

Д.И. Рябоконт

## Восстановление пространственной конфигурации объектов и сцен по их стереоизображениям

Предложены алгоритмы восстановления пространственной конфигурации трехмерных объектов и сцен по стереопаре снимков. Особенностью работы является постановка задач стереорекострукции как задач разметки. Предполагается, что все точки объекта, видимые в поле зрения камеры, имеют дальности, вероятностная модель которых определяется марковским случайным полем.

The algorithms for spatial reconstruction of 3D objects and scenes by their stereoisimages are suggested. A peculiarity of the work is the formulation of stereoreconstruction tasks as labeling ones. It is assumed that all scene points possess range values that can be defined by a random Markov field model.

Запропоновано алгоритми відновлення просторової конфігурації тривимірних об'єктів та сцен за стереопарою знімків. Особливістю роботи є постановка задач стереорекострукції як задач розмітки. Вважається, що всі точки об'єкта, видимі в полі зору камери, мають дальності, імовірнісна модель яких визначається марківським випадковим полем.

**Введение.** Проблема восстановления пространственной конфигурации объектов по их стереоизображениям появилась еще в начале прошлого века. Причиной ее возникновения было изобретение техники для съемки изображений окружающего мира. Актуальность исследований данной тематики объясняется стремительным ростом возможностей вычислительной и, прежде всего, мультимедийной техники. С помощью методов машинного стереозрения можно не только экономно хранить визуальную информацию, но и получать промежуточные данные, которые могут быть более важными, чем исходные изображения. Имея трехмерную модель сцены, можно генерировать такие изображения, которые невозможно было бы получить в процессе съемки. Именно эта особенность и используется сейчас во многих отраслях человеческой деятельности. К примеру, в геодезии, картографии и военном деле возникают задачи автоматического построения карты местности по результатам аэрофотосъемки. В медицине для диагностики заболеваний органов человеческого зрения используются трехмерные модели глаза, восстанавливаемые по его стереоизображениям. В системах безопасности и идентификации человека происходит анализ трехмерных моделей лиц. Навигационные автомобильные системы базируются на трехмерных моделях города для ориентации в нем. Техно-

логия стереорекострукции имеет также и большое культурологическое значение. С ее помощью сохраняют культурное наследие человечества – создают виртуальные музеи известнейших достижений архитектуры и искусства.

Существует множество подходов стереорекострукции, в том числе и таких, где используется сложное и дорогостоящее оборудование (приборы кодированного освещения объектов [1], лазерные измерители дальности и т.п.). Тем не менее, наиболее естественным и простым в технической реализации является фотограмметрический подход [2], использующий только изображения объектов, полученные с помощью двух фотоаппаратов или видеокамер. Именно этот подход имеет много общего с механизмом человеческого стереозрения. В отличие от лазерных дальномеров и приборов кодированного освещения, он позволяет получить не только пространственную конфигурацию объекта съемки, но и построить цветную модель объекта.

Особенностью предлагаемой работы является постановка задач стереозрения как задач разметки [3, 4]. Предполагается, что все точки объекта, видимые в поле зрения камеры, имеют дальности, вероятностная модель которых задается с помощью марковского случайного поля [5]. Принятые в теории разметок обозначения приведены в разд. 1. Методы построе-

ния пространственных моделей объектов по снимкам, поступающим с двух видеокамер, приведены в разд. 2–4. Исходными данными для обработки будут служить реальное и искусственные стереоизображения. Результатом их обработки будут трехмерные модели, визуализируемые с помощью стандартных программных средств персональных компьютеров.

## 1. Условные обозначения

Существуют две камеры (левая и правая), поле зрения которых есть множество  $T = \{t \equiv (i, j) | 0 \leq i < I, 0 \leq j < J\}$ . Горизонтальный размер поля зрения  $I$ , а вертикальный –  $J$ . Элементы поля зрения камер будем называть *пикселями* и обозначать парой переменных  $(i, j)$  или одной переменной  $t$ . Множество соседей пиксела  $t \equiv (i, j)$  обозначим  $\tau(t)$ . В зависимости от постановки задачи (построчная или целостная обработка стереопары), соседями пиксела  $t \equiv (i, j)$  будут  $\tau(t) = \{(i-1, j), (i+1, j)\}$  или  $\tau(t) = \{(i-1, j), (i+1, j), (i, j-1), (i, j+1)\}$ .

Изображением  $x: T \rightarrow [0 \dots 255]$  будем называть функцию, ставящую в соответствие каждому пикселу число – меру его яркости. На снимках, поступающих из камер, содержится изображение некоторой трехмерной сцены или объекта. Значения сигналов в элементах поля зрения камер обозначим  $x_i \equiv x(i, j)$  и  $x'_i \equiv x'(i, j)$  соответственно.

Идеальной стереопарой  $(x, x')$  будем называть такие два изображения, где каждому пикселу  $(i, j)$  в строке  $j$  левого изображения соответствует пиксел  $(i', j)$  в такой же строке  $j$  правого изображения, т.е. соответствующие друг другу элементы идеальной стереопары находятся только в одинаковых строках.

Параллаксом пиксела  $(i, j)$  левого изображения будем называть разность  $k \equiv (i-i', j-j')$ , где  $i'$  и  $j'$  – горизонтальные координаты соответствующего ему пиксела на правом изо-

бражении. Будем считать, что возможные значения параллакса известны и заданы множеством  $K$ .

Разметкой будем называть функцию  $\bar{k}: T \rightarrow K$ , которая ставит в соответствие каждому пикселу его параллакс. Значение параллакса в пикселе  $t \equiv (i, j)$  левого изображения будем обозначать  $k_t \equiv k(i, j)$ . Будем считать, что разметка  $\bar{k}$  является случайной реализацией марковского поля, задаваемого потенциалами Гиббса  $g: T \times T \times K \times K \rightarrow R$ . Значения функции  $g$  обозначим  $g_{it'}(k, k')$ , где  $t$  и  $t' \in \tau(t)$  – соседние пиксели левого изображения, имеющие параллаксы  $k$  и  $k'$  соответственно. Вероятность  $p(\bar{k})$  разметки  $\bar{k}$  вычисляется по формуле  $p(\bar{k}) = \prod_{t \in T, t' \in \tau(t)} g_{it'}(k(t), k(t'))$ .

Примем, что распределение яркости сигнала  $x_i, t \in T$  не зависит от пиксела  $t$ , и условная вероятность  $p(x|\bar{k})$  изображения  $x$  при условии известной разметки  $\bar{k}$  есть произведение  $\prod_{t \in T} p(x_i | k_i)$ .

## 2. Построчная обработка неидеальных стереопар

Широкую известность и признание среди фотограмметрических методов стереорекострукции получили так называемые *одномерные* алгоритмы [6–9]. Они осуществляют построчную обработку изображений и создают модель объекта из участков, рельеф которых восстановлен в результате обработки лишь одной пары строк на стереоснимках. Одномерные алгоритмы обладают высоким быстродействием и требуют малых объемов памяти. Их применение позволяет получать приемлемые результаты стереорекострукции при обработке идеальных стереопар, таких, где каждая точка объекта проецируется на левое и правое изображения в одинаковые строки. Если же стереопара не идеальна, то результат восстановления пространственной конфигу-

рации объекта бывает неудовлетворительным. Устранение этого недостатка одномерных алгоритмов заключается в том, что при обработке строки левого снимка следует использовать не одну строку правого снимка, а целую полосу строк. Далее приводится усовершенствование алгоритма построчной стереорекострукции, что в результате позволяет обрабатывать неидеальные стереопары.

### Постановка задачи

Пусть на основании стереопары  $(x, x')$  для каждого пиксела  $t \in T$  и значения параллакса  $k \in K$  заданы значения  $q_t(k)$ , равные условной вероятности  $p(x_t | k_t)$  яркости  $x_t$  в пикселе  $t$  левого снимка при условии, что пиксел  $t$  имеет параллакс  $k_t$ . Пусть также для всевозможных пар соседних пикселов, определяемых структурой соседства  $\tau(t) = \{(i-1, j), (i+1, j)\}$ , заданы значения функции  $g_{it'}(k, k')$ :

$$g_{it'}(k, k') = \begin{cases} C > 0, & \text{если } \|k - k'\| \leq \Delta \\ 0 & \text{иначе} \end{cases}$$

Задача стереорекострукции ставится как задача оптимальной разметки. Необходимо среди множества  $K^T$  всех возможных разметок найти разметку  $\bar{k}^*$  с наибольшей вероятностью  $p(\bar{k}^* | x)$ :

$$\begin{aligned} \bar{k}^* &= \underset{\bar{k} \in K^T}{\operatorname{argmax}} p(\bar{k} | x) = \underset{\bar{k} \in K^T}{\operatorname{argmax}} \frac{p(\bar{k}) \cdot p(x | \bar{k})}{p(x)} = \\ &= \underset{\bar{k} \in K^T}{\operatorname{argmax}} p(\bar{k}) \cdot p(x | \bar{k}) = \\ &= \underset{\bar{k} \in K^T}{\operatorname{argmax}} \prod_{\substack{t \in T \\ t' \in \tau(t)}} g_{it'}(k(t), k(t')) \cdot \prod_{i \in T} q_i(k_i). \end{aligned} \quad (1)$$

Преимущества приведенной постановки задачи перед известными [6–9] заключается в том, что множество  $K$  возможных значений параллакса допускает сопоставление пикселов левого и правого снимков, находящихся в различных строках. Пикселу  $(i, j)$  левого изо-

бражения может быть сопоставлен, вообще говоря, любой пиксел правого изображения, а не только такой, что имеет вертикальную координату  $j$ .

### Алгоритм решения задачи

Решение задачи (1) осуществляется методом динамического программирования [10]. Разметка элементов изображения происходит в каждой строке независимо. Рассмотрим процесс поиска разметки с наибольшей вероятностью в некоторой строке  $j$ . Он будет заключаться в последовательном отыскании наилучшей разметки для части пикселов строки.

Обозначим  $F_{ij}(k)$  качество наилучшей разметки, что присваивает пикселу  $(i, j)$  метку  $k$  и определяет некоторые метки для множества пикселов  $\{(0, j), (1, j), \dots, (i-1, j)\}$ . Величина  $F_{1j}(k)$ , к примеру, определяется как  $F_{0j}(k) = q_{(0,j)}(k)$ . Значение  $D_{ij}(k)$  будет хранить метку пиксела  $(i-1, j)$ , в которую нужно двигаться из метки  $k$  пиксела  $(i, j)$ .

Допустим, что для некоторого  $i$  и всех значений  $k \in K$  уже вычислены значения  $F_{ij}(k)$ . Тогда  $F_{i+1,j}(k)$  вычисляется по формуле  $F_{i+1,j}(k) = \max_{k' \in K} (F_{ij}(k') \cdot g_{(ij),(i+1,j)}(k', k) \cdot q_{i+1,j}(k))$ . Направление движения  $D_{i+1,j}(k)$  вычисляется по формуле  $D_{i+1,j}(k) = \operatorname{argmax}_{k' \in K} (F_{ij}(k') \cdot g_{(ij),(i+1,j)}(k', k) \times q_{i+1,j}(k'))$ .

Процесс поиска оптимальной разметки выглядит так. В результате последовательного прохода пикселов  $\{(0, j), (1, j), \dots, (I, j)\}$  строки  $j$  для каждого  $k \in K$  будет вычислено значение  $F_{I,j}(k)$ . Далее необходимо выбрать такое значение  $k \in K$ , для которого значение  $F_{I,j}(k)$  максимально. Это и будет результат разметки пиксела  $(I, j)$ . Метка пиксела  $(I-1, j)$  определяется значением  $D_{I,j}(k)$ . При выполнении последовательного обратного прохода

пикселей  $\{(I, j), (I-1, j), \dots, (0, j)\}$ , осуществляется их разметка по правилу  $k_{i-1, j} = D_{ij}(k_{ij})$ .

Временная сложность алгоритма составляет  $O(T \cdot K^2)$ . Его реализация требует памяти объемом  $O(T \cdot K^2)$  для хранения значений функции  $g$  и памяти объемом  $O(I \cdot K)$  для хранения вспомогательных данных – значений функций  $F_{i, j}(k)$  и  $D_{i, j}(k)$ .

### Результаты экспериментов

На рис. 1 представлено построение трехмерной модели человеческого лица: стереопара изображений и результат ее обработки – нераскрашенная «гипсовая» трехмерная модель. Размер каждого из снимков стереопары составлял  $6000 \times 800$  пикселей. Время обработки одной стереопары на персональном компьютере с тактовой частотой 1 ГГц не превышает 5 мин.



Рис. 1

### 3. Применение байесовских методов принятия решения для обработки стереопар

Рассмотренная выше задача стереозрения, заключающаяся в поиске разметки с  $\bar{k}^*$  наибольшей вероятностью  $p(\bar{k}^*|x)$ , является частным случаем задач стереозрения, основанных на байесовской теории принятия решений [7, 10]. Оказывается, если штраф  $W(\bar{k}, \bar{k}^*)$  за разметку  $\bar{k}$  выбрать по правилу  $W(\bar{k}, \bar{k}^*) = 1 - \delta(\bar{k}, \bar{k}^*)$  (здесь  $\bar{k}^*$  – истинная, но неизвестная разметка, а  $\delta$  – символ Кронекера), то байесовское решение  $\bar{k}$ , минимизирующее ма-

тематическое ожидание штрафа  $\sum_{k^* \in K^T} p(\bar{k}^*|x) \times W(\bar{k}, \bar{k}^*)$ , будет разметка с наибольшей вероятностью  $p(\bar{k}|x)$ . Далее рассматривается одна из возможных байесовских постановок задачи стереозрения для случая, когда штрафная функция имеет более естественный вид, а именно  $W(\bar{k}, \bar{k}^*) = \sum_{i \in T} (k_i - k_i^*)^2$ .

### Постановка задачи

Пусть, как сказано,  $W(\bar{k}, \bar{k}^*) = \sum_{i \in T} (k_i - k_i^*)^2$  – штраф за принятие решения о том, что разметка изображения есть  $\bar{k}$ , а истинная, но неизвестная разметка есть  $\bar{k}^*$ . Как и раньше, предполагается, что разметка является случайной реализацией марковского поля. Ее априорная вероятность задается с помощью потенциалов Гиббса  $g_{\tau'}(k, k')$  и определяется произведением  $p(\bar{k}) = \prod_{i \in T, \tau' \in \tau(i)} g_{\tau'}(k_i, k_{\tau'})$ , где  $\tau(i) = \{(i-1, j), (i+1, j), (i, j-1), (i, j+1)\}$ , что является множеством соседей пиксела  $i$ . Пусть заданы значения  $q_i(k)$ , равные условной вероятности  $p(x_i|k_i)$  яркости  $x_i$  в пикселе  $i$  левого снимка при условии, что пиксел  $i$  имеет параллакс  $k_i$ . Заданы значения функции  $g$ :

$$g_{\tau'}(k, k') = \begin{cases} C > 0, & \text{если } \|k - k'\| \leq 1 \\ 0 & \text{иначе} \end{cases}$$

Риском  $R(\bar{k})$  разметки  $\bar{k}$  будем называть математическое ожидание ее штрафа:

$$\begin{aligned} R(\bar{k}) &= \sum_{k^* \in K^T} p(\bar{k}^*|x) W(\bar{k}, \bar{k}^*) = \\ &= \sum_{i \in T} \sum_{k \in K} p(k_i = k|x_i) \cdot (k_i - k)^2. \end{aligned} \quad (2)$$

Разметка  $\bar{k}$ , имеющая наименьший риск, определяется по правилу  $k_i = \sum_{k \in K} k \cdot p(k_i = k|x_i)$ . Задача заключается в отыскании значений

маргинальных вероятностей  $p(k_i = k|x)$  меток  $k \in K$  во всех пикселах  $t \in T$ .

$$\begin{aligned}
 p(k_i = k|x) &= \\
 &= \sum_{\bar{k} \in K^T: k_i = k} p(\bar{k}|x) = \frac{1}{p(x)} \cdot \sum_{\bar{k} \in K^T: k_i = k} p(\bar{k}) \cdot p(x|\bar{k}) = \\
 &= \frac{1}{p(x)} \cdot \sum_{\bar{k} \in K^T: k_i = k} \left( \prod_{\substack{t \in T \\ t' \in (t)}} g_{t'}(k_t, k_{t'}) \cdot \prod_{t \in T} q_t(k_t) \right). \quad (3)
 \end{aligned}$$

### Приближенное решение задачи

Вычисление значения выражения (3) для каждой метки  $k \in K$  каждого пиксела  $t \in T$  требует суммирования вероятностей всех разметок поля зрения, имеющих в пикселе  $t$  состояние  $k$ . Для вычисления приближенного значения (3) можно просуммировать вероятности только тех разметок, которые определены в полосе размером  $I \times L$ , охватывающей пиксел  $t \equiv (i, j)$ , где  $I$  – это горизонтальный размер поля зрения, а  $L < J$  – высота полосы. Разобьем эту полосу, состоящую из  $I \cdot L$  пикселов, на  $I$  столбцов шириной в один пиксел и высотой в  $L$  пикселов. В каждом столбце размера  $1 \times L$  пикселов существует  $K^L$  различных разметок, из которых только  $M \equiv K \cdot 3^L$  имеют ненулевую вероятность.

Продвигаясь по описанным столбцам слева направо, выполним следующую процедуру. Подсчитаем в крайнем левом столбце  $i = 0$  вероятности  $p_{i=0}(\bar{k}_1|x)$ ,  $p_{i=0}(\bar{k}_2|x)$ , ...,  $p_{i=0}(\bar{k}_M|x)$   $M$  разметок, имеющих ненулевую вероятность. В столбце  $i = 1$  для всех  $M$  разметок  $\bar{k}_1, \bar{k}_2, \dots, \bar{k}_M$ , имеющих ненулевую вероятность, подсчитаем числа  $p_{i=1}(\bar{k}_1|x)$ ,  $p_{i=1}(\bar{k}_2|x)$ , ...,  $p_{i=1}(\bar{k}_M|x)$ , равные сумме вероятностей всевозможных разметок, охватывающих первые два столбца, и проходящие через соответствующие разметки  $\bar{k}_1, \bar{k}_2, \dots, \bar{k}_M$  второго столбца. В результате продолжения описанных вычислений, в каждом столбце  $i = 0, \dots, I$  будут храниться  $M$

значений  $p_i(\bar{k}_1|x)$ ,  $p_i(\bar{k}_2|x)$ , ...,  $p_i(\bar{k}_M|x)$ , содержащих сумму вероятностей всевозможных разметок фрагмента  $[0 \dots i] \times [j - \frac{L}{2} \dots j + \frac{L}{2}]$ , проходящих через такие разметки  $\bar{k}_1, \bar{k}_2, \dots, \bar{k}_M$  столбца  $i$ , которые имеют ненулевую вероятность.

Описанную процедуру также выполним, продвигаясь по столбцам справа налево. В результате в каждом столбце  $i = 0, \dots, I$  будут храниться  $M$  значений  $p'_i(\bar{k}_1|x)$ ,  $p'_i(\bar{k}_2|x)$ , ...,  $p'_i(\bar{k}_M|x)$ , содержащих сумму вероятностей всевозможных разметок фрагмента  $[i \dots I] \times [j - \frac{L}{2} \dots j + \frac{L}{2}]$ , проходящих через такие разметки  $\bar{k}_1, \bar{k}_2, \dots, \bar{k}_M$  столбца  $i$ , которые имеют ненулевую вероятность.

Теперь приближенное значение вероятности  $p(k_i = k|x_i)$  для пиксела  $t \equiv (i, j)$  вычисляется как сумма значений  $p_i(\bar{k}_m|x) + p'_i(\bar{k}_m|x)$  по разметкам  $\bar{k}_m, m = 1, \dots, M$ , имеющим в пикселе  $t \equiv (i, j)$  метку  $k$ :

$$\begin{aligned}
 p(k|x_i) &= \sum_{m=1}^M p_i(\bar{k}_m|x) \delta\{\bar{k}_m(t), k\} + \\
 &+ \sum_{m=1}^M p'_i(\bar{k}_m|x) \delta\{\bar{k}_m(t), k\}.
 \end{aligned}$$

В этой формуле символ  $\delta$  означает символ Кронекера. Вычислительная сложность алгоритма составляет  $O(T \cdot K^2 \cdot 3^{2L})$ , а затраты памяти  $O(I \cdot K \cdot 3^L)$ , где  $L$  есть толщина полосы прямоугольного участка поля зрения, используемого для приближенных вычислений значения (3).

### Результаты экспериментов

На рис. 2 представлена искусственная стереопара и результат ее обработки с помощью построчного алгоритма (внизу слева) и алгоритма, предлагаемого в данном разделе (внизу справа). Истинной моделью, на основании которой сгенерирована стереопара изображений, является сфера. Как видим, применение байесовских методов принятия решения для обра-

ботки стереопар позволяет улучшить точность распознавания и качество реконструируемых моделей.

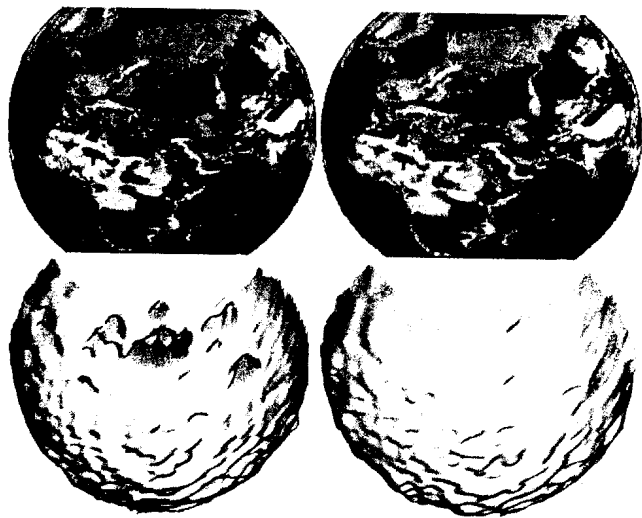


Рис. 2

#### 4. Целостная обработка стереопар, содержащих изображение кусочно-плоских объектов

В отличие от приведенного подхода построчной (одномерной) обработки стереопар, существуют целостные (двумерные) подходы [11–15]. Если при одномерной обработке решение о метке пиксела  $(i, j)$  принимается на основании распределения яркостей в строках  $j$  левого и правого снимков, то при двумерной обработке учитывается распределение яркости во всех строках. Целостная обработка стереопар повышает точность реконструкции и позволяет распознавать более широкий класс объектов. Например, трехмерные модели, восстанавливаемые в результате целостной обработки стереопар поверхностей, являются уже непрерывными в горизонтальном и вертикальном направлении, а не кусочно-непрерывными, как это было при построчной обработке. Особенности двумерных алгоритмов являются их высокие требования к вычислительным ресурсам: памяти и времени работы. Ниже предлагается новый алгоритм целостной обработки стереопар, основанный на предположении, что на снимках содержатся несколько плоских объектов.

#### Постановка задачи

Задана стереопара  $(x, x')$ . Известно, что объекты, изображенные на ней, имеют  $K$  уровней дальности. Разметкой левого изображения будет функция  $\bar{k}: T \rightarrow K$ , которая ставит в соответствие каждому пикселу номер его уровня дальности. Предполагается, что разметка является реализацией марковского случайного поля. Ее априорная вероятность задается с помощью потенциалов Гиббса  $g_{it'}(k, k')$  и определяется произведением  $p(\bar{k}) = \prod_{t \in T, t' \in \tau(t)} g_{it'}(k_t, k_{t'})$ , где  $\tau(t) = \{(i-1, j), (i+1, j), (i, j-1), (i, j+1)\}$  – это множество соседей пиксела  $t$ . Каждый из  $k \in K$  уровней дальности изображения определяется своим параметром – параллаксом  $d_k$ , который заранее не известен. Условная вероятность  $p(x_i | k_i)$  яркости  $x_i$  на левом снимке при условии, что пиксел  $t$  принадлежит уровню  $k_i$ , есть функция  $f(x_i, x', d_k)$ , зависящая от параллакса  $d_k$ .

Необходимо для каждого пиксела  $t \in T$  левого изображения найти его уровень дальности  $k_i^*$ , а также для каждого уровня  $k \in K$  указать его параллакс  $d_k^*$ . При этом вероятность  $p(x, d^*)$  наблюдения изображения  $x$  (функция правдоподобия выборки  $x$ ) и условная вероятность  $p(\bar{k}^* | x, d^*)$  разметки  $\bar{k}^*$  при условии заданного изображения  $x$  должны быть максимальными:

$$d^* = \operatorname{argmax}_d p(x, d) = \operatorname{argmax}_d \sum_{\bar{k} \in K^T} p(\bar{k}) \cdot p(x, d | \bar{k}) \quad (4)$$

$$\bar{k}^* = \operatorname{argmax}_{\bar{k} \in K^T} p(\bar{k} | x, d^*) = \operatorname{argmax}_{\bar{k} \in K^T} p(\bar{k}, x, d^*). \quad (5)$$

#### Решение задачи

Алгоритм решения задачи (4) основывается на наиболее правдоподобной оценке параметров системы в режиме самообучения [16] и оценке на этой основе параметров скрытого марковского поля [17]. Вначале с помощью

алгоритма самообучения будут оценены значения параметров  $d_k^*$ ,  $k \in K$ . После этого, на основании параметров  $d_k^*$  с помощью процедуры генератор Гиббса (Gibbs sampler) будет сгенерирована разметка  $\bar{k}^*$ , являющаяся решением задачи (5).

**Вычисление параметров  $d_k^*$  с помощью алгоритма самообучения.** Задача (4) состоит в вычислении параметров  $d_k^*$ ,  $k \in K$ . Алгоритм решения задачи работает итерационно. Каждая итерация состоит из двух шагов. Номер итерации обозначим символом  $r$ , а значения параметров на этой итерации  $d^{[r]}$ .

**Шаг 0.** Выбрать параметры  $d^{[0]}$  произвольным случайным образом.

**Шаг 1.** Для каждого пиксела  $t \in T$  и уровня  $k \in K$  вычислить величину  $\alpha_t(k)^{[r]}$ , которая имеет смысл приближенного значения вероятности  $p(k|x_t)$  реализации в пикселе  $t$  уровня  $k$  при условии известной яркости  $x_t$ :

$$\begin{aligned} \alpha_t(k)^{[r]} &= p(k|x_t, d^{[r]}) = \frac{p(x_t, k, d^{[r]})}{\sum_{k' \in K} p(x_t, k', d^{[r]})} = \\ &= \frac{p(k_t = k) \cdot p(x_t | k, d^{[r]})}{\sum_{k' \in K} p(k_t = k') \cdot p(x_t | k', d^{[r]})} = \\ &= \frac{p(k_t = k) \cdot f(x_t, x', d_k^{[r]})}{\sum_{k' \in K} p(k_t = k') \cdot f(x_t, x', d_{k'}^{[r]})}. \end{aligned}$$

Значение  $p(k_t = k)$  – априорная вероятность состояния  $k$  в пикселе  $t$ . Ее вычисление для заданного с помощью потенциалов Гиббса марковского поля – NP-полная задача. Для приближенного вычисления значений  $p(k_t = k)$ ,  $k \in K$ , используют генератор Гиббса (Gibbs sampler).

**Шаг 2.** Для каждого  $k \in K$  вычислить значение параметра  $d_k^{[r+1]}$ :

$$d^{[r+1]} = \arg \max_d \sum_{k \in K} \sum_{t \in T} \alpha_t(k) \cdot \log f(x_t, x', d_k).$$

**Шаг 3.** Если величина  $\|d^{[r+1]} - d^{[r]}\|$  больше, чем некий наперед заданный порог, то перейти к шагу 1, иначе завершить алгоритм.

**Обоснование алгоритма вычисления параметров  $d_k^*$ .** Обоснование алгоритма базируется на известной теореме М. Шлезингера о самообучении [10, 16]. После выполнения каждого цикла  $r$  справедливо соотношение  $\log p(x, d^{[r]}) \leq \log p(x, d^{[r+1]})$ . Однако, вообще говоря, сходимость последовательности  $\log p(x, d^{[r]})$ ,  $r = 0, 1, \dots, \infty$  к глобальному максимуму может не произойти.

Докажем соотношение  $\log p(x, d^{[r]}) \leq \log p(x, d^{[r+1]})$ . Для этого логарифм функции правдоподобия на шаге  $r$  запишем в виде

$$\begin{aligned} \log p(x, d^{[r]}) &= \sum_{\bar{k} \in K^T} \alpha(\bar{k})^{[r]} \cdot \log p(x, \bar{k}, d^{[r]}) - \\ &- \sum_{\bar{k} \in K^T} \alpha(\bar{k})^{[r]} \cdot \log \frac{p(x, \bar{k}, d^{[r]})}{p(x, d^{[r]})}, \end{aligned}$$

где числа  $\alpha(\bar{k})^{[r]}$  – неотрицательные числа, определенные для всевозможных разметок  $\bar{k} \in K^T$  и удовлетворяющие условию  $\sum_{\bar{k} \in K^T} \alpha(\bar{k})^{[r]} = 1$ .

Выберем их по правилу  $\alpha(\bar{k})^{[r]} = p(\bar{k}|x, d^{[r]})$ .

На шаге 2 алгоритма справедливо соотношение

$$\begin{aligned} d^{[r+1]} &= \arg \max_d \sum_{k \in K} \sum_{t \in T} \alpha_t(k) \cdot \log f(x_t, x', d_k) = \\ &= \arg \max_d \sum_{k \in K} \sum_{t \in T} \alpha_t(k) \cdot \log p(x_t | k, d) = \\ &= \arg \max_d \sum_{k \in K} \sum_{t \in T} \log p(x_t | k, d) \cdot \sum_{\bar{k} \in K^T: \bar{k}_t = k} \alpha(\bar{k})^{[r]} = \\ &= \arg \max_d \sum_{k \in K} \sum_{t \in T} \sum_{\bar{k} \in K^T: \bar{k}_t = k} \alpha(\bar{k})^{[r]} \cdot \log p(x_t | k, d) = \\ &= \arg \max_d \sum_{t \in T} \sum_{k \in K} \alpha(\bar{k})^{[r]} \cdot \log p(x_t | k, d). \end{aligned}$$

Воспользовавшись формулой Байеса, получим:

$$\begin{aligned} d^{[r+1]} &= \operatorname{argmax}_d \sum_{t \in T} \sum_{\bar{k} \in K^T} \alpha(\bar{k})^{[r]} \cdot \log \frac{p(x, k_t, d)}{p(k_t)} = \\ &= \operatorname{argmax}_d \sum_{\bar{k} \in K^T} \alpha(\bar{k})^{[r]} \cdot \sum_{t \in T} \log p(x_t | k_t, d) = \\ &= \operatorname{argmax}_d \sum_{\bar{k} \in K^T} \alpha(\bar{k})^{[r]} \cdot \log p(x | \bar{k}, d) = \\ &= \operatorname{argmax}_d \sum_{\bar{k} \in K^T} \alpha(\bar{k})^{[r]} \cdot \log p(x, \bar{k}, d). \end{aligned}$$

Таким образом, справедливо неравенство

$$\begin{aligned} \sum_{\bar{k} \in K^T} \alpha(\bar{k})^{[r]} \cdot \log p(x, \bar{k}, d^{[r]}) &\leq \\ &\leq \sum_{\bar{k} \in K^T} \alpha(\bar{k})^{[r]} \cdot \log p(x, \bar{k}, d^{[r+1]}). \end{aligned} \quad (6)$$

Поскольку числа  $\alpha(\bar{k})^{[r]}$ ,  $\bar{k} \in K^T$  выбраны как  $\alpha(\bar{k})^{[r]} = p(\bar{k} | x, d^{[r]})$ , на основании леммы Шеннона получим неравенство:

$$\begin{aligned} - \sum_{\bar{k} \in K^T} \alpha(\bar{k})^{[r]} \cdot \log \frac{p(x, \bar{k}, d^{[r]})}{p(x, d^{[r]})} &= \\ = - \sum_{\bar{k} \in K^T} \alpha(\bar{k})^{[r]} \cdot \log p(\bar{k} | x, d^{[r]}) &\leq \\ \leq - \sum_{\bar{k} \in K^T} \alpha(\bar{k})^{[r]} \cdot \log p(\bar{k} | x, d^{[r+1]}). \end{aligned} \quad (7)$$

Используя (6) и (7), получаем соотношение

$$\begin{aligned} \sum_{\bar{k} \in K^T} \alpha(\bar{k})^{[r]} \cdot \log p(x, \bar{k}, d^{[r]}) - \sum_{\bar{k} \in K^T} \alpha(\bar{k})^{[r]} \times \\ \times \log \frac{p(x, \bar{k}, d^{[r]})}{p(x, d^{[r]})} &\leq \sum_{\bar{k} \in K^T} \alpha(\bar{k})^{[r]} \cdot \log p(x, \bar{k}, d^{[r+1]}) - \\ - \sum_{\bar{k} \in K^T} \alpha(\bar{k})^{[r]} \cdot \log \frac{p(x, \bar{k}, d^{[r+1]})}{p(x, d^{[r+1]})}, \end{aligned}$$

которое и доказывает неравенство  $\log p(x, d^{[r]}) \leq \log p(x, d^{[r+1]})$ .

Алгоритм генерирования разметки  $\bar{k}^*$  для заданных параметров  $d_k^*$  с помощью генератора Гиббса. Задача состоит в поиске наиболее вероятной разметки  $\bar{k}^*$  для заданного изображения  $x$  и параметров  $d_k^*$ ,  $k \in K$ :

$$\begin{aligned} \bar{k}^* &= \operatorname{argmax}_{\bar{k} \in K^T} p(\bar{k} | x, d^*) = \\ &= \operatorname{argmax}_{\bar{k} \in K^T} p(\bar{k}, x, d^*) = \operatorname{argmax}_{\bar{k} \in K^T} p(\bar{k}) \cdot p(x | \bar{k}, d^*) = \\ &= \operatorname{argmax}_{\bar{k} \in K^T} \prod_{\substack{t \in T \\ t' \in \tau(t)}} g_{t'}(k_t, k_{t'}) \cdot p(x | \bar{k}, d^*) = \\ &= \operatorname{argmax}_{\bar{k} \in K^T} \prod_{\substack{t \in T \\ t' \in \tau(t)}} g_{t'}(k_t, k_{t'}) \cdot \prod_{t \in T} f(x_t, x_{t'}, d_{k_t}). \end{aligned} \quad (8)$$

В общем случае такая задача является *NP*-полной. Лишь в некоторых случаях, например, когда значения функции  $g$  удовлетворяют так называемому условию монотонности  $g_{t'}(k_1, k_2') \cdot g_{t'}(k_2, k_1') \leq g_{t'}(k_1, k_1') \cdot g_{t'}(k_2, k_2')$  для любых  $t, t' \in \tau(t)$  и  $k_1 < k_2$ ,  $k_1' < k_2'$ , задача (8) разрешима [3, 4]. Однако существуют методы поиска части ее решения [18] или приближенные методы решения. Ниже приводится один из них, известный как генератор Гиббса (*Gibbs sampler*).

Алгоритм работает циклически, генерируя последовательность разметок  $\bar{k}^{[0]}, \bar{k}^{[1]}, \dots, \bar{k}^{[r]}$ , сходящуюся к разметке  $\bar{k}^*$ , которая есть решением задачи.

**Шаг 0.** Произвольным образом сгенерировать случайную разметку  $\bar{k}^{[0]}$ , вероятность которой ненулевая.

**Шаг 1.** На основании разметки  $\bar{k}^{[r]}$  сгенерировать разметку  $\bar{k}^{[r+1]}$  по следующему правилу: для каждого пиксела  $t \in T$  вычислить условные вероятности  $p(k | x_t)$  реализации его возможных состояний  $k \in K$  и сгенерировать состояние  $k_t^{[r+1]}$  как случайную величину с распределением вероятности  $p(k | x_t)$ . Услов-

ная вероятность  $p(k|x_i)$ ,  $k \in K$  может быть рассчитана по приближенной формуле:

$$\begin{aligned} p(k|x_i) &= \frac{p(x_i, k_i = k)}{p(x_i)} = \frac{p(k_i = k) \cdot p(x_i | k_i = k)}{p(x_i)} = \\ &= \frac{p(k_i = k) \cdot f(x_i, x', d_{k_i^{[r]}})}{p(x_i)} \approx \\ &\approx \frac{\prod_{i' \in (i)} g_{i'}(k, k_i^{[r]}) \cdot f(x_i, x', d_{k_i^{[r]}})}{p(x_i)}. \end{aligned}$$

Следует отметить, что вероятность  $p(x_i)$  в этой формуле вычислять нет надобности, поскольку она входит как постоянный множитель для всех значений вероятности  $p(k|x_i)$ ,  $k \in K$ .

**Шаг 2.** Если величина  $\|k^{[r+1]} - k^{[r]}\|$  больше некоего наперед заданного порога, то перейти к шагу 1, иначе завершить алгоритм.

### Результаты экспериментов

На рис. 3 представлен результат стереоре-конструкции искусственной стереопары снимков. Истинная модель, на основании которой сгенерирована стереопара, содержится в правой нижней части рисунка. В левой нижней его части представлена модель, полученная в результате работы алгоритма, который описан в данном разделе. Как видно, результаты восстановления пространственной конфигурации сцены практически не отличаются от ее истинной модели.

Предложенный алгоритм обладает следующими достоинствами. В отличие от известных алгоритмов построчной (одномерной) обработки [6–9], он учитывает зависимость дальности в соседних пикселах реконструируемой сцены, как по горизонтали, так и по вертикали. Его применение возможно не только для объектов, характеризуемых одним параметром дальности  $d_k$ , но и для объектов, параметры дальности которых многомерны.

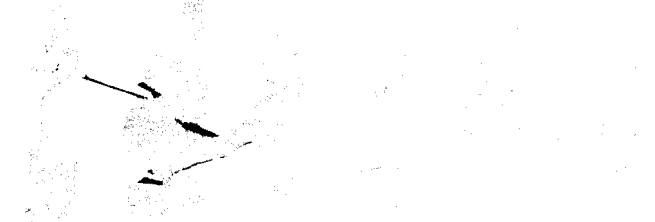


Рис. 3

**Заключение.** На основе предложенных алгоритмов, автором был создан аппаратно-программный комплекс «Стерео», реализующий технологию распознавания пространственных сцен и визуализацию их трехмерных моделей. Научная новизна работы состоит в создании новых эффективных алгоритмов распознавания стереопар, содержащих изображения непрерывных поверхностей. Результаты исследований автора были использованы при создании действующего макетного образца «Стереовизор», разрабатываемого в рамках Государственной научно-технической программы «Образный компьютер», а также при создании аппаратно-программного комплекса «Видеосекьюрити» для идентификации личности человека по изображениям его лица.

1. Пауль Л., Штанке Г. Трехмерные измерения, моделирование, обработка и визуализация объектов культуры (документирование, реставрация, презентация) // Электронні зображення та візуальні мистецтва EVA 2002: Пр. конф., Київ, 22–24 травня 2002 року. – Київ, 2002. – С. 52–28.
2. Гимельфарб Г.Л. Симметрический подход к задаче автоматических стереоскопических измерений в фотометрии // Кибернетика. – 1979. – № 2. – С. 73–82.
3. Schlesinger M.I., Flach B. Some solvable subclass of structural recognition problems // Czech Patt. Recog. Workshop 2000. – Praha. – 2000. – P. 55–62.
4. Schlesinger M.I., Flach B. Analysis of optimal labeling problems and their application to image segmen-

- tation and binocular stereovision // Шоста всеукр. міжнар. конф. з оброблення сигналів та розпізнавання образів (УкрОБРАЗ 2002). – Київ, 2002. – С. 15–18.
5. *Stan Z. Li*. Markov Random Field Modeling in Image Analysis // *Comp. Sci. Workbench*, Springer, 2001. – 323 p.
  6. *Шлезінгер М.І., Флах Б., Онищенко І.В.* Оптимізація кусково-лінійної деформації знімків у задачах машинного стереозору // П'ята всеукр. міжнар. конф. з оброблення сигналів та розпізнавання образів (УкрОБРАЗ 2000). – Київ, 2000. – С. 177–178.
  7. *Савчинський Б.Д.* Порівняльний аналіз алгоритмів стереозору в рамках байєсівської теорії прийняття рішень // Там же. – С. 25–28.
  8. *Рябокоть Д.І.* Створення тривимірних моделей об'єктів за стереопарами зображень для електронних музеїв // Електронні зображення та візуальні мистецтва EVA 2002, Київ, 22–24 травня 2002 року: Пр. конф. – Київ 2002. – С. 61–68.
  9. *Sara R.* Finding the Largest Unambiguous Component of Stereo Matching // *The 7-th Europ. Conf. on Comp. Vision*. – 2002. – 2. – P. 900–914.
  10. *Schlesinger M.I., Hlavac V.* Ten Lectures on Statistical and Structural Recognition. Computational Imaging and Vision, Kluwer Acad. Publ., Dordrecht–Boston–London, 2002. – 522 p.
  11. *Ishikawa H., Geiger D.* Segmentation by Grouping Junctions // *IEEE Comp. Soci. Conf. on Comp. Vision and Patt. Recog.*, Santa Barbara, CA, June, 1998.
  12. *Boykov Yu., Veksler O., Zabih R.* Fast approximate energy minimization via graph cuts. // *Intern. Conf. of Comp. Vision*, 1999. – P. 377–384.
  13. *Kolmogorov V., Zabih R.* Computing visual correspondence with occlusions via graph cuts. // *Ibid*, 2001. – P. 508–515.
  14. *Geman S., Geman D.* Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images // *IEEE Trans. on PAMI*. – 1984. – 6, № 6. – P. 721–741.
  15. *Рябокоть Д.І.* Постобработка результатов стереореконструкции поверхностей как восстановление скрытых параметров марковского поля // *УСиМ*. – 2003. – № 4. – С. 56–61.
  16. *Шлезінгер М.І.* Взаимосвязь обучения и самообучения в распознавании образов // *Кибернетика* – 1968. – № 2. – С. 81–88.
  17. *Unifying Registration and Segmentation for Multi-Sensor Images / B. Flach, E. Kask, D. Schlesinger et al.* // *DAGM2002 – Patt. Recog. Symp.* – Curriish, 2002.
  18. *Ковтун І.В.* Поиск части оптимальной разметки некоторого NP-полного подкласса (max,+) задач // *УСиМ*. – 2003. – № 6. – С. 33–38.

Поступила 14.09.2004

Тел. для справок: (044) 266-6208 (Київ)

e-mail: [dima@image.kiev.ua](mailto:dima@image.kiev.ua)

© Д.І. Рябокоть, 2005