

Пространственная реконструкция локально симметричных объектов по силуэтным изображениям

Леонид Местецкий, Арчил Цискаридзе
Кафедра Интеллектуальных систем
Московский физико-технический институт, Москва, Россия
l.mest@ru.net, achikotsi@gmail.com

Реферат

Рассматривается задача восстановления пространственных характеристик объектов с локально симметричными элементами. Предлагается подход к решению, основанный на модели пространственного гибкого объекта, определяемого как объединение семейства шаров с центрами на графе древовидной структуры. Предлагается метод идентификации пространственного гибкого объекта в реальном времени по стереопаре изображений силуэтов объекта. Метод основан на построении непрерывных скелетов силуэтов. Рассматривается приложение к задаче распознавания жестов.

Ключевые слова: *силуэт, стереопара, форма объекта, срединные оси, непрерывный скелет, калибровка камер, распознавание жестов, ладонь, фигура человека.*

1. ПОСТАНОВКА ЗАДАЧИ

Задача восстановления формы пространственного объекта по нескольким двумерным изображениям хорошо известна и имеет множество приложений. Особенность рассматриваемой нами постановки этой задачи состоит в том, что двумерные изображения являются бинарными и представляют собой лишь силуэты пространственного объекта. Такая задача, в частности, возникает при распознавании жестов с помощью стандартного недорогого оборудования. Исходными данными служат изображения низкого разрешения формата 480×640, полученные с использованием обычных WEB-камер. Такие камеры плохо передают текстурные особенности изображений и позволяют с достоверностью выявить лишь силуэты представленных на изображении объектов. Для распознавания жеста требуется восстановить пространственную форму столь сложного и изменчивого объекта как человеческая ладонь или фигура человека в целом. Актуальность такой постановки задачи объясняется тем, что круг потенциальных пользователей систем распознавания жестов включает в себя большое число людей, не способных приобрести дорогое оборудование, но весьма нуждающихся в системах, способных понимать жесты. Речь идёт об инвалидах, слабослышащих. Известны работы, целью которых является создание систем, понимающих азбуку глухонемых [4]. Также известны разработки соответствующих систем управления компьютером с помощью жестов [5].

Невозможность анализа изображений на уровне текстур не позволяет применить для решения задачи хорошо известные методы восстановления формы объектов по стереопаре изображений, основанные на автоматическом выявлении общих точек, присутствующих на обоих изображениях. Очевидно, что если на изображении представлен лишь силуэт

объекта, то достоверно на нём можно выявить лишь точки границы этого объекта. Но на двух картинках в стереопаре изображений силуэты полностью различаются, т.е. все точки границы одного силуэта отличаются от всех граничных точек другого силуэта. Поэтому выявление общих точек невозможно.

Подходы к решению этой задачи в полном объёме для сложных объектов в настоящее время только лишь формируются. Можно сослаться на работу [9], в которой описан метод восстановления поверхности сложного пространственного объекта (фигура лошади) по серии силуэтных изображений, полученных под разными ракурсами. Для этого метода требуются изображения хорошего качества, а также большие ресурсы процессорного времени.

Вместе с тем, существует определённый класс объектов, чьи структурные особенности позволяют решить задачу восстановления пространственной формы с приемлемой точностью в реальном масштабе времени работы системы машинного зрения. В частности, особенности таких объектов, как человеческая ладонь и фигура человека в целом, форму которых нужно восстановить при распознавании жестов, позволяют, решить эту задачу на основе использования скелетов силуэтных изображений. Эти особенности состоят в том, что человеческая ладонь и фигура человека приближённо могут быть представлены как объединение нескольких «цилиндрических» элементов, имеющих локальную осевую симметрию. Такие объекты некоторые авторы называют обобщёнными цилиндрами. Встречается также термин «трубчатые объекты» (tubular). Более строго под цилиндрическим элементом будем понимать пространственное тело, образованное семейством шаров, центры которых расположены на некоторой кривой. В работе [4] такие объекты называются пространственными жирными кривыми. Нас интересуют те объекты, которые могут быть представлены в виде объединения небольшого числа жирных кривых. Такие локально симметрические объекты могут использоваться в качестве приближённых моделей для описания ладони человека. Человеческая фигура также может быть составлена из жирных кривых, по аналогии с тем, как дети лепят человечков из пластилина.

2. МЕТОД РЕШЕНИЯ

Предлагаемый подход к решению задачи основан на идее выявления в структуре объекта точек, составляющих его оси локальной симметрии, которые хотя и не видны на изображениях стереопары, но могут быть вычислены на каждом изображении в результате анализа представленного на нём силуэта. Если рассматривать силуэты на двух снимках стереопары, как проекции пространственных жирных кривых

на соответствующие картинные плоскости изображений, то можно предположить, что оси этих жирных кривых проецируются в серединные оси соответствующих силуэтов. Это соображение предлагается положить в основу метода восстановления пространственной формы объекта по стереопаре его силуэтных изображений.

Таким образом, в роли общих так называемых реперных точек стереопары предлагается использовать некоторые (невидимые) точки, не являющиеся граничными точками силуэтов. В качестве такого множества реперных точек предлагается рассмотреть множество серединных осей силуэтного изображения, составляющих его скелет.

Реализация предлагаемого подхода ставит несколько задач. Во-первых, нужно построить скелеты силуэтов в такой форме, которая позволит идентифицировать точки разных скелетов. Во-вторых, нужно по результатам идентификации пары скелетов восстановить пространственную форму всего объекта. Особо следует отметить, что решение этих задач должно осуществляться в рамках системы машинного зрения в реальном времени работы этой системы, т.е. требуется обработка нескольких стереопар изображений в секунду. Это предъявляет высокие требования к вычислительной эффективности разрабатываемых алгоритмов.

В работе [1] введено понятие плоского гибкого объекта и предложен эффективный метод сравнения гибких объектов на основе гранично-скелетной модели. В настоящей работе предлагается обобщение плоского гибкого объекта на пространственный случай.

Пространственный гибкий объект определим, как семейство шаров различного размера с центрами на графе, имеющем структуру «дерево». С определёнными допущениями человеческая ладонь или фигура человека в целом может быть приближённо представлена в виде гибкого объекта такого вида. Таким образом, задача восстановления пространственной структуры такого объекта состоит в идентификации модели гибкого объекта по стереопаре его проекций.

Восстановление пространственных характеристик объекта даёт возможность отслеживать динамику перемещения элементов, составляющих объект, а также изменение его формы. В частности, применительно к ладони или фигуре человека это позволяет отслеживать жесты.

Реализация предложенного подхода включает решение нескольких задач.

2.1 Выделение силуэтов на изображениях

Предполагается, что имеется пара видеокамер, которые позволяют одновременно получать изображение объекта. Пример такой стереопары изображений представлен на рис.1. В наших экспериментах использовались обычные WEB-камеры, подключённые к настольному компьютеру. Каждое изображение отдельно сегментируется, выделяется силуэт, который затем представляется в виде бинарного растрового изображения. Существуют различные способы сегментации. Все они зависят от конкретных приложений. Следует отметить, что при распознавании жестов требования, предъявляемые к качеству силуэтных изображений, не очень высоки. На рис.2 показаны результаты сегментации изображений ладоней, полученные путём вычитания фона и пороговой бинаризации.

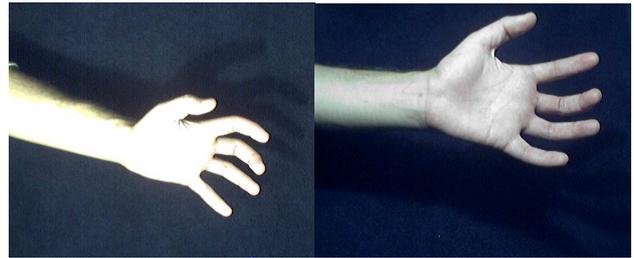


Рис.1: Стереопара ладони



Рис.2: Сегментация стереопары изображений

2.2 Построение непрерывного скелета для силуэта

Построение скелетов силуэтных бинарных изображений (рис.3) осуществляется методом, описанным в [2],[2]. Скелет представляет собой геометрическое место точек - центров вписанных в силуэт окружностей. Основным достоинством используемого метода скелетизации является представление скелета в виде графа, рёбра которого описываются непрерывными линиями. Как будет показано ниже, это свойство позволяет успешно решить задачу идентификации точек скелетов на разных изображениях. Кроме того, достоинством метода состоит в его высокой вычислительной эффективности, позволяющей решать задачу в реальном времени работы системы компьютерного зрения.

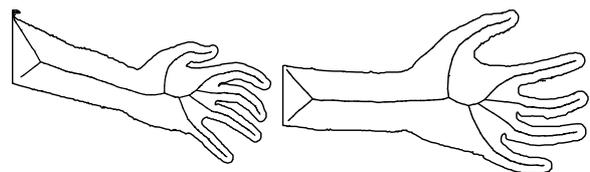


Рис.3: Скелеты силуэтных изображений.

2.3 Задача калибровки камер

Каждая точка в пространстве характеризуется координатами в некоторой фиксированной ортогональной системе координат, которую будем называть *лабораторной*. В то же время с каждой камерой связывается *собственная* ортогональная система координат, центр которой расположен в центре камеры, ось z направлена вдоль оптической оси камеры (*линии визирования*), а две другие параллельны координатным осям, заданным в плоскости изображения. Получаемая при этом *модель точечной перспективы*, несмотря на свою простоту, часто представляет собой приемлемое приближение к про-

цессу формирования изображения. Калибровка камеры подразумевает привязку положения камеры к лабораторной координатной системе и одновременно уточнение её внутренних параметров.

Существует метод калибровки непосредственно по стереопаре изображений, который требует идентификацию 5-8 точек в зависимости от выбранного метода. [6]. Наиболее сложной частью применения подхода является задача выявления и идентификации различных точечных объектов на снимках. Решение этой задачи традиционными методами требует большого объема вычислений и неизбежно сопровождается множеством ошибок. В том случае, когда задача должна решаться в реальном масштабе времени и с использованием web-камер, такой подход неприемлем. Качество получаемых изображений в силу низкого разрешения не позволяет надёжно выявлять и идентифицировать на стереопаре требуемое количество точек. Однако для локально симметричных объектов при использовании их скелетов задача эта существенно упрощается. Узлы скелета могут быть использованы в качестве реперных точек. Таким образом, задача сводится к идентификации узлов скелетов двух изображений стереопары.

2.4 Идентификация реперных точек на скелетах

Естественно предположить, что на изображении локально симметричного объекта проекция его осей совпадает со скелетом силуэта, что позволяет вычислить оси. Пусть C некоторая точка на одном из плоских изображений, образующих стереопару. В пространстве ей соответствует прямая, которая проектируется в эту точку. Образ этой прямой на другом изображении называется эпиллярной линией точки C . Для заданной точки на скелете её стереопара находится на пересечении другого скелета с эпиллярной линией этой точки (рис. 4).

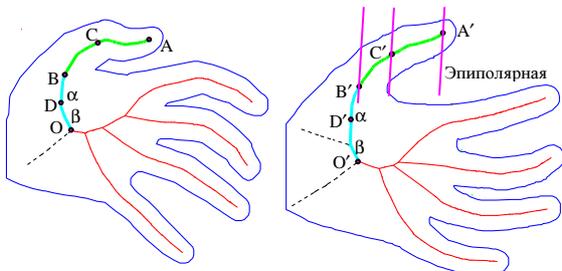


Рис.4: Поиск реперных точек на скелетах.

2.5 Построение пространственного объекта

Построив оси на основании идентификации стереопар точек, можно вычислить пространственную структуру скелета объекта. А затем, используя информацию о ширине объекта относительно срединных осей, можно восстановить поверхность пространственного объекта.

3. КАЛИБРОВКА КАМЕР

Используемый метод калибровки камер основан на принципах, описанных в работах [3][6][7].

Пусть $P = (x, y, z, 1)^T$ проективные координаты точки в лабораторной системе. Ее координаты $p = (p_x, p_y, 1)^T$ на

изображении первой камеры определяются выражением $p = K(R \cdot P + t)$, где (R, t) ориентация и центр камеры,

$K = ((f_x, 0, O_x), (0, f_y, O_y), (0, 0, 1), (0, 0, 0))^T$ здесь (O_x, O_y) координаты центра изображения, f_x и f_y коэффициенты

масштабирования. Тройка (K, R, t) определяет все параметры камеры. Возьмем в качестве лабораторной системы координат систему координат первой камеры. Тогда тройка примет вид $(K, I, 0)$. Для второй камеры будем иметь (K', R, t) , здесь (R, t) задает переход от системы координат

первой камеры во вторую. Пусть $C = (x, y, 1)^T$ некоторая точка на первом изображении. В пространстве ей соответствует прямая, которая проектируется в эту точку. Образ этой прямой на другом изображении является эпиллярной линией точки C . Ее уравнение $(F \cdot C, p') = 0$,

где $F = K'^{-T} [t]_{\times} R K^{-1}$ фундаментальная матрица (3×3)

двух камер $([t]_{\times})$ антисимметричная матрица

ассоциированная с t). Любая стереопара точек (p, p')

удовлетворяет уравнению $(p', F \cdot p) = 0$. Будем

предполагать, что матрицы K и K' известны (они содержат только внутренние параметры камер, и не зависят от расположения в пространстве, их можно отдельно вычислить [3]). Если предположить что F известно с точностью до

множителя, то можно вычислить $E = K'^T F K = [t]_{\times} R$ и тем

самым определить (R, t) . Для вычисления E существуют

различные методы [6]. Мы использовали алгоритм 6 точек.

Пусть известны шесть пар точек $(p_1, p'_1), \dots, (p_6, p'_6)$,

которые составляют стереопары. Тогда $(K'^{-1} p'_i, E \cdot K^{-1} p_i) = 0$, развернув E в вектор e и

перепишав предыдущие уравнения в матричной форме, получаем $Ae = 0$. Так как A имеет размерность (6×9) , то ее

нуль-пространство имеет размерность 3. Вычислим базис в этом пространстве и перепишем вектора в матричный формат Y, Z, W . Тогда $E = yY + zZ + W$. Остается определить

коэффициенты разложения. Так как $E = [t]_{\times} R$, два

собственных значения E совпадают, а третье равно нулю, поэтому E удовлетворяет уравнению

$$EE^T E - \frac{1}{2} \text{trace}(EE^T) E = 0.$$

Подставив сюда разложение E , получаем 9 полиномов третьей степени относительно y и z . Их можно переписать в матричный вид:

$$Bv = 0, v = (y^3, y^2z, yz^2, z^3, y^2, yz, z^2, y, z, 1)$$

Так как B есть матрица (9×10) , вычислив ее сингулярное разложение, найдем ее правый нуль-вектор b . Тогда

$$y = \frac{b_8}{b_{10}}, z = \frac{b_9}{b_{10}}.$$

Данный метод требует знания шести пар (p_i, p'_i) . Их можно брать в вершинах скелетов, что позволяет в автоматическом режиме калибровать камеры.

4. ИДЕНТИФИКАЦИЯ ТОЧЕК СКЕЛЕТОВ

Опишем способ построения на примере ладони. Рассмотрим ее стереопару (рис.1) и соответствующие осевые графы (рис.3).

Пальцы являются локально симметричными объектами В первом приближении, считая, что направление взгляда на руку не сильно отклоняется от оптической оси камеры, можно предполагать, что на плоском изображении образ пространственной оси пальца совпадает со скелетом силуэта пальца (рис. 2, кривые АВ и А'В').

Эксперименты показывают, что центры больших кругов на обоих силуэтах (точки O и O') с достаточной точностью являются стереопарными. Отсюда можно сделать предположение, что множество стереопарных точек поддерева OA осевого графа одного силуэта совпадает с поддеревом $O'A'$ осевого графа другого силуэта, что позволяет построить кривую в пространстве.

Если задать кривую OA как непрерывное отображение $f : [0,1] \rightarrow R^2$, а кривую $O'A'$ как $g : [0,1] \rightarrow R^2$, задача сводится к нахождению отображения, $w : [0,1] \rightarrow [0,1]$ которое для каждой точки $f(t)$ задает ее стереопару $g(w(t))$ (рис.5). Очевидно, что на w накладывается ограничения: зависимость должна быть монотонной и непрерывной.

Пусть C некоторая точка на одном из плоских изображений, образующих стереопару. В пространстве ей соответствует прямая, которая проектируется в эту точку. Образ этой прямой на другом изображении является эпиполярной линией точки C . Для заданной точки $C = f(t)$ на кривой OA , её стереопара $C' = g(w(t))$ находится на пересечении кривой $O'A'$ с эпиполярной линией этой точки. Таким образом, используя эпиполярные линии, можно идентифицировать точки осевых графов, составляющие стереопару, и тем самым определить пространственное расположение осей жирных линий.

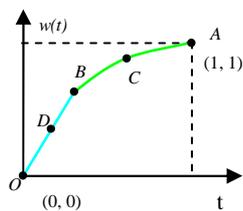


Рис.5: График $w(t)$

Однако здесь возникает трудность, если угол пересечения $\theta(t)$ кривой $O'A'$ с эпиполярной линией мал, то C'

определяется с большой погрешностью. Для устранения проблемы, выставим ограничение $\theta(t) > \theta_{\min}$. Значение $w(t)$ можно определить только в том случае если $\theta(t) > \theta_{\min}$. Для определения $w(t)$ при условии $\theta(t) \leq \theta_{\min}$ сделаем интерполяцию по уже найденным значениям W . Для нашей задачи линейная интерполяция вполне приемлема.

На рис. 4, ограничение $\theta(t) > \theta_{\min}$ нарушается на кривой OB , соответственно на графике $w(t)$ (рис. 5) кривая OB является отрезком.

На основе идентификации точек скелетов получается полная пространственная конфигурация осей локально симметричного объекта, представленного на стереопаре изображений. Во многих случаях этого представления достаточно для решения задач распознавания жестов. Однако скелетное представление содержит не только описание серединных осей объекта, но также и описание ширины объекта, поскольку известны радиусы вписанных в границы объекта кругов с центрами на серединных осях. Эта информация о ширине объекта даёт возможность не только визуализировать построенную модель гибкого объекта в виде его пространственных осей, но и построить его поверхность.

5. ВИЗУАЛИЗАЦИЯ ПРОСТРАНСТВЕННОГО ОБЪЕКТА

Построив пространственные оси и вычислив размеры шаров с центрами на этих осях, можно восстановить пространственный образ объекта.

С каждой точкой γ осевого графа свяжем сферу, найденную следующим образом: пусть Q – точка осевого графа одного из силуэтов, которая является образом точки γ . Ей соответствует максимальная вписанная окружность S с центром Q , которая целиком лежит в области силуэта. Данная окружность является образом сферы S_r с центром γ и радиусом r . Выберем произвольную точку $P \in S, P = (u, v)$.

Она задает луч l с началом в центре первой камеры и касающийся сферы S_r . Тогда r – это расстояние от точки γ до луча l .

Уравнение прямой, содержащей луч, имеет вид:

$$\begin{pmatrix} m_1 - u_1 m_3 \\ m_2 - v_1 m_3 \\ 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Таким образом вычисляется радиус сферы r . Модель объекта представляет собой огибающую поверхность этого множества сфер. Пример визуализации модели ладони, полученной по стереопаре изображений, представлен на рис.6. Как видно из этого примера, визуализация является не вполне реалистичной, поскольку в модели не только пальцы описываются жирными кривыми, но и пять (часть ладони

между пальцами и запястьем) также представляется жирной кривой. Этот недостаток визуализации может быть легко устранён, поскольку пространственное положение пальцев даёт возможность вычислить плоскость, в которой условно располагается пясть и «слегка «сплюснуть» сферы, относящиеся к этой части ладони. Результат такого преобразования представлен на рис.7.

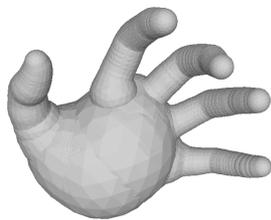


Рис.6: Визуализация пространственной модели

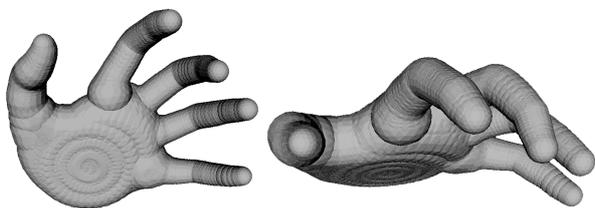


Рис.7: Коррекция пространственной модели

6. ВЫЧИСЛИТЕЛЬНЫЕ ЭКСПЕРИМЕНТЫ

Эксперименты с восстановлением пространственной модели фигуры человека проводились с куклами размером 30 см. Это объясняется лишь упрощением съёмки в лабораторных условиях. Результаты без особого труда экстраполируются на съёмку реальных человеческих фигур. На рисунке 8 показаны стереопары исходных изображений, их силуэты и полученные пространственные объекты. Эксперименты показывают, что модель гибкого объекта хорошо описывают ладонь и фигуру человека в целом. Скорость работы на компьютере Intel Pentium IV, Core 2 Duo, 2800 Mhz, 1Gb RAM составила более 5 кадров в секунду. Это даёт возможность использовать предложенный метод в системах компьютерного зрения в реальном масштабе времени их работы.

Следует отметить, что предложенный метод весьма чувствителен к структуре силуэтных изображений. В частности, если элементы изображения перекрываются, получение качественной картины срединных осей затруднено. Например, это происходит, если пальцы ладони закрывают друг друга и на силуэте вообще видны не 5 пальцев, а меньше. Однако в этом случае возможны различные усовершенствования метода. Поскольку скорость обработки стереопар достаточно высока, недостатки изображений в отдельных ракурсах могут компенсироваться использованием большего числа камер в расчёте на то, что найдётся пара камер, на которых силуэты имеют хорошую структуру. Другой путь состоит в использовании динамического видеоряда, поскольку в серии кадров появление приемлемых силуэтов более вероятно.

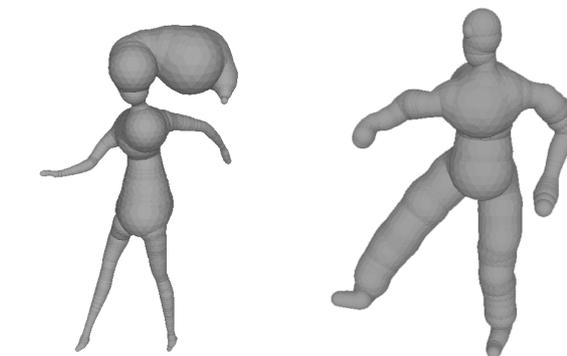
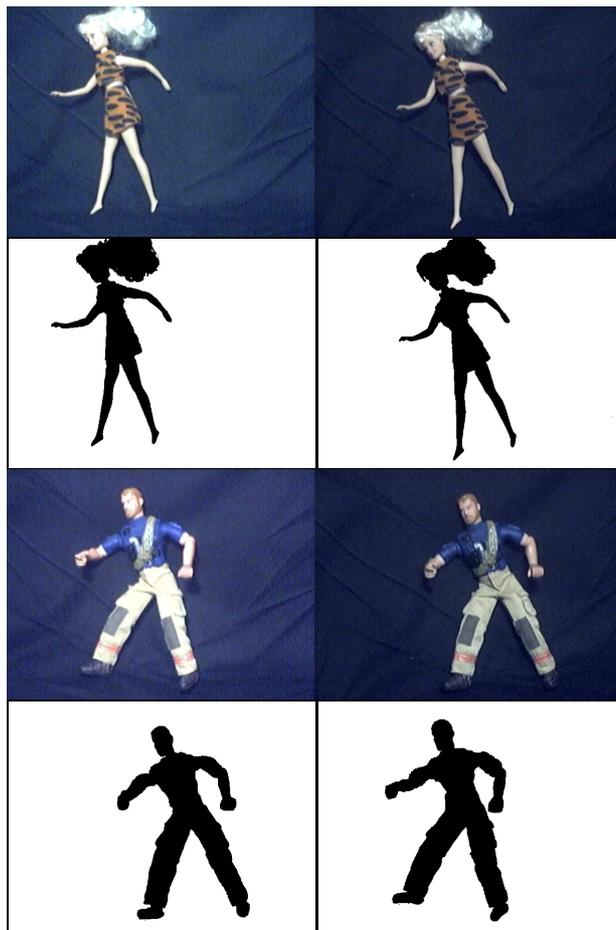


Рис.8: Стереопары исходных изображений, их силуэты и полученные пространственные объекты.

7. БЛАГОДАРНОСТИ

Авторы выражают благодарность Российскому фонду фундаментальных исследований, при поддержке которого выполнена данная работа (гранты 08-01-00670, 08-07-00270).

8. ЛИТЕРАТУРА

- [1] Mestetskiy L. *Shape comparison of flexible objects. International Conference on Computer Vision Theory and Applications, 2007.*
- [2] Местецкий Л. М. *Непрерывный скелет бинарного растрового изображения. Труды между. конф. Графикон-98*
- [3] Форсайт Д., Понс Ж. *Компьютерное зрение. Вильямс, 2004*
- [4] Thomas Burger, Alice Caplier and Stephane Mancini. *Cued speech hand gestures recognition tool. International Conference on Computer Vision Theory and Applications, 2007*
- [5] C. Keskin, O. Aran, L. Akarun. *Real time gestural interface for generic applications. European Signal Processing Conference, EUSIPCO 2005.*
- [6] Marcel Brückner, Ferid Bajramovic and Joachim Denzler. *Experimental evaluation of relative pose estimation algorithms. International Conference on Computer Vision Theory and Applications, 2008*
- [7] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision. Second Edition. Richard Hartley and Andrew Zisserman, Cambridge University Press, March 2004.*
- [8] Местецкий Л. М., Щетинин Д.В. Объемные примитивы Безье. Труды 11 международной конф. ГРАФИКОН-2001, Н.Новгород, с. 161-165.
- [9] Anthony J. Yezzi, Stefano Soatto. *Structure From Motion for Scenes Without Features. CVPR (1) 2003: 525-532.*
- [10] Mestetskiy L.M, Semenov A.B. *Binary image skeleton - continuous approach. Proceedings of the Third International conference on computer vision theory and applications (VISAPP 2008), Volume 1, 2008, p.251-258.*

COMPARISON OF ONE-SHEET SURFACES ACQUIRED BY 3D SCANNING

Spatial reconstruction of locally symmetrical objects through silhouette images

Leonid Mestetskiy, Arcil Tsiskaridze

Abstract

We consider the task of restoring the 3D characteristics of objects with the locally symmetrical elements. The approach to the solution based on the model of three-dimensional flexible object is proposed. The flexible object is defined as the union of the family of spheres with the centers on the graph of treelike structure. The method of the identification of three-dimensional flexible object in the real time on the stereo pair of the images of the silhouettes of object is proposed. Method is based on the construction of the continuous skeletons of silhouettes. Application to the task of the recognition of gestures is examined.

Keywords: *silhouette, stereo pair, Image shape, middle axis, continuous skeleton, camera calibration, gesture recognition, palm, human figure*

Авторы

Местецкий Леонид Моисеевич – доктор технических наук, профессор кафедры математических методов прогнозирования факультета Вычислительной математики и кибернетики Московского государственного университета им. М.В.Ломоносова и кафедры интеллектуальных систем Московского физико-технического института. Его адрес l.mest@ru.net.

Арчил Цискаридзе – аспирант кафедры интеллектуальных систем Московского физико-технического института. Его адрес AchikoTsi@gmail.com.