

Стереореко́нструкція та візуалізація тривимірних сцен, що складаються з декількох шарів дальності

Д. Рябокони

Міжнародний науково-навчальний центр
інформаційних технологій та систем
Київ, Україна
dima@image.kiev.ua

Анотація (Abstract)

An algorithm for spatial reconstruction of 3D scene that consist of several flat objects is considered. The algorithm is based on method of most probabal estimation of system parameters in unsuperwised mode and estimation of hidden markov field parameters as well. Also the technology for visualization of reconstructed 3D models by lenticular images is proposed.

Вступ

Новітні дослідження в галузі машинного стереозору обумовили появу низки так званих “двовимірних” алгоритмів. На відміну від “одновимірних” алгоритмів [1, 2], вони будують модель тривимірної сцени не у вигляді шматочків, отриманих в результаті незалежної обробки рядків стереопари зображень, а у вигляді цілісної тривимірної моделі, що враховує зв’язок між положенням як горизонтальних, так і вертикальних сусідів кожної точки простору. Переваги “двовимірних” алгоритмів [3-5] очевидні. За їх допомогою можна отримати більш досконалі та точні результати стереореко́нструкції, підвищити надійність результатів розпізнавання завдяки врахуванню конфігурацій більшої кількості сусідніх ділянок кожної точки об’єкта.

“Двовимірні” алгоритми існують зараз для досить широкого класу стереознімків. Зокрема, для стереознімків, що містять зображення неперервних поверхонь, як от земна поверхня чи людське обличчя. На таких стереознімках мають міститися лише точки, що спостерігаються обома камерами одночасно. Якщо деяка точка видима на лівому зображенні стереопари, то вона має бути видима і на правому зображенні.

Однак, все ще залишається багато стереозображень, для яких ефективні “двовимірні” алгоритми не досліджені належним чином. Наприклад, для стереозображень сцен, де містяться декілька об’єктів, кожен з яких є неперервною поверхнею. Найвідоміші алгоритми для обробки таких зображень залишаються “одновимірними” [2]. Вдосконалення існуючих алгоритмів обробки стереозображень вказаного типу дозволить не лише розширити теоретичні знання в галузі машинного стереозору, а й вирішити певні практичні задачі. Прикладом практичного застосування представленого в роботі алгоритму може бути технологія створення платівок із стереоефектом.

1. Прийняті позначення

Існують дві камери (ліва та права), *поле зору* яких є множина $T = \{t \equiv (i, j) / 0 \leq i < I, 0 \leq j < J\}$.

Горизонтальний розмір поля зору I , а вертикальний – J . Елементи поля зору камер називатимемо *пікселами* та позначати парю змінних (i, j) , або ж однією змінною t .

Множину *сусідів* пікселя $t \equiv (i, j)$ позначимо $\tau(t) = \{(i-1, j), (i+1, j), (i, j-1), (i, j+1)\}$.

Зображенням $x : T \rightarrow [0..255]$ будемо називати функцію, що ставить у відповідність кожному пікселю міру його яскравості. На знімках, що поступають з камер, міститься зображення деякої тривимірної сцени. Значення сигналів в елементах поля зору камер позначимо $x_t \equiv x(i, j)$ та $x'_t \equiv x'(i, j)$ відповідно.

Паралаксом пікселя (i, j) лівого зображення будемо називати різницю $k \equiv (i-i', j-j')$, де i' та j' – це горизонтальні координати відповідного йому пікселя на правому зображенні. Вважатимемо, що можливі значення паралакса відомі та визначені множиною K .

Розміткою будемо називати функцію $\bar{k} : T \rightarrow K$, що ставить у відповідність кожному пікселю лівого зображення його паралакс. Значення паралаксу в пікселі $t \equiv (i, j)$ лівого зображення позначатимемо $k_t \equiv k(i, j)$. Вважатимемо, що розмітка \bar{k} являється випадковою реалізацією марківського поля, що задається *потенціалами Гіббса* $g : T \times T \times K \times K \rightarrow R$.

Значення функції g позначимо $g_{t,t'}(k, k')$, де t та $t' \in \tau(t)$ – це сусідні піксели лівого зображення, що мають паралакси k та k' відповідно. Ймовірність $p(\bar{k})$ розмітки \bar{k} обчислюється за формулою
$$p(\bar{k}) = \prod_{t \in T, t' \in \tau(t)} g_{t,t'}(k(t), k(t')).$$

Вважатимемо, що розподіл яскравості сигналу $x_t, t \in T$ не залежить від пікселя t , і умовна ймовірність $p(x|\bar{k})$ зображення x за умови відомої розмітки \bar{k} є добуток
$$\prod_{t \in T} p(x_t | k_t).$$

2. Постановка задачі

Задана стереопара (x, x') . Кожна точка, що зображена на ній, належить до одного з K шарів (рівнів) дальності. Розміткою лівого зображення буде функція $\bar{k} : T \rightarrow K$, що ставить у відповідність кожному пікселю номер його рівня дальності. Вважається, що розмітка є реалізацією марківського випадкового поля. Її апіорна ймовірність задається за допомогою потенціалів Гіббса $g_{it'}(k, k')$ та визначається добутком $p(\bar{k}) = \prod_{t \in T, t' \in \tau(t)} g_{it'}(k_t, k_{t'})$, де

$$\tau(t) = \{(i-1, j), (i+1, j), (i, j-1), (i, j+1)\} \in$$

множиною сусідів пікселя t . Кожен із $k \in K$ рівнів дальності характеризується своїм параметром, а саме паралаксом d_k , що попередньо невідомий. Умовна ймовірність $p(x_t | k_t)$ яскравості x_t на лівому зображенні за умови, що піксель t належить рівню k_t , є функція $f(x_t, x', d_{k_t})$, що залежить від паралаксу d_{k_t} .

Необхідно для кожного пікселя $t \in T$ лівого зображення знайти його рівень дальності k_t^* , а також для кожного рівня $k \in K$ вказати його паралакс d_k^* . При цьому ймовірність $p(x, d^*)$ спостереження зображення x (це, до речі, є функція правдоподібності вибірки x) та умовна ймовірність $p(\bar{k}^* | x, d^*)$ розмітки \bar{k}^* за умови заданого зображення x мають бути максимальні:

$$d^* = \arg \max_d p(x, d) = \arg \max_d \sum_{k \in K^T} p(\bar{k}) \cdot p(x, d | \bar{k}) \quad (1)$$

$$\bar{k}^* = \arg \max_{k \in K^T} p(\bar{k} | x, d^*) = \arg \max_{k \in K^T} p(\bar{k}, x, d^*) \quad (2)$$

3. Алгоритм розв'язку задачі

Алгоритм розв'язку задачі (1) базується на методі найімовірнішої оцінки параметрів системи в режимі самонавчання та оцінці на цій основі параметрів прихованого марківського поля [6, 7].

Спочатку, за допомогою алгоритму самонавчання будуть оцінені значення параметрів $d_k^*, k \in K$. Далі, на основі параметрів d_k^* за допомогою процедури "генератор Гіббса" (Gibbs sampler) буде згенерована розмітка \bar{k}^* , що є розв'язком задачі (2).

Алгоритм обчислення параметрів $d_k^*, k \in K$ працює ітераційно. Кожна ітерація містить два кроки. Номер ітерації позначимо символом r , а значення параметрів на цій ітерації $d^{(r)}$.

Крок 0. Вибрати параметри $d^{(0)}$ довільним випадковим чином.

Крок 1. Для кожного пікселя $t \in T$ та рівня дальності $k \in K$ обчислити величину $\alpha_t(k)^{(r)}$, що має

сенса наближеного значення вірогідності $p(k | x_t)$ реалізації в пікселі t рівня k за умови відомої яскравості x_t .

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

Значення $p(k_t = k)$ є апіорною ймовірністю стану k в пікселі t . Її обчислення для заданого за допомогою потенціалів Гіббса марківського поля є NP-повною задачею. Для наближеного обчислення значень $p(k_t = k), k \in K$ використовують генератор Гіббса (Gibbs sampler).

Крок 2. Для кожного $k \in K$ обчислити значення параметру $d_k^{(r+1)}$.

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

Крок 3. Якщо величина $\|d^{(r+1)} - d^{(r)}\|$ більше, ніж певне наперед визначене значення, то перейти до кроку 1, інакше, завершити алгоритм.

Задача (2) в загальному випадку є NP-повною. Лише в деяких випадках, наприклад, коли значення функції g задовольняють так звану умову монотонності $g_{it'}(k_1, k_2') \cdot g_{it'}(k_2, k_1') \leq g_{it'}(k_1, k_1') \cdot g_{it'}(k_2, k_2')$ для будь-яких $t, t' \in \tau(t)$ і $k_1 < k_2, k_1' < k_2'$ задача (2) розв'язна [3]. Для наближеного розв'язку чи пошуку частини частини точного розв'язку задачі (2) розроблено багато алгоритмів [8]. Нижче наведено один із них, відомий як "генератор Гіббса" (Gibbs sampler).

Алгоритм працює циклічно та генерує послідовність розміток $\bar{k}^{(0)}, \bar{k}^{(1)}, \dots, \bar{k}^{(r)}, \dots$ і т.д., що збігається до розмітки \bar{k}^* , що є розв'язком задачі (2).

Крок 0. Довільним чином згенерувати випадкову розмітку $\bar{k}^{(0)}$, ймовірність якої ненульова.

Крок 1. На основі розмітки $\bar{k}^{(r)}$, згенерувати розмітку $\bar{k}^{(r+1)}$ за наступним правилом: для кожного пікселя $t \in T$ обчислити умовні вірогідності $p(k | x_t)$ реалізації його можливих станів $k \in K$ та згенерувати стан $k_t^{(r+1)}$ як випадкову величину з розподілом ймовірності $p(k | x_t)$. Умовна ймовірність $p(k | x_t)$, $k \in K$ може бути розрахована за наближеною формулою:

$$p(k | x_t) = \frac{p(x_t, k_t = k)}{p(x_t)} = \frac{p(k_t = k) \cdot p(x_t | k_t = k)}{p(x_t)}$$

$$\begin{aligned}
&= \frac{p(k_t = k) \cdot f(x_t, x', d_{k_t^{[r]}})}{p(x_t)} \approx \\
&\approx \frac{\prod_{t' \in \tau(t)} g_{tt'}(k, k_t^{[r]}) \cdot f(x_t, x', d_{k_t^{[r]}})}{p(x_t)}.
\end{aligned}$$

Слід відзначити, що ймовірність $p(x_t)$ в цій формулі обчислювати не слід, адже вона входить як постійний множник для всіх значень ймовірності $p(k|x_t), k \in K$.

Крок 2. Якщо величина $\|k^{[r+1]} - k^{[r]}\|$ більше, ніж певне наперед визначене значення, то перейти до кроку 1, інакше завершити алгоритм.

4. Приклад

На рис. 1 міститься штучне стереозображення тривимірної сцени. Результат її обробки з допомогою різних алгоритмів представлено на рис. 2. Перша тривимірна модель є результатом роботи "одновимірного" алгоритму [1]. Друга модель є результатом роботи "двовимірного" алгоритму, що відновлює просторову конфігурацію сцени у вигляді неперервної поверхні. Третя модель є результатом роботи представленого в даній роботі алгоритму. Четверта модель є дійсною конфігурацією сцени, на основі якої і створено штучні стереозображення.



Рис. 1: Стереопара зображень.

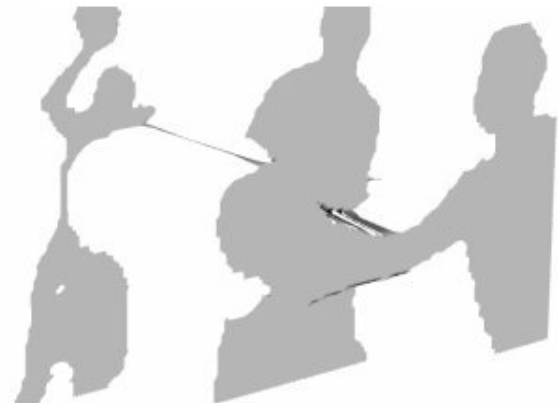
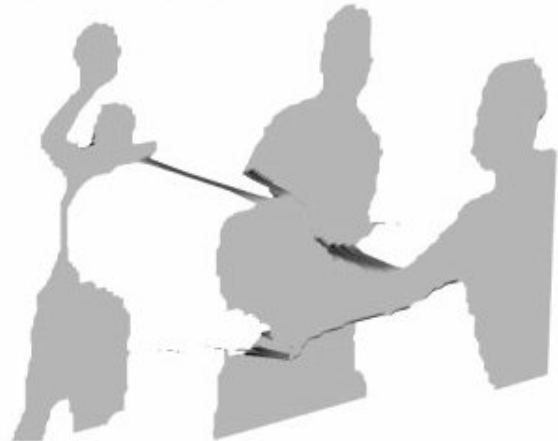


Рис. 2: Результати стереорекострукції.

5. Візуалізація тривимірних моделей за допомогою лентикулярних листівок

Сучасні апаратні та програмні засоби персональних комп'ютерів дозволяють доволі реалістично відтворювати складні тривимірні моделі. На екрані монітору з легкістю можна змінювати масштаб тривимірних моделей, їх забарвлення, освітлення, ракурс тощо. В той же час існують інші можливості для візуалізації тривимірних моделей, наприклад за допомогою відомих більше століття лентикулярних платівок [9]. Їх перевага в тому, що на них можна дивитися, як на звичайні фотографії та бачити при цьому стереоефект. Новизна запропонованого підходу полягає в тому, що для створення листівки використовується зображення, що надходять лише з 2-х камер, а не з 6-ти чи 8-ми, як це передбачено. Приклад створення лентикулярного зображення наведено на рис 3.



Рис. 3: Приклад лентикулярного зображення.

Листівка створюється за серією із $N = 6..8$ зображень $x^{[0]}, x^{[1]}, \dots, x^{[N-1]}$, що мають однаковий розмір та отримані в результаті зйомки деякої тривимірної сцени з різних точок зору. В рядку кожного зображення має бути стільки пікселів, скільки ребер містить платівка. На основі серії знімків створюється одне зображення (називатимемо його лентикулярним), зверху якого і накладається ребриста платівка. Під ребром з номером i має бути зображена одна смуга, що складається з N стовпчиків, кожен з яких міститься у відповідному зображенні в стовпчику i . Наприклад, під першим ребром платівки має міститися смуга, що складається із стовпчика 1 першого знімку, стовпчика 1 другого знімку, і т.д. до стовпчика 1 знімку N . Важливо зауважити, що кількість знімків має бути не менше 6-ти. Це мінімальна кількість знімків, за якої досягається гарний стереоефект.

Запропонований підхід для створення лентикулярних листівок полягає в тому, що він використовує лише два знімки (x, x') , які утворюють стереопару. Проміжні знімки, що необхідні для створення лентикулярного зображення, можуть бути відтворені за допомогою тривимірної моделі, що створюється за стереопарою зображень.

Нехай кожному пікселю зображення x поставлено у відповідність певний піксель на зображенні x' . Якщо

(i, j) – це піксель лівого зображення, а $k(i, j)$ – його паралакс, то пікселю (i, j) ні лівому зображенні відповідає піксель $(i, j) + k(i, j)$ на правому зображенні. І навпаки, якщо (i, j) – це піксель правого зображення, а $k'(i, j)$ – його паралакс, то пікселю (i, j) на правому зображенні відповідає піксель $(i, j) + k'(i, j)$ на лівому зображенні. Проміжні зображення $x^{[0]}, x^{[1]}, \dots, x^{[N-1]}$ генеруються за допомогою наступного алгоритму.

for $(i, j) \in T, r = 0, \dots, N-1$

{

$$(i', j') := (i, j) + k(i, j) \cdot \frac{r}{N-1}; \quad x^{[r]}(i', j') = \frac{r}{N-1} \cdot x(i, j);$$

$$(i', j') := (i, j) + k'(i, j) \cdot \frac{N-1-r}{N-1}; \quad x^{[r]}(i', j') = \frac{N-1-r}{N-1} \cdot x'(i, j);$$

}

Висновки

Запропоновано алгоритм обробки класу стереозображень навколишнього світу, на яких містяться декілька плоских об'єктів. Перевагою алгоритму є цілісний підхід до обробки стереозображень, за якого модель тривимірної сцени відновлюється в результаті обробки всієї стереопари, а не обробки окремих її рядків. Технологія створення лентикулярних листівок, що пропонується в даній роботі, може представляти інтерес для сучасних галузей поліграфії та реклами.

Література

- [1] Гимельфарб Г.Л. "Симметрический подход к задаче автоматических стереоскопических измерений в фотометрии", *Кибернетика*. (2), 73-82, 1979.
- [2] Sara R. "Finding the Largest Unambiguous Component of Stereo Matching". *The 7-th European Conf. on Comp. Vision*, vol. 2. – 2002. – P. 900-914.
- [3] Schlesinger M.I., Flach B. "Analysis of optimal labeling problems and their application to image segmentation and binocular stereovision". *Шоста всеукр. міжнар. кон. З оброблення сигналів та розпізнавання образів (УкрОБРАЗ 2002)*. – Київ, 2002. – С. 15-18.
- [4] Boykov Yu., Veksler O., Zabih R. "Fast Approximate Energy Minimization via Graph Cuts", *Intern. Conf. of Comp. Vision*, P. 377-384, 1999.
- [5] Рябоконт Д. "Постобработка результатов стереорекострукции поверхностей как восстановление скрытых параметров марковского поля", *УСМ*. – 2003(4). С.56-61.
- [6] Flach B., Kask E., Schlesinger D., Skulysh A. "Unifying Registration and Segmentation for Multi-Sensor Images" *DAGM2002 - Pattern Recognition Symposium*, Curish.
- [7] Schlesinger M.I., Hlavac V. *Ten Lectures on Statistical and Structural Recognition*. Computational Imaging and Vision, Kluwer Academic Publishers, Dordrecht/Boston/London, 2002. – 522 p.
- [8] Ковтун И.В. "Поиск части оптимальной разметки некоторого NP-полного подкласса (max,+) задач" *УСМ*. – 2003. – №6. С.33-38.
- [9] Соколов А.П. Автостереоскопия и интегральное фотографирование методом Липпманна // Издательство МГУ. – 1911.