

КЛАСТЕРИЗАЦИЯ МНОЖЕСТВА ОТОЖДЕСТВЛЕННЫХ ТОЧЕК НА ИЗОБРАЖЕНИЯХ ДИНАМИЧЕСКИХ СЦЕН НА ОСНОВЕ ПРИНЦИПА МИНИМАЛЬНОЙ ДЛИНЫ ОПИСАНИЯ

© 2010 г. М. В. Петерсон

Санкт-Петербургский государственный университет информационных технологий,
механики и оптики, Санкт-Петербург

E-mail: maxim.peterson@gmail.com

В работе рассмотрена задача разделения набора отождествленных опорных точек на паре изображений динамической сцены по кластерам, которые соответствуют наблюдаемым перемещающимся объектам. На основе принципа минимальной длины описания предложен критерий, который позволяет осуществить выбор между различными классами преобразований и оценить качество кластеризации. С помощью данного критерия можно оптимальным образом подобрать модель пространственного преобразования для каждого выделенного кластера и избежать влияния выбросов в виде некорректно отождествленных точек.

Ключевые слова: сегментация движений, минимальная длина описания, аффинное преобразование, фундаментальная матрица.

Коды OCIS: 100.5760

Поступила в редакцию 20.04.2010

Введение

По серии изображений, полученных с помощью перемещающейся камеры, можно оценивать трехмерную структуру снимаемой сцены, а также восстанавливать траекторию движения самой камеры, что может применяться в области компьютерного зрения, например, для навигации мобильных роботов внутри помещений или системах дополненной реальности. Данная задача делится на два основных этапа. На первом этапе на паре изображений проводится выделение и сопоставление ключевых точек по их локальным инвариантным признакам, после чего по выделенным точкам выполняется анализ структуры наблюдаемой сцены.

В последние годы решением задачи восстановления моделей перемещений камеры и объектов наблюдаемой сцены занимается широкий круг исследователей, однако большинство работ основано на включении в последовательность изображений только одной модели движения твердого тела: наблюдаемая сцена статична, а перемещается только наблюдатель, или, наоборот, камера – статична, а в наблюдаемой сцене присутствует только один движущийся объект.

Проблема восстановления структуры сцены и перемещений камеры при названных условиях хорошо исследована, и методики ее решения достаточно подробно описаны в работах [1–3].

В этой связи представляет интерес более общий случай динамической сцены, когда в ней присутствуют независимо движущиеся объекты, а сама камера также совершает некоторые перемещения. Существует алгебраический подход к решению данной проблемы, основанный на полиномиальном представлении и обобщенном анализе главных компонентов (GPCA) [4, 5], однако данные методы не слишком устойчивы к выбросам, возникающим в результате некорректного сопоставления точек на паре изображений, и с увеличением числа моделей движения резко возрастает количество вычислений.

Альтернативой алгебраическому подходу является геометрический подход, основанный на уравнениях эпиполярной геометрии. В работе [6] представлен метод, позволяющий разделять ключевые точки на кластеры, соответствующие независимо движущимся объектам, движение которых может описываться моделями с разным числом параметров. Здесь применяется байесовский подход к решению данной задачи. При

сравнении моделей с разным числом параметров критерии наподобие СКО недостаточны для оценки их качества. Поэтому помимо ошибок описания в критерий качества модели включается ее сложность, т. е. число параметров, что задает априорные вероятности модели.

В том случае, когда в наблюдаемой сцене присутствуют независимо движущиеся объекты, приходится выбирать модель, которая наилучшим образом описывает распределение набора точечных соответствий между имеющимися моделями движения. Можно считать, что перемещения каждого объекта описываются отдельной моделью с фиксированным числом параметров, что позволяет сравнивать модели исходя из ошибки (например, среднеквадратического отклонения), с которой они описывают точечные соответствия на объектах. Таким образом, в рамках построения параметрических моделей движения необходимо найти оптимальное распределение выделенных точечных признаков по кластерам, соответствующим отдельным моделям, которое минимизирует эти невязки.

В данной работе критерий, по которому принимается решение о принадлежности точки к отдельному кластеру, описываемому своей моделью движения, представляется в виде суммарного числа битов, необходимого для описания параметров модели и возникающих невязок между точечными соответствиями на паре изображений. Данный подход характеризуется принципом минимальной длины описания (МДО) [7] применительно к выбору модели движения. Оптимальным распределением точечных соответствий по кластерам считается то, которое минимизирует суммарную длину описания всех выделенных моделей.

В работе [8] принцип минимизации длины описания применен к выбору моделей движения на серии изображений при условии перспективной проекции, где перемещение каждого объекта описывается существенной матрицей [1, 2]. Предложенное решение основывается на произвольных выборках ключевых опорных точек на паре изображений, после чего точки разделяются по кластерам и отслеживаются на последующих изображениях в серии.

В настоящей статье рассматривается применение принципа МДО для решения задачи нахождения пространственных преобразований между наборами точек на паре изображений динамической сцены. На основе принципа МДО вводится критерий, позволяющий разделить массив соответствующих точек на кластеры,

каждый из которых описывается своей матрицей преобразования и характеризует отдельный движущийся объект наблюдаемой сцены.

Модели преобразования точек на паре изображений

Точка на плоскости первого изображения может быть представлена в виде вектора-столбца $\mathbf{x}_i = (x_i, y_i)^T$, где x_i, y_i – координаты по осям абсцисс и ординат. Соответственно отождествленная с ней точка на втором изображении будет записываться как $\mathbf{x}'_i = (x'_i, y'_i)^T$. Для удобства расчетов вводится представление точки в однородных координатах [1]. Точка $\mathbf{x}_i = (x_i, y_i)^T$ лежит на некоторой линии $\mathbf{l} = (a, b, c)^T$ в плоскости изображения, если выполняется равенство $ax + by + c = 0$, которое может быть записано в виде скалярного произведения векторов $(x, y, 1)(a, b, c)^T = (x, y, 1)\mathbf{l} = 0$. Можно заметить, что для любого ненулевого k и линии \mathbf{l} равенство $(kx, ky, k)\mathbf{l} = 0$ будет выполняться, если $(x, y, 1)\mathbf{l} = 0$. Таким образом, точка на изображении с координатами (x, y) может описываться произвольным однородным вектором $\mathbf{x} = (x_1, x_2, x_3)^T$, т. е. $(x, y) = (x_1/x_3, x_2/x_3)$.

Предположим, что задан набор из N опорных точек $\mathbf{x}_i = (x_i, y_i, 1)^T$ на плоскости первого изображения и соответствующий ему набор из N точек $\mathbf{x}'_i = (x'_i, y'_i, 1)^T$ на втором изображении. Пусть также задано множество некоторых преобразований, переводящих точки с плоскости первого изображения на плоскость второго изображения $\{T_m(\mathbf{x}, \mathbf{p}_m)\}_{m=1}^M$, где \mathbf{p}_m – вектор параметров преобразования m -го класса, а M – общее количество классов преобразований.

В качестве классов преобразований рассматриваются класс аффинного преобразования и класс преобразований перспективной проекции, которые описываются фундаментальной матрицей [1, 2]. *A priori* неизвестно, каким из данных преобразований можно описать перенос точек между первым и вторым изображениями в паре. Также заранее неизвестно, сколько движущихся объектов может присутствовать в наблюдаемой сцене. Помимо этого, в наборах точек предполагается наличие некоторой погрешности в задании координат, а также ложных соответствий, ввиду возможных ошибок в процессе сопоставления.

При представлении точек на изображении в однородных координатах рассматриваемые преобразования описываются матрицей из 3×3 элементов.

Аффинное преобразование точек имеет вид $\mathbf{x}' = \mathbf{H}_A \mathbf{x}$, где \mathbf{H}_A – матрица аффинного преобразования. В развернутом виде его можно представить как

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} h_{A_1} & h_{A_2} & h_{A_3} \\ h_{A_4} & h_{A_5} & h_{A_6} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}. \quad (1)$$

Данный вид преобразования представляет собой комбинацию из преобразования поворота в сочетании с анизотропным масштабированием (коэффициенты h_{A_i} в первых двух столбцах матрицы \mathbf{H}_A) и последующим переносом (коэффициенты h_{A_i} в третьем столбце матрицы \mathbf{H}_A). Модель аффинного преобразования между точками первого и второго изображений может рассматриваться в том случае, когда глубина наблюдаемой сцены мала по сравнению с расстоянием до камеры. Как видно из выражения (1), аффинное преобразование имеет шесть независимых параметров, а одна пара отождествленных точек задает два уравнения относительно элементов h_{A_i} , поэтому для нахождения матрицы \mathbf{H}_A необходимо как минимум три пары точек. Если имеется более трех пар точек, то задача может быть решена методом наименьших квадратов.

В более общем случае перспективной проекции взаимосвязь отождествленных опорных точек на паре изображений может быть описана фундаментальной матрицей \mathbf{F} , которая является алгебраическим представлением эпиполярной геометрии [1, 2].

Пусть имеется пара изображений сцены, снятой с разных ракурсов (рис. 1). Точка X на наблюдаемой сцене проецируется в точки x и x' на первом и втором изображении соответственно и вместе с центрами проекций камеры C и C' образует эпиполярную плоскость π . Эпиполярная плоскость пересекает плоскость каждого изображения в эпиполярных линиях l и l' [1]. Таким образом, для каждой точки x на одном изображении существует соответствующая ей эпиполярная линия на другом изображении. Любая точка x' на втором изображении, соответствующая точке x , должна лежать на эпиполярной линии l' .

В отличие от аффинного преобразования, фундаментальная матрица задает преобразования из двумерного пространства в одномерное, иными словами, точка на первом изображении переносится не в отдельную точку на втором изображении, а на эпиполярную линию $l' = \mathbf{F}x$. Базовым свойством фундаментальной матрицы является то, что для любого точечного соотв-

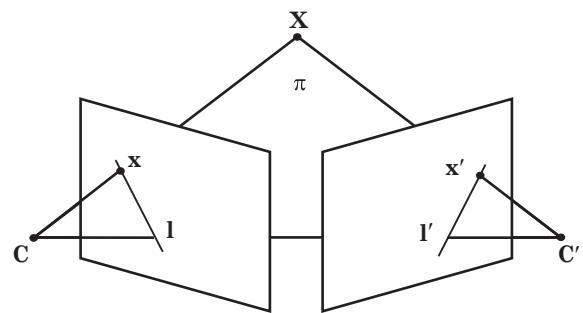


Рис. 1. Эпиполярная геометрия.

ствия на паре изображений $x \leftrightarrow x'$ она удовлетворяет следующему условию [1, 2]:

$$\mathbf{x}'^T \mathbf{F} \mathbf{x} = 0. \quad (2)$$

Соотношение (2) дает возможность определить фундаментальную матрицу без использования параметров камеры, т. е. только с помощью точечных соответствий на паре изображений. Фундаментальная матрица имеет ранг 2 и включает семь независимых элементов. Каждая пара отождествленных точек задает одно уравнение относительно элементов матрицы \mathbf{F} , при наложении дополнительного условия на вырожденность фундаментальной матрицы ($\det \mathbf{F} = 0$) она может быть вычислена по семи парам отождествленных точек. Методы вычисления фундаментальной матрицы подробно описаны в работах [1, 2].

Кластеризация точек по моделям преобразований

Согласно принципу минимальной длины описания оптимальной моделью должна считаться та, которая минимизирует суммарную длину описания модели и данных, закодированных с помощью данной модели. Таким образом, для того чтобы применить принцип МДО к задаче кластеризации отождествленных точек в соответствии с их моделями движения, необходимо оценить суммарную длину описания кластеров точек для каждого класса преобразований. Для произвольного преобразования соотношение между координатами отождествленных точек может быть записано в виде $\mathbf{x}'_i = T_m(\mathbf{x}_i, \mathbf{p}_m) + \mathbf{\varepsilon}_i$, где \mathbf{p}_m – вектор параметров модели преобразования m -го класса, а $\mathbf{\varepsilon}_i$ – векторы невязок ($i = 1, \dots, N_k$, где N_k – номер ключевой точки в k -м кластере). Векторы невязок должны кодироваться по отдельности таким образом, чтобы была возможность получить координаты точки \mathbf{x}'_i по

имеющимся \mathbf{x}_i , \mathbf{p}_m и $\boldsymbol{\varepsilon}_i$. Таким образом можно оценить длину описания координат точек на втором изображении при условии, что координаты соответствующих точек на первом изображении считаются заданными *a priori*. То есть для расчета длины описания отдельного класса преобразования необходимо оценить

- L_m – длину описания вектора параметров модели каждого класса \mathbf{p}_m ,
- L_{ind} – длину описания номеров отождествленных ключевых точек, которые принадлежат к рассматриваемому кластеру,
- L_e – длину описания вектора невязок $\boldsymbol{\varepsilon}_i$.

При оценке длины описания вектора параметров преобразования можно руководствоваться минимальным количеством точечных соответствий, которое необходимо для его вычисления. Как отмечалось выше, в случае аффинного преобразования необходимо три пары точек, которые задают шесть уравнений относительно параметров преобразования, а для фундаментальной матрицы достаточно семи пар соответствий, которые задают семь уравнений. Помимо этого для фундаментальной матрицы необходимо учитывать еще один дополнительный параметр, описывающий позицию точки с нулевой диспаратностью на эпиполярной линии. Каждый компонент вектора параметров должен быть описан некоторым числом битов. В общем случае точность данного описания зависит от количества элементов данных, которые используются в модели. Стандартная оценка [9] для длины описания вектора параметров \mathbf{p} , состоящего из n_p элементов, при имеющихся N точках в кластере будет иметь следующий вид:

$$L_p = \frac{n_p}{2} \log_2 N_k. \quad (3)$$

В свою очередь номера отождествленных ключевых точек в кластере могут быть описаны как

$$L_{\text{ind}} = \log_2 C_N^{N_k}, \quad (4)$$

ввиду того, что число различных вариантов формирования кластера, состоящего из N_k точек, равно числу сочетаний из N по N_k .

Длина описания невязок зависит от класса преобразования. Предположим, что в случае аффинного преобразования компоненты вектора $\boldsymbol{\varepsilon}_i$ независимы и являются равномерно распределенными. В таком случае длину описания невязок можно оценить следующим образом:

$$L_e = 2N_k \log_2 \sigma_e / \sqrt{2}, \quad (5)$$

где $\sigma_e / \sqrt{2}$ – значение среднеквадратичного отклонения невязок по оси абсцисс и ординат – $\boldsymbol{\varepsilon}_x$ и $\boldsymbol{\varepsilon}_y$.

В случае фундаментальной матрицы можно охарактеризовать невязки только относительно принадлежности точки к эпиполярной линии.

Тогда, для того чтобы описать координаты точки на изображении, необходимо описать ее отклонения от эпиполярной линии $\boldsymbol{\varepsilon}_l$ и ее положение вдоль этой линии $\boldsymbol{\varepsilon}_a$. Отклонения от эпиполярной линии могут иметь малые значения, в то время как положения точки на самой линии могут очень сильно отличаться из-за различной глубины сцены. Поэтому данные невязки будут иметь существенно различные распределения. Положение на эпиполярной линии можно описать как смещение относительно некоторого среднего значения (например, вычисленного через гомографию), которое также может рассматриваться как точка нулевой диспаратности. Данную опорную точку необходимо включить в модель в качестве дополнительного параметра, поэтому длина описания невязок выражается как

$$L_e = N_k (\log_2 \sigma_l + \log_2 \sigma_a). \quad (6)$$

Выбор между моделями движений различных классов может осуществляться по суммарной длине описания

$$L = L_p + L_{\text{ind}} + L_e. \quad (7)$$

Для выполнения кластеризации необходима оценка выигрыша в длине описания, получаемого при различных моделях движения. Предполагается, что координаты точек, которые не подходят под описание выбранной модели, являются независимыми и равномерно распределенными, поэтому каждая координата задается $2 \log_2 S$ битами, где S – линейный размер изображений. Общий выигрыш в битах на символ для кластера из N_k точек будет

$$2N_k \log_2 S - L. \quad (8)$$

Точки должны быть включены в кластер, если это дает максимальный положительный выигрыш в длине описания. Критерий качества кластеризации будет иметь вид

$$L_{\text{tot}} = \sum_{k=1}^K (2N_k \log_2 S - L_k), \quad (9)$$

где L_k – длина описания k -го кластера из точек, описываемых выбранной моделью. Ввиду возможных выбросов в виде некорректно отождествленных пар точек, необходимо обязательно учитывать выигрыши в длине описания при добавлении точки в кластер.

Таким образом, задача заключается в выборе распределения массива точечных соответствий по кластерам, описывающимся отдельной моделью преобразования, при котором значение суммарной длины описания было бы максимально. Так как предполагается, что среди выделенных соответствий могут присутствовать ложные пары точек, был применен следующий алгоритм, позволяющий исключать выбросы в процессе кластеризации.

На первом шаге из массива соответствий N делается минимальная произвольная выборка из n точек, которые необходимы для оценки параметров рассматриваемых моделей. Так как для оценки фундаментальной матрицы требуется больше точек, чем для аффинного преобразования, то в данной работе считалось, что $n = 7$. Для повышения вероятности принадлежности точек в выборке одному объекту брались только те соответствия, которые на обоих изображениях находятся друг от друга в пределах некоторого порогового расстояния t . После чего по полученной выборке оценивался вектор параметров каждого класса преобразований (матрица аффинного преобразования рассчитывалась по методу наименьших квадратов, фундаментальная матрица рассчитывалась по 7-точечному алгоритму [1]). Далее, исходя из полученных на данном этапе параметров преобразований и сделанной выборке, определялись невязки, которые образуются в рамках данного класса, и определялось число бит для описания каждой модели.

На следующем этапе работы делается проход по всем оставшимся отождествленным точкам и для каждой из них оценивается степень ее соответствия выделенным на предыдущем шаге

моделям перемещения. Для этого исходные параметры модели пересчитываются с учетом выбранной точки и с учетом вносимых новой точкой погрешностей переоценивается длина описания модели. Если при добавлении точки получается выигрыш в длине описания, то данная точка приписывается к кластеру, описывающемуся этой моделью, и исключается из дальнейшего рассмотрения.

В итоге, по окончании проходов по массиву отождествленных точек получаем наборы кластеров, каждый из которых характеризуется своей длиной описания в виде некоторого числа бит. Оптимальным считается то распределение точек по кластерам, которое имеет наиболее оптимальное значение критерия (9).

Экспериментальная проверка

Представленный метод кластеризации отождествленных точек по различным моделям пространственного преобразования был экспериментально проверен на реальных изображениях сцены внутри помещений. На первом этапе выполнялась проверка критерия для выбора оптимальной модели движения отдельно от задачи кластеризации точек. Для этого использовались изображения статичной сцены, снятой с разных ракурсов. В результате было подтверждено, что, в рамках решения данной задачи, выведенный критерий минимальной длины описания позволяет проводить выбор модели с соответствующей сложностью, несмотря на различные виды проективных искажений. В частности, на рис. 2 изображен пример выбора менее сложной модели, а на рис. 3 – пример выбора более сложной модели преобразования. На рис. 2 приведена пара изображений, полученных

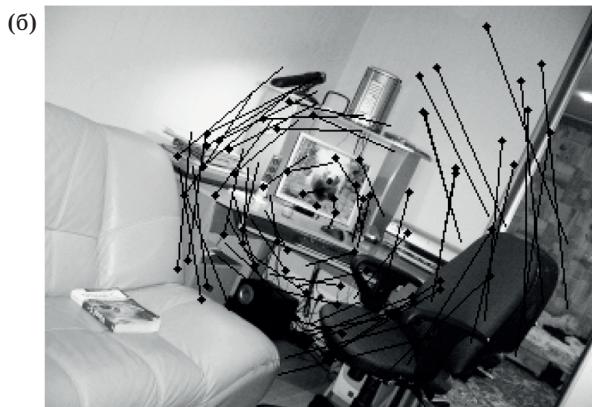


Рис. 2. Исходное изображение (а) и изображение после поворота камеры (б) с выделенными ключевыми точками, смещение которых описывается аффинным преобразованием.

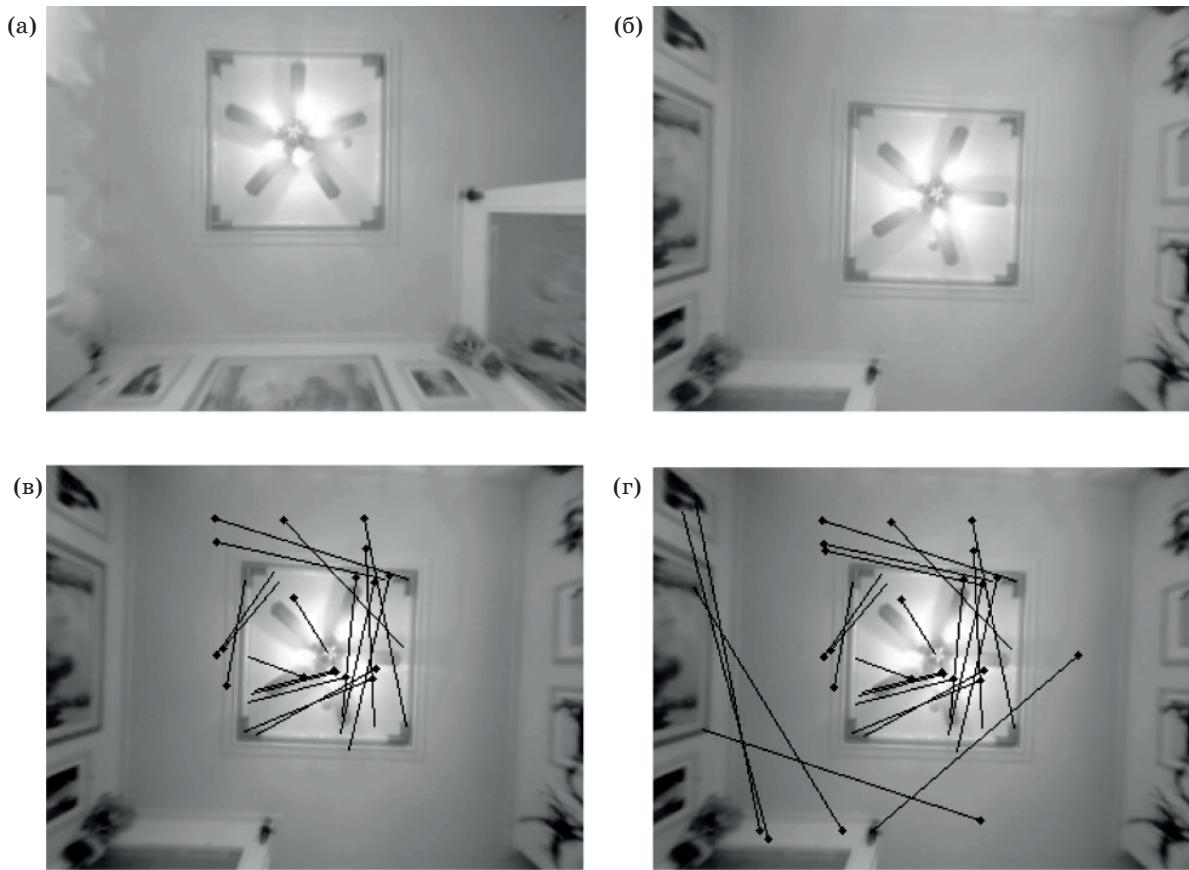


Рис. 3. Пара изображений (а, б) и отождествленные ключевые точки со смещениями, описываемыми аффинным преобразованием (в) и фундаментальной матрицей (г).

при повороте камеры. Оптимальный набор точек для моделей, которые описываются аффинным преобразованием и фундаментальной матрицей, приведен на рис. 2б. В данном примере аффинная модель дает выигрыш в 1085 битов, в то время как модель фундаментальной матрицы обеспечивает выигрыш в 1070 битов. Таким образом, может быть выбрана более корректная модель на основе аффинного преобразования.

На рис. 3 приведены изображения, полученные при перспективной проекции наблюдаемой сцены. В данном случае оптимальный набор точек для аффинной модели отличается от набора точек для модели на основе фундаментальной матрицы. Набор точек на рис. 3в дает выигрыши в 281 бит для аффинного преобразования и 269 бит для фундаментальной матрицы. В то же время набор точек на рис. 3г дает выигрыши в 278 бит для аффинной матрицы и 333 бита для фундаментальной матрицы. Таким образом, модель на основе фундаментальной матрицы может быть выбрана в качестве оптимальной для данной пары изображений, так как модель аффинного преобразования описывает меньшее количество отождествленных точек.

На втором этапе осуществлялась проверка алгоритма кластеризации на изображениях динамических сцен, полученных при перемещении камеры. На рис. 4 приведены пары изображений, на которых имеются движущиеся объекты, и результаты процедуры кластеризации для этих изображений.

Заключение

В работе было показано, что на основе принципа минимальной длины описания можно осуществлять выбор между семействами пространственных преобразований, которые содержат различное число параметров (рассматривались модель аффинного преобразования и модель перспективной проекции, которая описывается фундаментальной матрицей), а также проводить разделение отождествленных ключевых точек на кластеры в соответствии с различными моделями движения.

Предложенный критерий качества кластеризации на основе принципа МДО позволяет избежать влияния выбросов в виде некорректно отождествленных точек и подобрать модель про-



Рис. 4. Результаты кластеризации отождествленных точек по различным моделям движения.

страннычного преобразования для каждого выделенного кластера с наименьшими погрешностями. Работа предложенного алгоритма кластеризации, который оптимизирует выведенный критерий качества, была проверена на серии тестовых изображений. Полученные результаты подтверждают корректность предложенного подхода к решению описанной задачи.

Дальнейшие исследования должны быть направлены на решение проблемы поиска оптимальной кластерной гипотезы ввиду того, что использование произвольных выборок, позволяющих избежать локальных экстремумов в цепевой функции, может привести к уменьшению скорости кластеризации.

ЛИТЕРАТУРА

1. Hartley R., Zisserman A. *Multiple view geometry in computer vision*. Cambridge: University Press, 2003. 655 p.
2. Faugeras O., Luong Q.-T., Papadopoulou T. *The Geometry of Multiple Images: The Laws That Govern the Formation of Multiple Images of a Scene and Some of Their Applications*. London: MIT Press, 2004. 639 p.

3. Lourakis V., Argyros A. Efficient, Causal Camera Tracking in Unprepared Environments // *Computer Vision and Image Understanding Journal*. 2005. V. 99. № 2. P. 259–290.
4. Vidal R., Soatto S., Ma Y., Sastry S. Segmentation of dynamic scenes from the multibody fundamental matrix // *ECCV Workshop on Visual Modeling of Dynamic Scenes*. Copenhagen: Springer, 2002. P. 16–21.
5. Vidal R., Soatto S., Ma Y., Sastry S. Two-view multi-body structure from motion // *International Journal of Computer Vision*. 2006. V. 68. № 1. P. 7–25.
6. Torr P.H.S. Geometric motion segmentation and model selection // *Philosophical Transactions of the Royal Society*. London: Royal Society Publishing, 1998. P. 1321–1340.
7. Vitanyi P.M.B., Li M. Minimum description length induction, Bayesianism and Kolmogorov complexity // *IEEE Transactions on Information Theory*. 2000. V. 46. № 2. P. 446–464.
8. Schindler K., James U., Wang H. Perspective n -view Multibody Structure-and-Motion through Model Selection // *Proc. of the 9th European Conference on Computer Vision*. Graz: Springer, 2006. P. 606–619.
9. Rissanen J.J. Modeling by the shortest data description // *Automatica*. 1978. V. 14. P. 465–471.