты области интереса вычисляются путем отображения модели в режиме отклика OpenGL (feedback mode). В режиме отклика рисования модели в буфер кадра не происходит, но вместо этого выдается информация обо всех отображаемых вершинах, включающая оконные координаты проецируемых вершин и соответствующие им значения цвета RGB. Отобразив модель в режиме отклика, определяем область интереса путем нахождения минимальных и максимальных оконных координат вершин, проецируемых в область вывода.

Режим отклика используется и для построения карты уверенности. Получив в режиме отклика оконные координаты вершин каждой простейшей геометрической фигуры, составляющей модель головы, мы самостоятельно проводим процедуру растеризации, используя алгоритм растеризации треугольников с барицентрическими координатами, описанный в спецификации OpenGL. Если фигуры представляют собой не треугольники, а многоугольники, они разбиваются на треугольники, и каждый из треугольников обрабатывается отдельно. Карта уверенности отражает величину вклада каждого пикселя шаблона в процесс оптимизации. Чем выше интенсивность пикселя – тем больше его вес (пиксели с нулевым весом не принадлежат проекции модели).

Для каждого пикселя проекции модели головы необходимо сохранять соответствующие ему трехмерные координаты относительно наблюдателя<sup>‡</sup>, то есть 3D координаты относительно системы координат наблюдателя (СКН) точки поверхности модели  $\mathbf{X}_e = (x_e; y_e; z_e; w_e)^T$ , которая проецируется в соответствующую точку окна  $\mathbf{X}_w = (x_w; y_w; z_w)^T$ .

Имея оконные координаты вершины фигуры, можно вычислить координаты наблюдателя путем применения к оконным координатам обратного оконного преобразования и умножения на обращенную матрицу проекции (см. OpenGL функцию `gluUnproject()`):

<sup>‡</sup> Здесь и далее для обозначения координат используется терминология, введенная в [1].

$$(x_e; y_e; z_e; w_e)^T = \mathbf{P}^{-1}(2x_w/p_x - 1; 2y_w/p_y - 1; 2z_w - 1; 1)^T,$$

где  $\mathbf{P}$  – матрица проекции;

$(x_w; y_w; z_w)^T$  – координаты точки окна;

$p_x$  и  $p_y$  – ширина и высота окна соответственно.

3D координаты используются в процедуре оптимизации. Буфер 3D координат можно представить себе в виде совокупности трех матриц. Обозначим через  $a_{ij}^k$  элемент матрицы под номером  $k$ , где  $k = 0, 1, 2$ . Тогда элементы  $a_{ij}^0$ ,  $a_{ij}^1$  и  $a_{ij}^2$  содержат соответственно координаты  $x_e$ ,  $y_e$  и  $z_e$  точки поверхности модели относительно СКН, проецируемой в пиксель видеокадра, имеющий координаты  $\mathbf{X}_w = (i; j)^T$  (здесь компонента глубины  $z_w$  опущена и не используется).

После подготовки данных проводим процедуру сопоставления изображений, которая подробно описана в [1].

## 2. Задание моделей головы человека

Хотя многие объекты, с которыми мы встречаемся в реальном мире, в общем случае не являются твердыми телами, мы можем предположить, что они являются таковыми, когда выполняются определенные условия. Например, голова человека по своей природе не является абсолютно твердым телом (может двигаться челюсть, изменяться выражение лица), но при незначительных движениях рта и небольших изменениях выражения лица можно упрощенно считать, что голова – твердое недеформируемое тело. Твердое тело может иметь шесть степеней свободы движения, которые складываются из трех степеней свободы перемещения тела и трех степеней свободы вращения тела.

Пространственное положение жесткой модели на  $i$ -м кадре можем задать набором из шести параметров (вектором конфигурации):

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

где  $x_o$ ,  $y_o$  и  $z_o$  – координаты центра системы координат объекта (СКО) относительно СКН;  $\alpha$ ,  $\beta$  и  $\gamma$  – углы Эйлера, задающие поворот СКО относительно СКН (их еще называют углами соответственно тангажа, рысканья и крена).

Точки поверхности модели задаются выражением вида

$$\mathbf{X}_o = f(s; t), \quad (1)$$

где  $\mathbf{X}_o$  – вектор объектных координат точки поверхности модели;  $f(\cdot)$  – выражение, задающее соответствие между точкой двумерной текстуры и точкой трехмерной поверхности;  $s$  и  $t$  – нормализованные координаты точки текстуры ( $s \in [0; 1]$ ,  $t \in [0; 1]$ ).

Плоская модель головы (рис. 2, а) представляет собой плоский прямоугольник, параметрами которого являются ширина  $w$  и высота  $h$ .

Выражение для задания точек поверхности  $\mathbf{X}_o$  плоской модели имеет вид

$$(x_o; y_o; z_o; w_o)^T = \left( w \left( s - \frac{1}{2} \right); h \left( t - \frac{1}{2} \right); 0; 1 \right)^T.$$

Эллипсоидальная модель головы (рис. 2, б) задается тремя радиусами  $P_x$ ,  $P_y$  и  $P_z$  (вдоль осей  $X_o$ ,  $Y_o$  и  $Z_o$  соответственно). Выражение для задания поверхности будет иметь вид

$$(x_o; y_o; z_o; w_o)^T = (-R_x \sin(\pi t) \cos(2\pi s); -R_y \cos(\pi t); R_z \sin(\pi t) \sin(2\pi s); 1)^T.$$

Цилиндрическая модель головы представляет собой трехмерную цилиндрическую поверхность. Параметрами модели являются радиус цилиндра  $R$  и его высота  $h$ , а выражение для задания его поверхности имеет вид

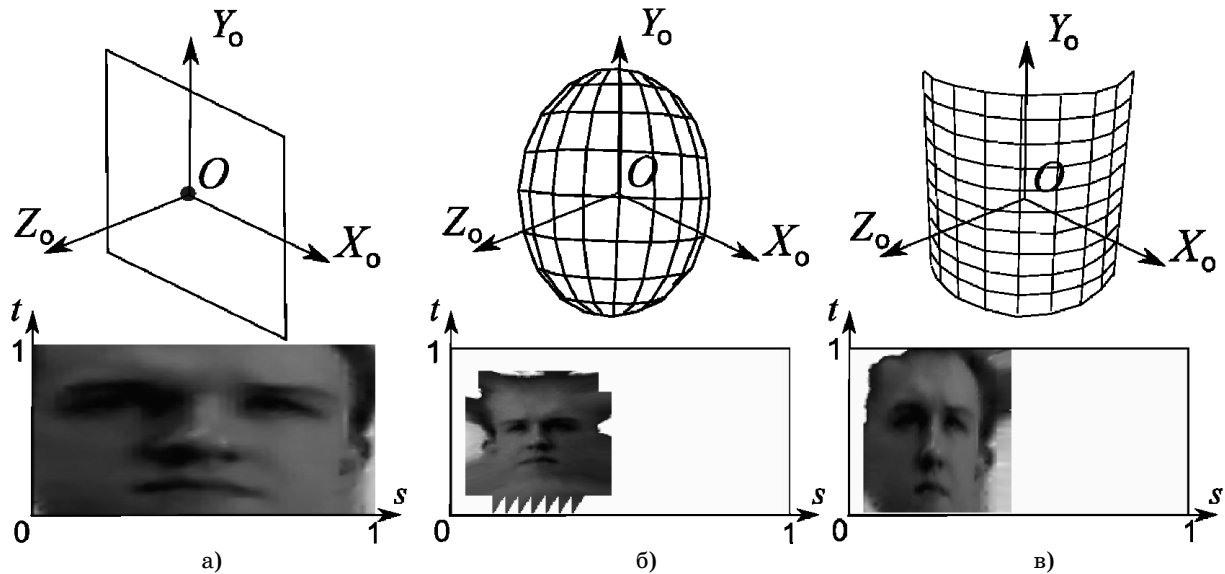


Рис. 2. Модели головы и накладываемые на них текстуры

$$(x_o; y_o; z_o; w_o)^T = \left( -R \cos(2\pi s); h \left( t - \frac{1}{2} \right); R \sin(2\pi s); 1 \right)^T.$$

Нас интересует только боковая поверхность цилиндра, повернутая к наблюдателю (рис. 2в). Крышки и дна у цилиндра нет.

### 3. Задание текстуры для модели головы

Важной задачей является построение текстурного изображения для поверхности модели головы, имея на текущем видеокadre изображение, которое является проекцией головы пользователя (реального объемного тела). Для этого необходимо задать соответствие, обратное (1):  $(s;t)^T = f^{-1}(\mathbf{X}_o)$ . Аналитическое задание сложно, поэтому поступим следующим образом.

Пусть модель состоит из простых геометрических фигур, причем каждой вершине фигуры соответствует пара текстурных координат. Отобразим модель в режиме отклика OpenGL. Получив данные о том, в какие координаты окна  $\mathbf{X}_w$  проецируется каждая вершина  $\mathbf{X}_o$  (и соответствующая ей точка текстуры  $(s;t)^T$ ), построим текстурное изображение по проекции модели.

Для кодирования информации о координатах текстуры для каждой вершины треугольника используется RGB значение цвета этой вершины (R:=s, G:=t, B не используется). Соответствующий фрагмент кода представлен в таблице 1.

Таблица 1

#### Фрагмент кода C++, отвечающий за задание вершины

```
/* Фрагмент кода, отвечающий за задание вершины простейшей фигуры */
// 1. Сохраняем текущие текстурные координаты в виде RGB цвета вершины,
// чтобы потом достать их из буфера отклика
glColor3f(s, t, 1); // R = s; G = t; B - не используется

// 2. Задаем текстурные координаты вершины
glTexCoord2f(s, t);

// 3. Задаем объектные 3D координаты вершины
glVertex3f(x, y, z);
```

При этом необходимо установить режим сглаживания цвета путем вызова функции `glShadeModel()` с параметром `GL_SMOOTH`. Если при растеризации часть фигуры выходит за границы кадра, OpenGL может разбить фигуру на несколько частей, но за счет сглаживания цвета каждая из этих частей будет иметь корректные текстурные координаты.

Итак, определив соответствие между точками экрана, точками объекта и точками текстуры, можно перевести соответствующие фрагменты экрана во фрагменты текстуры с помощью функций OpenGL.

#### 4. Экспериментальные результаты

Реализовав описанные выше модели головы в составе системы [1], которая доступна по адресу <http://headtracking.googlepages.com>, проведем эксперимент для сравнения точности этих моделей при поворотах головы. Этот эксперимент аналогичен опыту, проведенному в [4], но дополнительно используется эллипсоидальная модель.

Воспользуемся видеоклипами, доступными на сайте университета Бостона (<http://www.cs.bu.edu/groups/ivc/HeadTracking>). На этих видеопоследовательностях с названиями `jam`, `jim`, `llm`, `ssm` и `vam` сняты пять разных людей, сидящих перед камерой, укрепленной на мониторе компьютера. Люди произвольно перемещают голову и вращают ей. Условия освещения постоянны. Для каждого человека снято по 9 клипов одинаковой длины (по 200 кадров), итого в сумме имеем 45 клипов. С помощью специального устройства «Flock-of-Birds Head Tracker», прикрепляемого на голову пользователя, для каждого кадра последовательностей были определены оценки углов поворота головы, которые считаются истинными значениями углов. Рассмотрим, как будет вести себя каждая из моделей головы при поворотах головы. Для сравнительной оценки моделей предлагается следующий критерий:

$$E = \frac{1}{N} \sum_{i=1}^N \left( (\hat{\alpha}_i - \alpha_i)^2 + (\hat{\beta}_i - \beta_i)^2 + (\hat{\gamma}_i - \gamma_i)^2 \right), \quad (2)$$

где  $N$  – число кадров видеопоследовательности ( $N=200$ );

$\alpha_i$ ,  $\beta_i$ ,  $\gamma_i$  – соответственно истинные значения углов поворота вокруг осей координат (тангаж, рысканье и крен) для  $i$ -го кадра, полученные с помощью устройства;

$\hat{\alpha}_i$ ,  $\hat{\beta}_i$ ,  $\hat{\gamma}_i$  – оценки значений соответствующих углов, полученных в нашей системе с использованием заданной модели головы.

Вычислим значения критерия (2) для всех клипов. Результаты сведены в таблице 2, а также показаны в виде гистограммы на рисунке 3.

Таблица 2

Значения критерия для каждого видеоклипа

№	Ци- линдр	Эллип- соид	Плос- кость	№	Ци- линдр	Эллип- соид	Плос- кость	№	Ци- линдр	Эллип- соид	Плос- кость
jam1	38,3	33,8	72,7	jim7	1356,8	679,5	1891	ssm4	81,4	74,2	412,3
jam2	33,8	23	332,1	jim8	63,4	288,2	361,3	ssm5	55,8	260,6	195,4
jam3	23,5	20,5	34,5	jim9	206,6	174,9	459	ssm6	41,1	177,1	80,6
jam4	18,4	18,1	76,9	llm1	67,1	44,3	317,9	ssm7	189,3	138	446,2
jam5	32,7	31	228,2	llm2	22,9	107,6	164,3	ssm8	771,2	152,3	292,1
jam6	60,3	316,7	137,8	llm3	70,1	106,8	322,5	ssm9	158	87,1	240,8
jam7	22,3	23,6	387,8	llm4	280,2	288,2	1717,8	vam1	13	13,3	232,3
jam8	293,1	79	1764,3	llm5	124,8	143	202,7	vam2	40,9	40,1	102,9
jam9	115,1	142,6	531,4	llm6	40,2	106,5	230,6	vam3	130,3	86,9	918,7
jim1	32,4	30,6	257,1	llm7	64,7	62,6	263	vam4	467,6	710,6	1442,8
jim2	101,1	114	368,3	llm8	187,7	126,1	327,4	vam5	169,7	314,4	360,3
jim3	24,9	40,7	307,7	llm9	77,5	101,6	962	vam6	132,6	140,3	282
jim4	30,3	52	287,1	ssm1	33	45,6	320,3	vam7	196,4	195,5	367,8
jim5	86,9	24	194,6	ssm2	48,9	15,7	313,7	vam8	130,7	54,1	359,5
jim6	155,4	197,1	521,9	ssm3	63,4	72,2	130,7	vam9	63,6	129,2	124,7

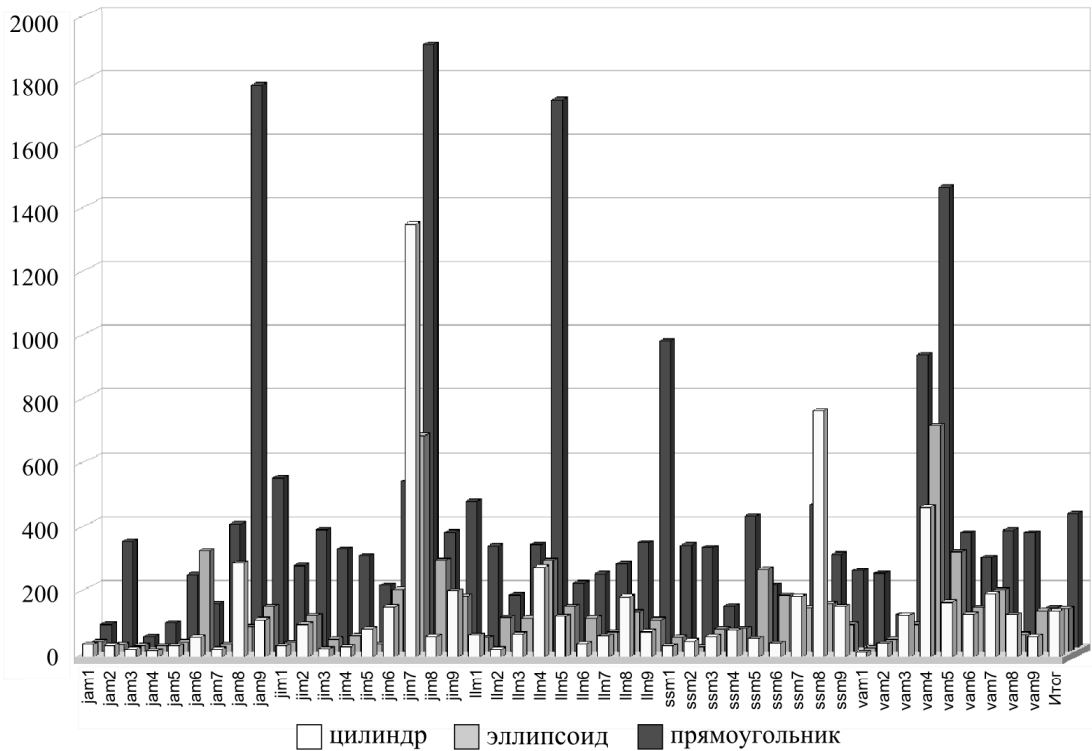
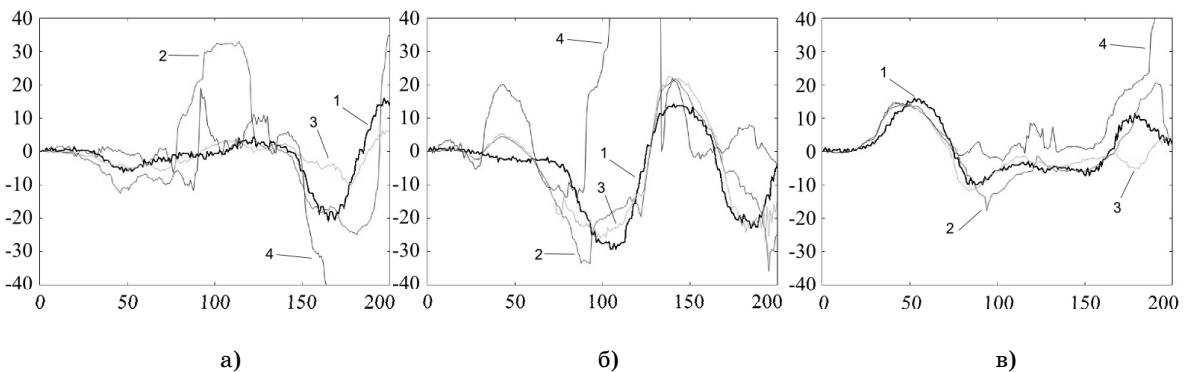


Рис. 3. Значения критерия для 45 видеопоследовательностей

Рис. 4. Значения углов поворота головы для клипа jam8: а) угол  $\alpha$ , б) угол  $\beta$ , в) угол  $\gamma$ .

По оси абсцисс отложены номера видеок кадров, по оси ординат – значения углов.

Цифрами обозначены: 1) истинное значение угла; 2) оценка угла для цилиндра;

3) оценка угла для эллипсоида; 4) оценка угла для плоской модели

Итоговое среднее значение критерия является наименьшим (135,3) для эллипсоидальной модели головы, для цилиндрической модели оно равно 142,6, а для плоской модели – 419,7. Плоская модель уступает, за незначительным исключением, первым двум моделям головы практически в каждом клипе. Эллипсоид и цилиндр в качестве модели головы показали примерно одинаковые результаты, с незначительным превосходством эллипсоида.

Как видно на рисунке 4в, все три модели хорошо аппроксимируют голову при поворотах в плоскости изображения (вокруг оси  $Z_0$ ), так как в этом случае лицо расположено фронтально к камере. Трудности возникают при повороте головы вне плоскости изображения, когда определенные части головы могут быть скрыты из поля зрения наблюдателя. Например, плоская модель часто дает неверный результат из-за того, что нос скрывает часть щеки. В этом случае интенсивность некоторых пикселей изображения головы может значительно от-

личаться от интенсивностей пикселей проекции модели, в результате получим неадекватный результат (рис. 4, а, б).

Очевидно, что плоская модель головы наименее других подходит к случаю, когда угол поворота головы вне плоскости кадра стремится к девяноста градусам, так как в этом случае проекция модели вырождается в отрезок линии. На клипах jam8, jim7, llm4, vam4 (рис. 3), где наблюдаются повороты головы в сочетании с перемещением, плоская модель показывает наихудшие результаты.

Итак, из проведенного эксперимента следует, что для решения задачи отслеживания пространственного положения головы человека лучший результат обеспечивают эллипсоидальная и цилиндрическая модели головы.

### Заключение

В данной работе проведен сравнительный анализ трех простых жестких моделей головы человека (цилиндр, эллипсоид и плоский прямоугольник) для решения задачи отслеживания положения головы человека на видеокдрах. В качестве входных данных для эксперимента использованы 45 видеоклипов, для которых были известны истинные значения углов поворота головы. Предложен количественный критерий оценки моделей, на основе которого сделан вывод о том, что наилучшие результаты для решения задачи отслеживания положения головы человека дают эллипсоидальная и цилиндрическая модели головы. Плоская модель существенно уступает им, особенно при наличии значительного поворота головы вне плоскости изображения.

### Литература

1. Кривцов О. Отслеживание положения головы человека в пространстве на основе анализа видеокдраов / О. Кривцов, А. Корииков // Доклады ТУСУРа. – Томск : Изд-во ТУСУРа. – 2008. – №2(18), часть 2. – С. 32–39.
2. Hager G. Efficient Region Tracking With Parametric Models of Geometry and Illumination / G. Hager, P. Belhumeur // IEEE Transactions on Pattern Analysis and Machine Intelligence. – Washington, DC, USA : IEEE Computer Society. – 1998. – Vol.20. – P. 1025–1039.
3. 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 // IEEE Transactions on Pattern Analysis and Machine Intelligence. – Washington, DC, USA : IEEE Computer Society. – 2000. – Vol.22, No.4. – P. 322–336.
4. 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. – New York, NY, USA : Wiley. – 2003. – Vol.13. – P. 85–94.
5. Bottino A. Real time head and facial features tracking from uncalibrated monocular views // Proceedings of 5th Asian Conference on Computer Vision. – Melbourne, Australia : Asian Federation of Computer Vision Societies – 2002. – Vol.2. – P. 776–781.

---

### Кривцов Олег Александрович

Аспирант кафедры автоматизированных систем управления ТУСУРа

Тел.: (3822) 41-42-79

Эл. почта: olegkrivtsov@mail.ru

О.А. Krivtsov

### Modeling human's head for its tracking on a video sequence

A comparative analysis of three simple rigid head models in application to head tracking is presented. The models are: 3D cylinder, 3D ellipsoid and 2D flat rectangle.

**Keywords:** cylinder head model, ellipsoid head model, planar face model, head tracking, image alignment.