

Фундаментальные и прикладные проблемы Computer Science

УДК 004.93'1 : 519.157

М.И. Шлезингер, В.В. Гигиняк

Решение (max,+)-задач структурного распознавания с помощью их эквивалентных преобразований.* II

Исследованы задачи структурного распознавания, которые сводятся к задачам дискретной оптимизации, известным как (max, +)-разметки. Описан механизм эквивалентных преобразований (max, +)-задач и их решение на основе таких преобразований. Класс решаемых таким методом задач включает в себя известные ациклические и супермодулярные (max, +)-задачи.

Structural recognition problems are analyzed which can be reduced to optimization problems known as a (max, +)-labeling. A mechanism of equivalent transformations of (max, +)-problems is described as well as their solution based on such transformations. The domain of the problems solvable in such a way includes the known acyclic and supermodular (max, +)-labeling problems.

Досліджено задачі структурного розпізнавання, що зводяться до задач дискретної оптимізації, відомих як (max, +)-розмітки. Описано механізм еквівалентних перетворень (max, +)-задач та їх розв'язання на основі таких перетворень. Клас розв'язних таким методом задач охоплює відомі ациклічні та супермодулярні (max, +)-задачі.

Введение. Статья является продолжением работы, первая часть которой опубликована [1]. В данной части сохранены терминология и обозначения, введенные в первой части. Нумерация разделов и теорем автономная и не продолжает нумерацию первой части. Ссылки на разделы даны в пределах данной части. Ссылки на предыдущую часть будут оговорены.

Объектом исследований является (max,+)-задача структурного распознавания, формулируемая на основании следующих понятий.

Пусть заданы T – множество объектов, K – множество меток, $\mathfrak{Z} \subset K \times K$ – подмножество пар объектов, называемых соседними. Пусть для каждой пары $tt' \in \mathfrak{Z}$ соседних объектов и каждой пары $k \in K, k' \in K$ меток задано число $g_{tt'}(k, k')$, а для каждого объекта $t \in T$ и каждой метки $k \in K$ – число $q_t(k)$. Разметка – это функция \bar{k} вида $T \rightarrow K$, качество которой определяется как

$$G(\bar{k}) = \sum_{tt' \in \mathfrak{Z}} g_{tt'}(k(t), k(t')) + \sum_{t \in T} q_t(k(t)).$$

* Работа выполнена в рамках проекта EU INTAS PRINCESS 04-77-7347.

Задача (max,+)-разметки состоит в отыскании разметки с наилучшим качеством.

В первой части работы описан подход к решению этой задачи, основанный на понятиях эквивалентных и тривиальных (max,+)-задач и понятии энергии задачи. В данной части будут использованы следующие два результата, доказанные в первой части.

Теорема 1. Пусть $N(t) = \{t' \in T | tt' \in \mathfrak{Z}\}$ – множество объектов, соседних с объектом t , а Φ – массив $(\varphi_{tt'}(k) | t \in T, t' \in N(t), k \in K)$ чисел, называемых потенциалами.

Пусть z и z' – две (max,+)-задачи, определенные на множествах T, K и \mathfrak{Z} , первая из которых задана числами $g_{tt'}(k, k')$ и $q_t(k)$, а вторая – числами $g'_{tt'}(k, k')$ и $q'_t(k)$.

Задачи z и z' эквивалентны тогда и только тогда, когда существует массив Φ потенциалов, такой, что

$$g'_{tt'}(k, k') = g_{tt'}(k, k') + \varphi_{tt'}(k) + \varphi_{t't}(k'),$$
$$tt' \in \mathfrak{Z}, \quad k \in K, \quad k' \in K,$$

$$q'_t(k) = q_t(k) - \sum_{t' \in N(t)} \varphi_{tt'}(k), \quad t \in T, \quad k \in K. \quad \blacksquare$$

Теорема 2. Пусть z – это $(\max,+)$ -задача на множествах T, K, \mathfrak{Z} , характеризуемая числами $g_{tt'}(k, k')$ и $q_t(k)$. Пусть для задачи z существует эквивалентная ей тривиальная задача. Пусть Φ' – массив потенциалов, минимизирующий энергию

$$E(\Phi) = \sum_{tt' \in \mathfrak{Z}} \max_{k, k'} (g_{tt'}(k, k') + \varphi_{tt'}(k) + \varphi_{tt'}(k')) + \sum_{t \in T} \max_k \left(q_t(k) - \sum_{t' \in N(t)} \varphi_{tt'}(k) \right),$$

$$E(\Phi') = \min_{\Phi} E(\Phi).$$

В таком случае задача z' на множествах T, K, \mathfrak{Z} с числами

$$g'_{tt'}(k, k') = g_{tt'}(k, k') + \varphi_{tt'}(k) + \varphi_{tt'}(k'),$$

$$tt' \in \mathfrak{Z}, k \in K, k' \in K,$$

$$q'_t(k) = q_t(k) - \sum_{t' \in N(t)} \varphi_{tt'}(k), t \in T, k \in K,$$

тривиальна. ■

В первой части был описан алгоритм минимизации энергии произвольной $(\max,+)$ -задачи, основанный на том факте, что энергия является выпуклой, хотя и недифференцируемой функцией от потенциалов. В данной части будет исследована задача минимизации энергии для частного случая соседства \mathfrak{Z} , характерного для изображений.

1. Особенности минимизации энергии $(\max,+)$ -задач на двумерном поле зрения

Рассмотрим частный случай $(\max,+)$ -задач, когда множество T объектов есть множество пикселей, образующее прямоугольный участок двумерной целочисленной решетки. Именно,

$$T = \{(i, j) | 1 \leq i \leq m, 1 \leq j \leq n\},$$

где m – количество строк, а n – количество столбцов в изображении. Будем считать также, что отношение соседства \mathfrak{Z} есть множество пар

$$\{(i-1, j), (i, j) | 1 < i \leq m, 1 \leq j \leq n\} \cup \{(i, j-1), (i, j) | 1 \leq i \leq m, 1 < j \leq n\},$$

естественное для изображений.

Как видим, множество \mathfrak{Z} распадается на два непересекающиеся подмножества: подмножество

$$\mathfrak{Z}^B = \{(i-1, j), (i, j) | 1 < i \leq m, 1 \leq j \leq n\}$$

пар вертикально соседствующих пикселей и подмножество

$$\mathfrak{Z}^G = \{(i, j-1), (i, j) | 1 \leq i \leq m, 1 < j \leq n\}$$

пар горизонтально соседствующих пикселей,

$$\mathfrak{Z} = \mathfrak{Z}^G \cup \mathfrak{Z}^B.$$

Подобным образом множество $N(t)$ пикселей, соседних с t , распадается на подмножество $N^G(t)$ пикселей, соседних с t по горизонтали, и $N^B(t)$ – по вертикали. Существенно, что каждое из соседств \mathfrak{Z}^G и \mathfrak{Z}^B не образует циклов, в связи с чем вычисление величин

$$\max_{\bar{k} \in K^T} \left[\sum_{tt' \in \mathfrak{Z}^G} g_{tt'}(k(t), k(t')) + \sum_{t \in T} q_t(k(t)) \right],$$

$$\max_{\bar{k} \in K^T} \left[\sum_{tt' \in \mathfrak{Z}^B} g_{tt'}(k(t), k(t')) + \sum_{t \in T} q_t(k(t)) \right]$$

не составляет труда и легко реализуется с помощью динамического программирования.

В связи с указанными особенностями задачу минимизации энергии

$$E(\Phi) = \sum_{tt' \in \mathfrak{Z}} \max_{k, k'} (g_{tt'}(k, k') + \varphi_{tt'}(k) + \varphi_{tt'}(k')) + \sum_{t \in T} \max_k \left(q_t(k) - \sum_{t' \in N(t)} \varphi_{tt'}(k) \right)$$

можно видоизменить так, что для ее решения можно будет применить алгоритмы, более эффективные, чем алгоритмы минимизации энергии $(\max,+)$ -задач общего вида.

Введем в рассмотрение массив Ψ чисел $\psi_t(k)$, $t \in T, k \in K$, которые также назовем потенциалами.

Теорема 3. Справедливо равенство

$$\min_{\Phi} \left[\sum_{tt' \in \mathfrak{Z}} \max_{k, k'} (g_{tt'}(k, k') + \varphi_{tt'}(k) + \varphi_{tt'}(k')) + \sum_{t \in T} \max_k (\psi_t(k) - \sum_{t' \in N(t)} \varphi_{tt'}(k)) \right]$$

$$\begin{aligned}
& + \sum_{t \in T} \max_k \left(q_t(k) - \sum_{t' \in N(t)} \varphi_{t'}(k) \right) = \\
& = \min_{\Psi} \left[\max_{k \in K^T} \left(\sum_{t' \in \mathfrak{S}^T} g_{t'}(k(t), k(t')) + \right. \right. \\
& \quad \left. \left. + \frac{1}{2} \sum_{t \in T} q_t(k(t)) + \sum_{t \in T} \psi_t(k(t)) \right) + \right. \\
& \quad \left. + \max_{k \in K^T} \left(\sum_{t' \in \mathfrak{S}^B} g_{t'}(k(t), k(t')) + \right. \right. \\
& \quad \left. \left. + \frac{1}{2} \sum_{t \in T} q_t(k(t)) - \sum_{t \in T} \psi_t(k(t)) \right) \right]. \quad (1)
\end{aligned}$$

Доказательство. 1. Рассмотрим вторую сумму в левой части (1). Для нее справедливы следующие равенство и неравенство:

$$\begin{aligned}
& \sum_{t \in T} \max_k \left(q_t(k) - \sum_{t' \in N(t)} \varphi_{t'}(k) \right) = \\
& = \sum_{t \in T} \max_k \left(\frac{1}{2} q_t(k) + \psi_t(k) - \sum_{t' \in N^T(t)} \varphi_{t'}(k) + \right. \\
& \quad \left. + \frac{1}{2} q_t(k) - \psi_t(k) - \sum_{t' \in N^B(t)} \varphi_{t'}(k) \right) \leq \\
& \leq \sum_{t \in T} \max_k \left(\frac{1}{2} q_t(k) + \psi_t(k) - \sum_{t' \in N^T(t)} \varphi_{t'}(k) \right) + \\
& \quad + \sum_{t \in T} \max_k \left(\frac{1}{2} q_t(k) - \psi_t(k) - \sum_{t' \in N^B(t)} \varphi_{t'}(k) \right). \quad (2)
\end{aligned}$$

2. Неравенство (2) справедливо при любых значениях чисел $\psi_t(k)$. Однако при определенных их значениях это неравенство обращается в равенство. А именно, это происходит при

$$\psi_t(k) = \frac{1}{2} \left[\sum_{t' \in N^T(t)} \varphi_{t'}(k) - \sum_{t' \in N^B(t)} \varphi_{t'}(k) \right].$$

Поэтому справедливо равенство

$$\sum_{t \in T} \max_k \left(q_t(k) - \sum_{t' \in N(t)} \varphi_{t'}(k) \right) =$$

$$\begin{aligned}
& = \min_{\Psi} \left[\sum_{t \in T} \max_k \left(\frac{1}{2} q_t(k) + \psi_t(k) - \sum_{t' \in N^T(t)} \varphi_{t'}(k) \right) + \right. \\
& \quad \left. + \sum_{t \in T} \max_k \left(\frac{1}{2} q_t(k) - \psi_t(k) - \sum_{t' \in N^B(t)} \varphi_{t'}(k) \right) \right]. \quad (3)
\end{aligned}$$

3. Представим массив

$$\Phi = (\varphi_{t'}(k) \mid t \in T, t' \in N(t), k \in K)$$

в виде пары (Φ^T, Φ^B) двух массивов

$$\Phi^T = (\varphi_{t'}(k) \mid t \in T, t' \in N^T(t), k \in K),$$

$$\Phi^B = (\varphi_{t'}(k) \mid t \in T, t' \in N^B(t), k \in K).$$

Обозначим Q левую часть равенства (1) и запишем ее с учетом доказанного равенства (3):

$$\begin{aligned}
Q & = \min_{\Phi} \left\{ \sum_{t' \in \mathfrak{S}} \max_{k, k'} (g_{t'}(k, k') + \varphi_{t'}(k) + \varphi_{t'}(k')) + \right. \\
& \quad \left. + \min_{\Psi} \left[\sum_{t \in T} \max_k \left(\frac{1}{2} q_t(k) + \psi_t(k) - \sum_{t' \in N^T(t)} \varphi_{t'}(k) \right) + \right. \right. \\
& \quad \left. \left. + \sum_{t \in T} \max_k \left(\frac{1}{2} q_t(k) - \psi_t(k) - \sum_{t' \in N^B(t)} \varphi_{t'}(k) \right) \right] \right\}. \quad (4)
\end{aligned}$$

4. Преобразуем выражение (4), используя очевидные правила

$$\min_{\Phi} [f_1(\Phi) + \min_{\Psi} f_2(\Phi, \Psi)] =$$

$$= \min_{\Psi} \min_{\Phi} [f_1(\Phi) + f_2(\Phi, \Psi)],$$

$$\min_{\Phi} f(\Phi) = \min_{\Phi^T} \min_{\Phi^B} f(\Phi^T, \Phi^B),$$

$$\sum_{t' \in \mathfrak{S}} f_{t'} = \sum_{t' \in \mathfrak{S}^T} f_{t'} + \sum_{t' \in \mathfrak{S}^B} f_{t'},$$

и получим

$$\begin{aligned}
Q & = \min_{\Psi} \min_{\Phi^T} \min_{\Phi^B} \left\{ \sum_{t' \in \mathfrak{S}^T} \max_{k, k'} (g_{t'}(k, k') + \right. \\
& \quad \left. + \varphi_{t'}(k) + \varphi_{t'}(k')) + \right. \\
& \quad \left. + \sum_{t \in T} \max_k \left[\left(\frac{1}{2} q_t(k) + \psi_t(k) - \sum_{t' \in N^T(t)} \varphi_{t'}(k) \right) + \right. \right.
\end{aligned}$$

$$+ \sum_{t' \in \mathfrak{Z}^B} \max_{k, k'} (g_{t'}(k, k') + \varphi_{t'}(k) + \varphi_{t'}(k')) + \\ + \sum_{t \in T} \max_k \left[\frac{1}{2} q_t(k) - \psi_t(k) - \sum_{t' \in N^B(t)} \varphi_{t'}(k) \right] \Bigg\}.$$

5. Преобразуем последнее выражение, используя очевидное правило,

$$\min_{\Phi^\Gamma} \min_{\Phi^B} (f_1(\Phi^\Gamma) + f_2(\Phi^B)) = \\ = \min_{\Phi^\Gamma} f_1(\Phi^\Gamma) + \min_{\Phi^B} f_2(\Phi^B):$$

$$Q = \min_{\Psi} \left\{ \min_{\Phi^\Gamma} \left[\sum_{t' \in \mathfrak{Z}^\Gamma} \max_{k, k'} (g_{t'}(k, k') + \varphi_{t'}(k) + \varphi_{t'}(k')) + \right. \right. \\ \left. \left. + \sum_{t \in T} \max_k \left(\frac{1}{2} q_t(k) + \psi_t(k) - \sum_{t' \in N^\Gamma(t)} \varphi_{t'}(k) \right) \right] + \right. \\ \left. + \min_{\Phi^B} \left[\sum_{t' \in \mathfrak{Z}^B} \max_{k, k'} (g_{t'}(k, k') + \varphi_{t'}(k) + \varphi_{t'}(k')) + \right. \right. \\ \left. \left. + \sum_{t \in T} \max_k \left(\frac{1}{2} q_t(k) - \psi_t(k) - \sum_{t' \in N^B(t)} \varphi_{t'}(k) \right) \right] \right\}. \quad (5)$$

6. Рассмотрим первое слагаемое

$$\min_{\Phi^\Gamma} \left[\sum_{t' \in \mathfrak{Z}^\Gamma} \max_{k, k'} (g_{t'}(k, k') + \varphi_{t'}(k) + \varphi_{t'}(k')) + \right. \\ \left. + \sum_{t \in T} \max_k \left(\frac{1}{2} q_t(k) + \psi_t(k) - \sum_{t' \in N^\Gamma(t)} \varphi_{t'}(k) \right) \right] \quad (6)$$

в выражении (5). Оно относится к (max,+)-задаче на множестве T с ациклическим соседством \mathfrak{Z}^Γ и числами

$$g_{t'}(k, k'), \quad t' \in \mathfrak{Z}^\Gamma, \quad k \in K, \quad k' \in K, \quad \text{и} \\ \frac{1}{2} q_t(k) + \psi_t(k), \quad t \in T, \quad k \in K.$$

Выражение (6) обозначает минимальную энергию на множестве задач, эквивалентных этой задаче. В силу ациклическости соседства \mathfrak{Z}^Γ и теоремы 1, доказанной в первой части работы, такая задача эквивалентна некоторой

тривиальной задаче, и искомым минимумом (6) равен качеству

$$\max_{k \in K^T} \left[\sum_{t' \in \mathfrak{Z}^\Gamma} g_{t'}(k(t), k(t')) + \right. \\ \left. + \frac{1}{2} \sum_{t \in T} q_t(k(t)) + \sum_{t \in T} \psi_t(k(t)) \right]$$

оптимальной в этой задаче разметки. Подобным образом доказывается, что второе слагаемое

$$\min_{\Phi^B} \left[\sum_{t' \in \mathfrak{Z}^B} \max_{k, k'} (g_{t'}(k, k') + \varphi_{t'}(k) + \varphi_{t'}(k')) + \right. \\ \left. + \sum_{t \in T} \max_k \left(\frac{1}{2} q_t(k) - \psi_t(k) - \sum_{t' \in N^B(t)} \varphi_{t'}(k) \right) \right]$$

в выражении (5) равно качеству

$$\max_{k \in K^T} \left[\sum_{t' \in \mathfrak{Z}^B} g_{t'}(k(t), k(t')) + \right. \\ \left. + \frac{1}{2} \sum_{t \in T} q_t(k(t)) - \sum_{t \in T} \psi_t(k(t)) \right]$$

оптимальной разметки в другой (max,+)-задаче, а именно задаче с соседством \mathfrak{Z}^B .

Таким образом,

$$Q = \min_{\Psi} \left\{ \max_{k \in K^T} \left[\sum_{t' \in \mathfrak{Z}^\Gamma} g_{t'}(k(t), k(t')) + \right. \right. \\ \left. \left. + \frac{1}{2} \sum_{t \in T} q_t(k(t)) + \sum_{t \in T} \psi_t(k(t)) \right] + \right. \\ \left. + \max_{k \in K^T} \left[\sum_{t' \in \mathfrak{Z}^B} g_{t'}(k(t), k(t')) + \right. \right. \\ \left. \left. + \frac{1}{2} \sum_{t \in T} q_t(k(t)) - \sum_{t \in T} \psi_t(k(t)) \right] \right\}.$$

2. Алгоритм минимизации энергии

В силу справедливости теоремы 3 задача минимизации энергии

$$E(\Phi) = \sum_{t' \in \mathfrak{Z}} \max_{k, k'} (g_{t'}(k, k') + \varphi_{t'}(k) + \varphi_{t'}(k')) +$$

$$+ \sum_{t \in T} \max_k \left(q_t(k) - \sum_{t' \in N(t)} \varphi_{t'}(k) \right)$$

по потенциалам $\varphi_{t'}(k)$, $t \in T$, $t' \in N(t)$, $k \in K$, сведена к минимизации энергии

$$\begin{aligned} Q(\Psi) = & \max_{\bar{k} \in K^T} \left[\sum_{t' \in \mathfrak{Z}^T} g_{t'}(k(t), k(t')) + \right. \\ & \left. + \frac{1}{2} \sum_{t \in T} (q_t(k(t)) + \psi_t(k(t))) \right] + \\ & + \max_{\bar{k} \in K^T} \left[\sum_{t' \in \mathfrak{Z}^B} g_{t'}(k(t), k(t')) + \right. \\ & \left. + \frac{1}{2} \sum_{t \in T} (q_t(k(t)) - \psi_t(k(t))) \right] \end{aligned} \quad (7)$$

по потенциалам $\psi_t(k)$, $t \in T$, $k \in K$.

Каждая разметка $\bar{k} : T \rightarrow K$ характеризуется двумя качествами: горизонтальным

$$\begin{aligned} G^\Gamma(\bar{k}, \Psi) = & \sum_{t' \in \mathfrak{Z}^\Gamma} g_{t'}(k(t), k(t')) + \\ & + \frac{1}{2} \sum_{t \in T} (q_t(k(t)) + \psi_t(k(t))) \end{aligned}$$

и вертикальным

$$\begin{aligned} G^B(\bar{k}, \Psi) = & \sum_{t' \in \mathfrak{Z}^B} g_{t'}(k(t), k(t')) + \\ & + \frac{1}{2} \sum_{t \in T} (q_t(k(t)) - \psi_t(k(t))). \end{aligned}$$

Горизонтальное качество $G^\Gamma(\bar{k}, \Psi)$ любой разметки \bar{k} является линейной, а следовательно, и выпуклой функцией потенциалов Ψ , равно как и ее вертикальное качество $G^B(\bar{k}, \Psi)$.

Значит, и функции

$$G^\Gamma(\Psi) = \max_{\bar{k} \in K^T} G^\Gamma(\bar{k}, \Psi), \quad G^B(\Psi) = \max_{\bar{k} \in K^T} G^B(\bar{k}, \Psi)$$

являются выпуклыми функциями потенциалов Ψ . И наконец, выпуклой функцией потенциалов является и энергия

$$Q(\Psi) = \max_{\bar{k} \in K^T} G^\Gamma(\bar{k}, \Psi) + \max_{\bar{k} \in K^T} G^B(\bar{k}, \Psi) \quad (8)$$

как сумма двух выпуклых функций. Существенно, что каждое из слагаемых в выражениях (7) и (8), равно как и разметки

$$\bar{k}^{*\Gamma} = \arg \max_{\bar{k} \in K^T} G^\Gamma(\bar{k}, \Psi), \quad \bar{k}^{*B} = \arg \max_{\bar{k} \in K^T} G^B(\bar{k}, \Psi)$$

легко вычисляются, так как здесь речь идет о решении двух ациклических (max,+)-задач.

В силу выпуклости энергии $Q(\Psi)$ ее минимизацию можно выполнить методом обобщенного градиентного спуска, подобно тому, как это было сделано в первой части работы для минимизации энергии (max,+)-задач общего вида. Применительно к рассматриваемому здесь случаю, обобщенный градиентный спуск заключается в следующем.

Пусть γ_i , $i = 0, 1, \dots$, – последовательность чисел такая, что

$$\lim_{i \rightarrow \infty} \gamma_i = 0, \quad \lim_{n \rightarrow \infty} \sum_{i=0}^n \gamma_i = \infty.$$

Пусть Ψ_0 – произвольный набор потенциалов $\psi_t(k)$, $t \in T$, $k \in K$, например $\psi_t(k) = 0$.

Пусть Ψ_i , $i = 1, 2, \dots$, – последовательность наборов потенциалов такая, что

$$\Psi_{i+1} = \Psi_i - \gamma_i \cdot \Delta \Psi_i,$$

где $\Delta \Psi_i$ – обобщенный градиент функции $Q(\Psi)$ в точке Ψ_i .

В таком случае

$$\lim_{i \rightarrow \infty} Q(\Psi_i) = \min_{\Psi} Q(\Psi).$$

Для применения этого общего принципа в рассмотренном контексте следует лишь указать алгоритм вычисления градиента функции $Q(\Psi)$ для произвольного набора Ψ потенциалов. Этот алгоритм исключительно прост.

Градиент $\Delta \Psi$ есть массив чисел $\Delta \psi_t(k)$, $t \in T$, $k \in K$. Одна компонента этого массива определяется следующим образом.

Пусть

$$\bar{k}^\Gamma = \arg \max_{\bar{k} \in K^T} G^\Gamma(\bar{k}, \Psi), \quad \bar{k}^B = \arg \max_{\bar{k} \in K^T} G^B(\bar{k}, \Psi).$$

В таком случае

$$\Delta\psi_t(k) = 1, \text{ если } k^\Gamma(t) = k, k^B(t) \neq k,$$

$$\Delta\psi_t(k) = -1, \text{ если } k^\Gamma(t) \neq k, k^B(t) = k,$$

$$\Delta\psi_t(k) = 0 \text{ во всех прочих случаях.}$$

3. Иллюстрации разработанного подхода Конкретизация разработанных алгоритмов

В данной работе описаны два алгоритма решения (max,+)-задач методом эквивалентного их преобразования. Алгоритм 1, описанный в [1] выполняет минимизацию энергии для любой (max,+)-задачи, поданной на его вход. Алгоритм 2 выполняет минимизацию энергии (max,+)-задач, определенных для множеств T объектов в виде прямоугольного участка целочисленной решетки. Оба алгоритма сформулированы с точностью до последовательности $\gamma_i, i=1,2,\dots$ шагов. Кроме того, приведенное описание алгоритмов не является полным, так как не указано условие их останова.

Приведенные далее примеры, за исключением описанного в следующем подразделе, получены с помощью алгоритмов, в которых последовательность $\gamma_i, i=1,2,\dots$ была гармоническим рядом $\gamma_i = 1/i$. Условием останова являлось нулевое значение обобщенного градиента энергии, вычисленного так, как описано в разд. 4 первой части статьи или в предыдущем разделе данной части. Авторы не утверждают, что именно такая конкретизация описанных алгоритмов оптимальна в каком-либо смысле. Такая конкретизация алгоритмов сделана потому, что при их реализации величины шагов γ_i и условие останова следовало по крайней мере как-нибудь определить. Что касается оптимального выбора шагов γ_i и условия останова, то это потребует дальнейших, по-видимому, немалых усилий.

Пример решения (v, &)-задач с помощью минимизации энергии

В конце первой части работы сказано, что разработанный метод минимизации энергии есть средство решения не только (max,+)-задач, но и (v, &)-задач, причем более сильное,

чем обычно применяемый для этой цели метод, известный как релаксационная разметка [2]. Приведем пример, иллюстрирующий это утверждение.

Рассмотрим простейшую (v, &)-задачу на поле зрения T , состоящем всего из двух столбцов и четырех строк. Элементы этого поля (пиксели) представлены на рис. 1 в виде квадратов, пронумерованных числами от единицы до восьми. Множество K состоит из трех меток: a, b и c .

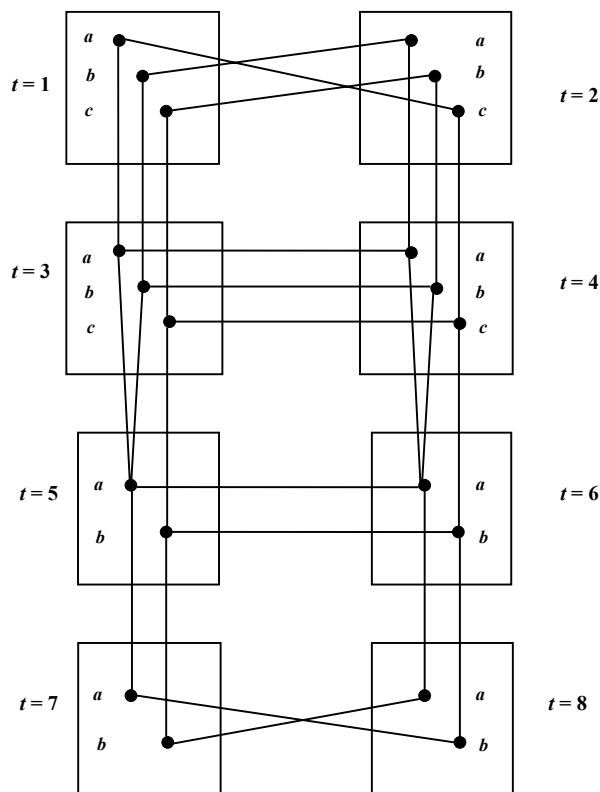


Рис. 1

Внутри каждого квадрата t изображены черные кружочки, каждый из которых соответствует той или иной метке k , обозначенной рядом с кружочком. В каждом квадрате t нарисованы те и только те кружочки, которые соответствуют таким меткам k , что $q_t(k) = 1$. Так, например, по рисунку видно, что для $t=1, 2, 3, 4$ число $q_t(k)$ равно единице для любого $k = a, b, c$, а для $t=5, 6, 7, 8$ число $q_t(k)$ равно единице только для $k = a, b$.

Числа $g_{tt'}(k, k')$ представлены на рисунке отрезками, которые соединяют те или иные кружочки в соседних квадратах. А именно, если $g_{tt'}(k, k') = 1$, то кружочек в квадрате t , соответствующий метке k , соединен с кружочком в квадрате t' , соответствующим метке k' . Так, по рисунку видно, например, что $q_{12}(a, a) = 0$, а $q_{12}(a, c) = 1$.

В терминах приведенного графического представления разметкой является выбор в каждом квадрате одного и только одного кружочка, а допустимая разметка соответствует такому выбору, при котором кружочки, выбранные в соседних квадратах, соединены отрезком. Представленная на рис. 1 задача состоит в ответе на вопрос, возможен ли такой выбор для этого рисунка.

Общеизвестный метод решения этой задачи основан на использовании необходимого, но не достаточного условия положительного решения $(\vee, \&)$ -задачи, которое было сформулировано в первой части работы: для того, чтобы $(\vee, \&)$ -задача имела положительное решение, необходимо, чтобы она содержала в себе непустую согласованную часть. Это условие проверяется с помощью общеизвестной релаксационной разметки. Оно выполняется для поставленной задачи, так как сама задача уже является непустой и согласованной. Нетрудно также увидеть, что представленная задача не имеет положительного решения, что свидетельствует о недостаточности этого условия. Описанная идея минимизации энергии $(\max, +)$ -задачи позволяет проверять положительность решения $(\vee, \&)$ -задачи более сильными методами, чем общеизвестные. Эта проверка состоит в следующем.

По представленной для решения $(\vee, \&)$ -задаче с величинами $g_{tt'}(k, k') \in \{0, 1\}$ и $q_t(k) \in \{0, 1\}$ формулируется $(\max, +)$ -задача с величинами $\bar{g}_{tt'}(k, k') \in R$ и $\bar{q}_t(k) \in R$, такая, что

$$\begin{aligned} \bar{g}_{tt'}(k, k') &= 0, \text{ если } g_{tt'}(k, k') = 1, \\ \bar{g}_{tt'}(k, k') &= -L, \text{ если } g_{tt'}(k, k') = 0, \\ \bar{q}_t(k) &= 0, \text{ если } q_t(k) = 1, \\ \bar{q}_t(k) &= -L, \text{ если } q_t(k) = 0, \end{aligned}$$

где L – любое положительное число, например 100.

Затем проверяется, существует ли $(\max, +)$ -задача, эквивалентная полученной, но с отрицательной энергией. Если такое эквивалентное преобразование возможно, то очевидно, что в полученной $(\max, +)$ -задаче отсутствует разметка с нулевым качеством, а в исходной $(\vee, \&)$ -задаче отсутствует допустимая разметка.

Реализация этого нового способа для $(\vee, \&)$ -задачи, представленной на рис. 1, приведена на рис. 2. На этом рисунке определенный набор потенциалов $\varphi_{tt'}(k)$ показан графически в виде наборов треугольников, расположенных рядом с кружочками. Потенциалы принимают значения $-3, -2, -1, 0, +1, +2, +3$. Эти числа изображены на рисунке в виде одного, двух или трех треугольников, расположенных рядом с кружочком, к которому они относятся. Отсутствие треугольников рядом с кружочком предполагает нулевое значение потенциала. Треугольники рядом с кружочками направлены либо вершиной к кружочкам, либо вершиной от кружочка. В первом случае треугольники отображают отрицательные значения потенциалов, т.е. увеличивающие качество кружочка. Треугольники, ориентированные вершиной от кружочка, отображают положительные значения потенциалов, т.е. уменьшающие качество кружочка.

Потенциалы, представленные на рис. 2, преобразуют исходную $(\max, +)$ -задачу в эквивалентную ей, но с энергией, равной (-1) . Действительно, для всех пар tt' соседних квадратов величина $\max_{k, k'} [\bar{g}_{tt'}(k, k') + \varphi_{tt'}(k) + \varphi_{t't}(k')]$ осталась равной нулю. Для второго сверху квадрата t^* в левом столбце величина

$$\max_k \left[\bar{q}_{t^*}(k) - \sum_{t' \in N(t^*)} \varphi_{t^*}(k) \right] \text{ стала равной } (-1).$$

Для всех остальных квадратов $t \neq t^*$ величина

$$\max_k \left[\bar{q}_t(k) - \sum_{t' \in N(t)} \varphi_{t^*}(k) \right] \text{ осталась рав-$$

ной нулю.

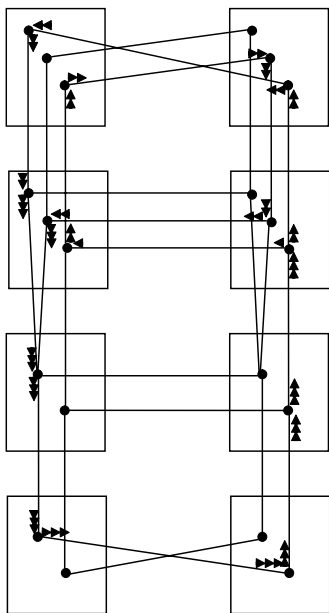


Рис. 2

Показанные на рис. 2 значения потенциалов были найдены вручную. Естественно, они могут быть найдены и автоматически с помощью предложенных Алгоритма 1 или Алгоритма 2. Для решения (\vee , $\&$)-задачи, представленной на рис. 1, был использован Алгоритм 1. После 11 итераций этого алгоритма был найден набор потенциалов, обеспечивающий отрицательную энергию соответствующей (\max , $+$)-задачи.

При некотором незначительном усложнении рассмотренного примера можно показать существенное отличие предложенного в работе подхода к решению (\max , $+$)-задач от известных подходов, идейно близких к предложенному.

Рассмотрим (\max , $+$)-задачу, представленную на рис. 3.

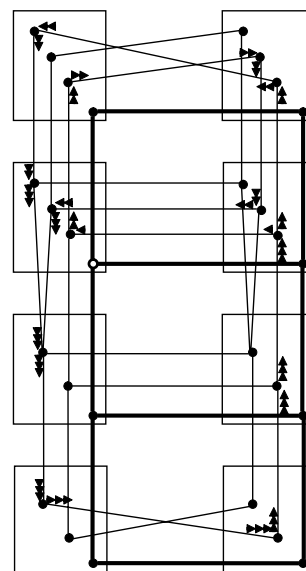


Рис. 3

Множество T представлено здесь восемью квадратами, как и в предыдущем примере, с тем же соседством \mathfrak{Z} . В каждом квадрате t изображены кружочки, представляющие различные пары (t, k) . Пусть t^* – второй сверху квадрат в левом столбце, k^* – метка, обозначенная в виде белого кружочка. Пусть $q_{t^*}(k^*) = -1$, а величины $q_t(k)$ равны нулю для всех прочих пар (t, k) . Величины $g_{t'}(k, k')$ равны нулю для всех тех четверок t, t', k, k' , которые обозначены на рисунке отрезками, соединяющими метку k в квадрате t с меткой k' в квадрате t' . Числа $g_{t'}(k, k')$ для всех прочих четверок t, t', k, k' равны некоторому большому по абсолютной величине отрицательному числу, скажем, -100 .

Сформулированная таким образом (\max , $+$)-задача не является тривиальной, но эквивалентна таковой. Перевод ее в тривиальную осуществляет набор потенциалов, показанный на рис. 3. Этот набор обеспечивает энергию задачи, равную (-1) и качеству оптимальной разметки, показанной на рис. 3 утолщенными линиями. Такой или какой-нибудь другой перевод исходной задачи в тривиальную обязательно будет реализован алгоритмом, который минимизирует энергию задачи. В то же время

ни один из известных до настоящего времени алгоритмов такую тривиализацию задачи не выполнит. Действительно, исходная (max,+)-задача такова, что соответствующая ей ($\vee, \&$)-задача содержит непустую согласованную часть. Известные к настоящему времени алгоритмы выходят на останов при выполнении этого условия и, таким образом, могут не найти тривиальный эквивалент решений (max,+)-задачи, даже если такой существует. Для алгоритмов же, минимизирующих энергию задачи, такая ситуация невозможна.

Для тривиализации представленной задачи использовался Алгоритм 1, время работы которого, необходимое для поиска оптимальной разметки, существенно зависело от выбора шагов γ_i . В зависимости от того, как выбраны эти шаги, время решения задачи менялось от нескольких сотен итераций до очень большого их количества. Вполне приемлемое время достигалось при $\gamma_i = 100/i$ и даже при постоянных значениях γ_i , равных, например, 0,2. Таким образом, рекомендация $\lim_{i \rightarrow \infty} \gamma_i = 0$,

$\lim_{n \rightarrow \infty} \sum_{i=0}^n \gamma_i = \infty$, вытекающая из общей теории обобщенного градиентного спуска, является слишком слабой. В рамках этой общей рекомендации существуют как вполне приемлемые решения, так и практически неприемлемые. Как было отмечено, вопрос об оптимальном выборе шагов γ_i в контексте наших задач остается открытым, и ответ на него потребует дальнейших усилий.

Примеры решения (max,+)-задач, эквивалентных тривиальным

Областью применения разработанных алгоритмов является подкласс (max,+)-задач, эквивалентных тривиальным. Повторим, что этот подкласс достаточно обширный, так как включает в себя все ациклические и все супермодулярные (max,+)-задачи.

Разработанные алгоритмы тестировались прежде всего именно на тех (max,+)-задачах,

которые заведомо входят в область их применения. Эксперименты проводились не с целью проверки, преобразуется ли та или иная задача в тривиальную, так как этот факт доказан теоретически. Целью экспериментов была оценка быстродействия алгоритмов и сравнения Алгоритма 1 и Алгоритма 2, т.е. ответы на вопросы, которые теоретически не исследовались. Кроме того, экспериментально проверялось принятое условие останова алгоритмов, которое также не обосновано теоретически.

Результаты экспериментов представлены в таблице.

Сравнение эффективности алгоритмов

Размер задачи			Алгоритм 1		Алгоритм 2	
Строк	Столбцов	Меток	Количество итераций	Время (с)	Количество итераций	Время (с)
5	5	2	194,5	1	81,7	$\ll 1$
5	5	5	415,3	6	180,0	$\ll 1$
5	5	10	506,1	26	192,4	$\ll 1$
10	10	2	251,3	13	155,8	$\ll 1$
10	10	5	421,7	59	198,6	$\ll 1$
10	10	10	534,0	199	201,6	$\ll 1$
50	50	2	1243,0	876	160,0	1
50	50	5	–	–	202,3	2
50	50	10	–	–	204,5	7
100	100	10	–	–	205,2	33

Опишем их на примере пятой строки в таблице.

Множество T объектов – это множество пикселей изображения, состоящего из 10 строк и 10 столбцов. Множество \mathfrak{Z} соседних пар объектов определено так, как это указано в данной части работы. Множество K меток состоит из пяти меток. В каждом выполненном эксперименте заранее определялась некоторая разметка \bar{k}^* , а затем – числа $q_t(k)$ и $g_{t'}(k, k')$ так, чтобы именно выбранная разметка \bar{k}^* оказалась оптимальной. Эти числа выбирались так, что

$$\begin{aligned}
 q_t(k) &= 0 \text{ для всех } t \in T \text{ и } k \in K, \\
 g_{t'}(k, k') &= 0, \text{ если } k = k^*(t), k' = k^*(t'), \\
 g_{t'}(k, k') &= -1 \text{ для всех прочих пар } k \text{ и } k'.
 \end{aligned}$$

Очевидно, что определенная таким способом задача тривиальна. Эта задача преобразо-

валась в эквивалентную ей с помощью случайных значений потенциалов $\varphi_{u'}(k)$, имеющих равномерное распределение в интервале от -50 до $+50$. Преобразованная задача, таким образом, определялась числами

$$g'_{u'}(k, k') = g_{u'}(k, k') + \varphi_{u'}(k) + \varphi_{i_t}(k'),$$

$$q'_i(k) = - \sum_{i' \in N(i^*)} \varphi_{u'}(k),$$

и, как правило, оказывалась нетривиальной. Эта задача подавалась на входы Алгоритма 1 и Алгоритма 2, каждый из которых выполнял ее тривиализацию после некоторого количества итераций, т.е. выходил на останов. Этот эксперимент для данных размеров изображения и количества меток выполнялся 20 раз при различных случайных значениях потенциалов $\varphi_{u'}(k)$. Среднее время, затрачиваемое Алгоритмом 1 на тривиализацию каждой из 20 представленных задач, равнялось 59 с, а среднее количество итераций равнялось 421,7. Время тривиализации этих же задач Алгоритмом 2 было достаточно малым, так что не поддавалось непосредственному измерению. Это отображено в соответствующей клетке таблицы знаком $\ll 1$. Среднее количество итераций равнялось 198,6.

Эксперименты выполнялись для различных размеров поля зрения T от 5×5 до 100×100 и для количества меток 2, 5 и 10. По первым семи строкам табл. 1 видно, что Алгоритм 2 имеет заметное преимущество перед Алгоритмом 1. Поэтому для последних трех строчек таблицы Алгоритм 1 не проверялся.

Распознавание одного класса изображений

Сформулируем определенное множество изображений, которые будем называть *идеальными*. Пусть T – прямоугольное поле зрения, состоящее из m строк и n столбцов, $T = \{(i, j) | 1 \leq i \leq m, 1 \leq j \leq n\}$. Определим $m + n$ изображений вида $T \rightarrow \{0, 1\}$, которые обозначим $e_{j^*}^B, j^* = 1, 2, \dots, n$, и $e_{i^*}^\Gamma, i^* = 1, 2, \dots, m$, так, что

$$e_{j^*}^B(i, j) = 1, \text{ если } j = j^*, i = 1, 2, \dots, m,$$

$$e_{j^*}^B(i, j) = 0, \text{ если } j \neq j^*, i = 1, 2, \dots, m,$$

$$e_{i^*}^\Gamma(i, j) = 1, \text{ если } i = i^*, j = 1, 2, \dots, n,$$

$$e_{i^*}^\Gamma(i, j) = 0, \text{ если } i \neq i^*, j = 1, 2, \dots, n.$$

Иными словами, $e_{j^*}^B$ – изображение вертикальной линии, в котором все пиксеты j^* -го столбца черные, а все прочие пиксеты – белые; $e_{i^*}^\Gamma$ – изображение горизонтальной линии, в котором все пиксеты i^* -й строки черные, а все прочие – белые. Определим также полностью белое изображение e_0 , такое, что $e_0(i, j) = 0$ для всех пикселов. Определим дизъюнкцию (наложение) изображений $e_1 : T \rightarrow \{0, 1\}$ и $e_2 : T \rightarrow \{0, 1\}$ как изображение $e : T \rightarrow \{0, 1\}$, такое, что

$$e(i, j) = e_1(i, j) \vee e_2(i, j).$$

Определим множество E изображений, как множество, замкнутое по дизъюнкции и содержащее белое изображение e_0 , изображения $e_{j^*}^B, j^* = 1, 2, \dots, n$, и изображения $e_{i^*}^\Gamma, i^* = 1, 2, \dots, m$. Как видно, это множество состоит из 2^{n+m} изображений, которые назовем *идеальными*.

Рассмотрим ситуацию, когда изображение $e \in E$ непосредственно не наблюдается, а наблюдается лишь результат его случайного искажения в соответствии со следующей простейшей моделью. Изображение $e : T \rightarrow \{0, 1\}$ преобразуется в изображение $x : T \rightarrow \{0, 1\}$ так, что в каждом пикселе $t \in T$ величина $e(t)$ заменяется на величину $x(t) = 1 - e(t)$ с вероятностью α и остается неизменной, $x(t) = e(t)$, с вероятностью $1 - \alpha$.

В такой ситуации естественна формулировка задачи распознавания, как отыскание наиболее правдоподобного идеального изображения $e \in E$ по его искаженному наблю-

дению x . При $\alpha < 0,5$ эта задача эквивалентна задаче поиска идеального изображения

$$e^* = \arg \min_{e \in E} \sum_{t \in T} |e(t) - x(t)|. \quad (9)$$

Неформально говоря, задача состоит в обнаружении совокупности вертикальных линий на фоне той или иной, заранее не заданной, совокупности горизонтальных линий, причем в условиях случайных искажений. При таком неформальном, поверхностном взгляде на задачу она может показаться недопустимой симплификацией реальных задач распознавания изображений, в которых нужно отыскивать значительно более сложные объекты, чем строго вертикальные линии, причем на значительно более сложном фоне, чем строго горизонтальные линии. Не отрицая такого мнения, следует учитывать, что кажущаяся простота сформулированной задачи в высшей степени обманчива хотя бы потому, что множество всех возможных задач вида (9) образует NP -полный класс.

Задача (9) достаточно естественно формулируется как $(\max,+)$ -задача разметки. Множество T есть множество пикселей, а соседство \mathfrak{Z} есть объединение двух соседств: \mathfrak{Z}^{Γ} по горизонтали и \mathfrak{Z}^{B} по вертикали. Множество K состоит из четырех меток: $\text{B}\Gamma$, $\bar{\text{B}}\Gamma$, $\text{B}\bar{\Gamma}$ и $\bar{\text{B}}\bar{\Gamma}$, имеющих следующий смысл. Если $k(t) = \text{B}\Gamma$, это значит, что через пиксел t проходит как вертикальная, так и горизонтальная линии, $k(t) = \bar{\text{B}}\Gamma$ обозначает, что через пиксел проходит горизонтальная линия, но не проходит вертикальная. Подобный смысл имеют метки $\text{B}\bar{\Gamma}$ и $\bar{\text{B}}\bar{\Gamma}$.

Для пикселей $tt' \in \mathfrak{Z}^{\Gamma}$, соседних по горизонтали, функцию $g_{tt'}(k, k')$ выбирают так, что

$$\begin{aligned} g_{tt'}(\text{B}\Gamma, \text{B}\Gamma) &= g_{tt'}(\bar{\text{B}}\Gamma, \text{B}\Gamma) = \\ &= g_{tt'}(\text{B}\Gamma, \bar{\text{B}}\Gamma) = g_{tt'}(\bar{\text{B}}\Gamma, \bar{\text{B}}\Gamma) = \\ &= g_{tt'}(\text{B}\bar{\Gamma}, \text{B}\bar{\Gamma}) = g_{tt'}(\bar{\text{B}}\bar{\Gamma}, \bar{\text{B}}\bar{\Gamma}) = \\ &= g_{tt'}(\text{B}\bar{\Gamma}, \bar{\text{B}}\bar{\Gamma}) = g_{tt'}(\bar{\text{B}}\bar{\Gamma}, \text{B}\bar{\Gamma}) = 0, \end{aligned}$$

и $g_{tt'}(k, k') = -\infty$ для всех прочих пар меток. Такой выбор чисел $g_{tt'}(k, k')$ приводит к тому, что если для какого-то пикселя принято решение, что через него проходит (или не проходит) горизонтальная линия, то такое же решение должно быть принято для всех пикселей в этой же строке. Подобным образом определены числа $g_{tt'}(k, k')$ для всех пар вертикально соседствующих пикселей.

Числа $q_t(k)$ выбираются так, что

$$\begin{aligned} q_t(\bar{\text{B}}\bar{\Gamma}) &= 0 \text{ при } x(t) = 0, \\ q_t(k) &= -1 \text{ при } k \neq \bar{\text{B}}\bar{\Gamma}, x(t) = 0, \\ q_t(\bar{\text{B}}\bar{\Gamma}) &= -1 \text{ при } x(t) = 1, \\ q_t(k) &= 0 \text{ при } k \neq \bar{\text{B}}\bar{\Gamma}, x(t) = 1. \end{aligned}$$

Эксперименты по решению $(\max,+)$ -задач, сформулированных указанным способом, проводились с целью проверки их быстродействия, как и в предыдущей серии экспериментов. Однако в данной серии, в отличие от предыдущей, речь идет о решении $(\max,+)$ -задач, которые совсем не обязательно эквивалентны тривиальным. Поэтому дополнительная цель этой серии экспериментов заключалась в том, чтобы определить, насколько часто в этой серии случайных $(\max,+)$ -задач возникают задачи, не эквивалентные тривиальным. Результат этой проверки таков, что эта неблагоприятная ситуация не возникла ни разу.

Эксперименты по решению $(\max,+)$ -задач, сформулированных приведенным способом, выполнялись для изображений размером 10×10 , 50×50 , 100×100 и 500×500 при вероятности $\alpha = 0,1$ искажения сигнала в пикселе. Задачи, связанные с распознаванием изображений размером 10×10 , решались как Алгоритмом 1, так и Алгоритмом 2. Среднее время решения задач Алгоритмом 1 равнялось приблизительно 40 с, а среднее количество итераций – около 140. Время решения задач Алгоритмом 2 было настолько малым, что не поддавалось измерению. Среднее количество итераций равнялось приблизительно 10.

Эксперименты по распознаванию изображений размером 50×50 , 100×100 и 500×500 выполнялись только Алгоритмом 2. Далее приведены результаты работы Алгоритма 2 на трех примерах, по одному для каждого размера изображений.

На рис. 4,а, 5,а и 6,а показаны идеальные изображения совокупности горизонтальных и вертикальных линий, а на рис. 4,б, 5,б и 6,б – результаты их искажения. Эти искаженные изображения подавались на вход программы распознавания, которая на их основе формировала соответствующую (max,+)-задачу, решаемую Алгоритмом 2. Результаты распознавания на рисунках не показаны, так как все они точно совпадают с идеальными изображениями.

Время работы Алгоритма 2 при распознавании изображения размером 50×50 (см. рис. 4,б) составляет 1 с, количество итераций – 17.

Время распознавания изображения размером 100×100 (см. рис. 5,б) составляет 3 с, количество итераций – 16.

Время распознавания изображения размером 500×500 (см. рис. 6,б) составляет 124 с, количество итераций – 36.

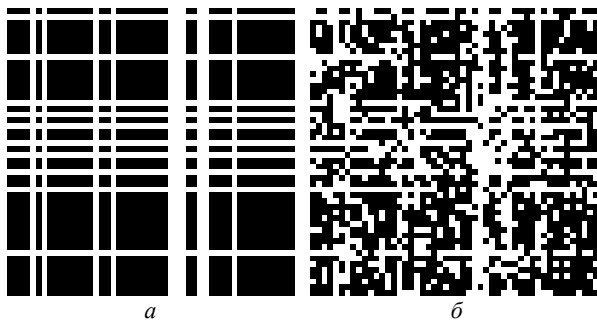


Рис. 4

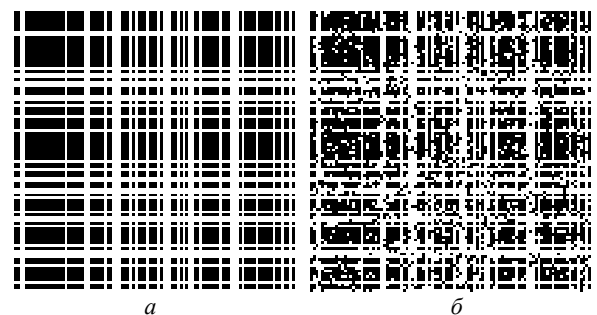


Рис. 5

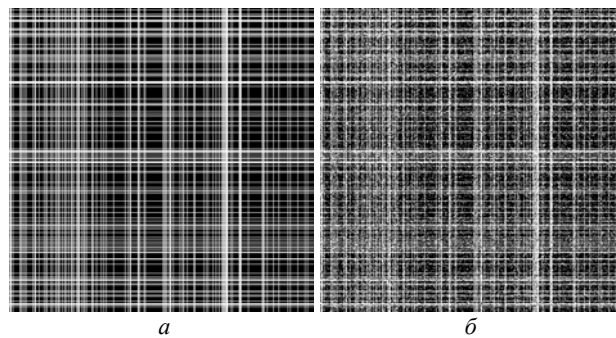


Рис. 6

Сегментация изображений

Рассмотрим два примера распространенной обработки изображений, известных как текстурная сегментация. Цель и сущность этой обработки вытекают из следующей порождающей модели [5]. Пусть T – поле зрения с обычным для изображений соседством \mathfrak{Z} . Пусть $\bar{x} : T \rightarrow \{0,1\}$ – бинарное изображение на этом поле зрения, а $X^T = \{0,1\}^T$ – множество всех возможных изображений, имеющих областью определения множество T . Пусть на множестве X^T заданы n Гиббсовских распределений вероятностей $p^i : X^T \rightarrow R$, $i = 1, 2, \dots, n$, так что

$$\log p^i(\bar{x}) = \sum_{t,t' \in \mathfrak{Z}} g_{tt'}^i(x(t), x(t')).$$

Пусть $\bar{k} : T \rightarrow \{1, 2, \dots, n\}$ – разметка, называемая сегментацией поля зрения T , такая, что $k(t) = i$ обозначает, что пиксел t принадлежит сегменту i . На множестве $\{1, 2, \dots, n\}^T$ всех возможных сегментаций задано Гиббсово распределение вероятностей $p^\circ : \{1, 2, \dots, n\}^T \rightarrow R$, так, что логарифм вероятности конкретной сегментации \bar{k} есть число

$$\log p^\circ(\bar{k}) = \sum_{t,t' \in \mathfrak{Z}} g_{tt'}^\circ(k(t), k(t')).$$

Числа $g_{tt'}^\circ(k, k')$ выбраны в соответствии с общепринятой диагональной моделью [3]. А именно, для всех $tt' \in \mathfrak{Z}$

$$g_{u'}^{\circ}(k, k') = 1 \text{ при } k = k',$$

$$g_{u'}^{\circ}(k, k') = 0 \text{ при } k \neq k',$$

что отражает тот интуитивно предполагаемый факт, что принадлежность двух соседних пикселей одному сегменту более вероятна, чем принадлежность разным.

Пусть $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n$ – n различных реализаций случайных изображений, причем каждое изображение \bar{x}_i получено в соответствии со своим распределением вероятностей p^i . Пусть \bar{k} – реализация случайной сегментации, полученная в соответствии с распределением вероятностей p° . Из этих n изображений и сегментации составляется изображение $\bar{x}: T \rightarrow \{0,1\}$ так, что $x(t) = x_{k(t)}(t)$. При этом говорят, что изображение \bar{x} составлено из изображений \bar{x}_i , $i = 1, 2, \dots, n$, с помощью сегментации \bar{k} . Задача сегментации изображений возникает тогда, когда процесс порождения составного изображения \bar{x} скрыт от наблюдателя, а наблюдаемым является только результат этого процесса, т.е. собственно составное изображение \bar{x} . Заключается задача в том, чтобы при известном составном изображении \bar{x} принять решение о сегментации \bar{k}^* , максимизирующей качество

$$G(\bar{k}) = \sum_{t' \in \mathcal{E}} g_{u'}^{\circ}(k(t), k(t')) + \sum_{t \in T} q_t^{\circ}(k(t), \bar{x}). \quad (10)$$

Числа $q_t^{\circ}(k(t), \bar{x})$ определялись на основании аргументов, описанных в работе [4]. В контексте данной работы эта аргументация не играет роли и здесь не приводится. Существенно то, что при $|K| = 2$, независимо от чисел $q_t(k, \bar{x})$, задача (10) супермодулярна и, следовательно, эквивалентна некоторой тривиальной задаче. Однако при количестве меток, равном трем и более, задача (10) существенно усложняется. Как известно [6], уже при $|K| = 3$ и даже при фиксированной диагональной функции $g_{u'}^{\circ}(k, k')$ множество задач

вида (10), соответствующих различным числам $q_t(k)$, образует NP -полный класс. Поэтому при $|K| = 3$ для задач вида (10) может и не существовать тривиального эквивалента. Однако в проведенных экспериментах такая ситуация не произошла.

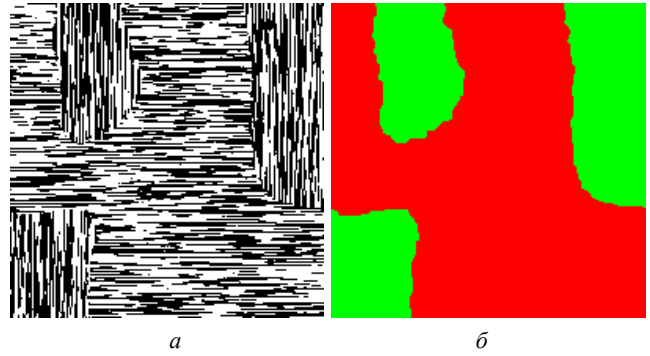


Рис. 7

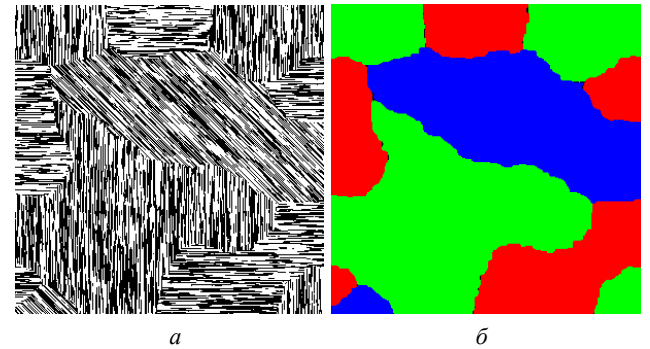


Рис. 8

На рис. 7 и 8 приведены примеры сегментации изображений при $|K| = 2$ и $|K| = 3$ соответственно. На рис. 7,а и 8,а представлены исходные изображения, а на рис. 7,б и 8,б – их сегментация, полученная как разметка, оптимизирующая качество (10). Соответствующие (max,+)-задачи решались Алгоритмом 2. Размер изображения на рис. 7,а равен 200×200 пикселей, время решения соответствующей (max,+)-задачи равно 2 с, количество итераций – 20. Изображение на рис. 8,а имеет размеры 300×300 пикселей, время его обработки – 25 с, 112 итераций.

Окончание на стр. 32

1. Шлезингер М.И., Гигиняк В.В. Решение (max,+)-задач структурного распознавания с помощью их эквивалентных преобразований. I // УСиМ. – 2007. – № 1. – С. 3–15.
2. Rosenfeld F., Hummel R., Zucker S. Scene Labeling by Relaxation Operations // IEEE Transactions on Syst. Man and Cyber. – 1976. – SMC-6. – № 6. – P. 420–433.
3. Potts R. Some generalized order-disorder transformation // Proc. of the Cambridge Philosophical Society. – 1952. – 48. – P. 106–109.
4. Ковтун И.В. Технология текстурной сегментации изображений на основании Марковских случайных полей и решения (max,+)-задач // УСиМ. – 2004. – № 2. – С. 61–66.
5. Ковтун И.В. Текстурная сегментация изображений на основании марковских случайных полей // Там же. – 2003. – № 4. – С. 46–55.
6. Boykov Y., Veksler O., Zabih R. Fast approximate energy minimization via graph cuts // IEEE Trans. Pattern Analysis and Machine Intell. – 2001. – 23, № 11. – P. 1222–1239.
7. Шлезингер М.И. Синтаксический анализ двумерных зрительных сигналов в условиях помех // Кибернетика. – 1976. – № 4. – С.113–130.
8. Коваль В.К., Шлезингер М.И. Двумерное программирование в задачах анализа изображений // Автоматика и телемеханика. – 1976. – № 8. – С. 149–168.
9. Wainwright M., Jaakkola T., Willsky A. MAP estimation via agreement on (hyper) trees: message passing and linear programming approaches // Allerton Conf. on Communication, Control and Computing, 2002.
10. Kolmogorov V. Convergent tree-reweighted message passing for energy minimization // Microsoft Research. Tech. Rep. MSR-TR-1005-38, 2005.
11. Kolmogorov V. Convergent tree-reweighted message passing for energy minimization // IEEE Trans. Pattern Analysis and Machine Intell. – 2006. – 28, № 10. – P. 1568–1583.
12. Werner T. A Linear Programming Approach to Maxsum Problem: A review. <ftp://cmp.felk.cvut.cz/pub/cmp/articles/werner/Werner-TR-2005-25.pdf>.

Поступила 26.01.2007

Тел. для справок: (044) 526-6208 (Киев)

© М.И. Шлезингер, В.В. Гигиняк, 2007

