

Побудова тривимірної моделі сцени з площин за знімками

Панфілов О.С.

Відділ обробки та розпізнавання зображень
МННЦ ІТС, Київ
alexey.panfilov@gmail.com

Анотація

В статті запропонована технологія автоматизованої побудови моделей просторових сцен за серією знімків. Відновлена модель складається з площин з нанесеними на них реалістичними текстурами. Технологія базується на пошуку проєктивних перетворень між зображеннями цих площин на знімках. За знайденими перетвореннями виконується метрична реконструкція й відповідна ректифікація текстур площин.

Технологія може бути використана для отримання початкових моделей при створенні систем віртуальної реальності.

1. Вступ

В останні роки все більше постає проблема часовитратного створення реалістичних моделей в комп'ютерній 3D-графіці. Багато віртуальних об'єктів походять із реальних і тому раціональніше було б отримувати моделі з них безпосередньо. Враховуючи прогрес цифрової фототехніки, найбільш доречним є використання фотознімків.

В цій статті описана саме така технологія створення моделей за знімками. Використано припущення, що об'єкти на сцені складаються з площин. Зазвичай, це має місце при зйомці приміщень або будівель, зсередини або зовні. Побудова моделей з площин вже описана в деяких роботах. Найближчою до цієї статі є [1]. Та вона заснована на автоматичному пошуку прямих ліній на аерознімках будівель, що не завжди застосовне для складних сцен.

В цій роботі для відновлення моделі пропонується використовувати знайдені перетворення між зображеннями однієї площини на різних знімках. Точність метричної реконструкції в описаному способі не залежить від людського фактору. Від оператора лише залежить яка кількість й які саме площини будуть відновлені у створюваній моделі. Звичайно, в подальшому в таку модель можуть бути додані об'єкти з поверхнями складної форми.

2. Обчислення проєктивних перетворень

2.1. Випадки проєктивних перетворень

Проєктивним перетворенням або гомографією називатимемо будь-яке перетворення площини у площину, при якому колінеарність будь-якої множини точок зберігається. Таке перетворення може бути здійснене множенням на матрицю H розміром 3×3 . Кожна точка p перетворюється у точку p' :

$$p' = Hp, \quad (1)$$

де $p = (x, y, 1)^T$ – однорідний запис координат точки, а H – матриця, що має обернену й важлива з точністю до деякого скалярного множника.

Припустимо 2 знімки зроблені однією або різними камерами з однієї точки простору. Або знімки зроблені камерами на відстані значно меншій ніж відстань до об'єктів зйомки. Тоді між знімками існуватиме проєктивне перетворення.

Справді, нехай проєкційні матриця першої та другої камери $P = KR[I \ -C]$, $P' = K'R'[I \ -C]$ відповідно. Тоді для кожної просторової точки X , що можна побачити з обох камер:

$$\begin{aligned} x' &= P'X = K'R'[I \ -C]X = \\ &= K'R'(KR)^{-1}PX = K'R'(KR)^{-1}x. \end{aligned}$$

Тобто маємо гомографію

$$H = K'R'(KR)^{-1}. \quad (2)$$

Іншим прикладом проєктивних перетворень є гомографія між площиною у просторі та її зображенням на знімку. Справді, нехай на площині у просторі введена система координат з початком у точці X_0 й базисними векторами e_x, e_y . Тоді

$$\begin{aligned} x &= PX = P(X_0 + \alpha e_x + \beta e_y) = \\ &= P[e_x \ e_y \ X_0] \begin{pmatrix} \alpha & \beta & 1 \end{pmatrix}^T. \end{aligned}$$

Тобто існує гомографія

$$H = P[e_x \ e_y \ X_0]. \quad (3)$$

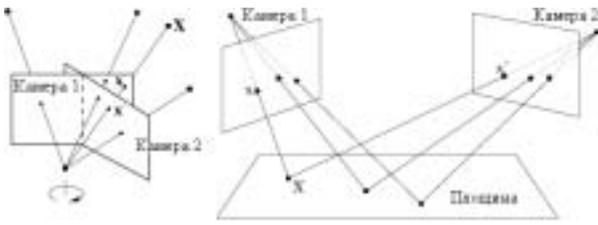


Рис. 1: Випадки проєктивних перетворень.

Зрозуміло також, що буде існувати проєктивне перетворення між зображеннями однієї площини на двох знімках. Далі такі зображення будемо називати відповідними площинами. Справді, якщо H_1 та H_2 – гомографії між площиною та її зображеннями на двох знімках, то проєктивним перетворенням між відповідними площинами буде

$$H = H_1 H_2^{-1}. \quad (4)$$

2.2. Автоматичне знаходження відповідних точок

Припустимо маємо два зображення між якими існує проєктивне перетворення. Для його знаходження необхідно знайти відповідні точки. Звичайно, це може зробити оператор. Та більш раціональний автоматичний спосіб знаходження гомографії. Використаємо такий метод заснований на SIFT (Scale-invariant Feature Transform).

Взагалі, виходячи з назви SIFT, метод призначений для пошуку особливих точок на зображенні які стійкі до повороту або масштабування. Це дуже частковий випадок, та як показує практика ці точки виявляються досить стійкими до зміни освітлення та довільних проєктивних перетворень.

Для кожної особливої точки можна обчислити локальний дескриптор – вектор у 128-вимірному просторі (докладніше в [2]). Цей дескриптор дозволяє порівнювати особливі точки на різних знімках. Вважається, що чим менший кут між векторами локальних дескрипторів двох особливих точок на різних знімках, тим більша впевненість, що це відповідні точки. Тому для кожної пари особливих точок з різних знімків обчислюється кут між їхніми дескрипторами. Для кожної особливої точки обчислюється відношення між двома найменшими кутами, і якщо воно менше за деякий поріг μ і це ж саме виконується для точки з іншого знімку, то такі точки вважаються відповідними. Звичайно, для мінімуму хибних відповідностей обирається поріг 0.5. Та вже маючи деяке обчислене проєктивне перетворення, можна визначити область пошуку, і збільшити значення μ до 0.8. Це дасть значно більшу кількість відповідностей, майже не збільшуючи кількість хибних. Ви-

користавши метод RANSAC (описаний в [3]), можна відсіяти хибні відповідності.

2.3. Знаходження гомографії

Описаний в підрозділі 2.2 спосіб автоматичного пошуку відповідностей дає змогу обчислювати гомографію між знімками за відповідними точками. Звичайно одержані відповідні точки мають деяку похибку в отриманні. Тому необхідно визначитися яку помилку необхідно мінімізувати.

Нехай x_1, x_2, \dots, x_n – знайдені точки на першому знімку, що відповідають точкам x'_1, x'_2, \dots, x'_n на другому знімку. Припустимо існує гомографія між цими знімками. Тоді в ідеалі повинно виконуватися $x'_i = H x_i$ для кожного $i = 1, 2, \dots, n$. Тому перша функція помилки яка спадає на думку

$$d_{STE}^2 = \sum_{i=1}^n \left[d(x_i, H^{-1}x'_i)^2 + d(x'_i, Hx_i)^2 \right], \quad (5)$$

де $d(x, y)$ – звичайна евклідова відстань між точками x та y на зображенні. Таку функцію називатимемо помилкою симетричного переносу.

Та насправді така функція помилки, що здається логічною, не має теоретичного обґрунтування. Згідно принципу максимальної правдоподібності необхідно обчислювати мінімум, так званої, помилки репроєктування

$$d_{RE}^2 = \sum_{i=1}^n \left[d(x_i, \hat{x}_i)^2 + d(x'_i, \hat{H}\hat{x}_i)^2 \right], \quad (6)$$

але не тільки за H , а вже за \hat{H} та \hat{x}_i , $i = 1, 2, \dots, n$. Врахувавши також радіальну дисторсію, звичайно присутню на знімках отримаємо, що замість x_i слід використати $x_i + L(r_i)(x_i - x_C)$, де $r^2 = (x - x_C)^T(x - x_C)$, $L(r) = k_1 r^2 + k_2 r^4$.

Перевагою мінімізації помилки симетричного переносу є невелика обчислювальна складність через те, що оптимізація ведеться лише за 9-ма параметрами, що утворюють матри-

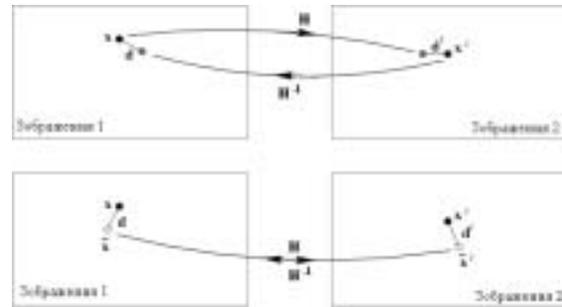


Рис. 2: Порівняння помилки симетричного переносу (зверху) з помилкою репроєктування (знизу).

цю H . Та для мінімізації помилки репроекування можна використати модифікацію методу Левенберга-Марквардта для блочних якобіанів (див. в [4]). Тоді найбільш складною операцією буде обернення матриці 17×17 і така мінімізація буде не набагато складнішою ніж у випадку мінімізації помилки симетричного переносу.

3. Побудова моделі за відповідними площинами

3.1. Відновлення конфігурації площин за гомографіями між ними

Як показано в 2.1 між зображеннями площини на двох знімках існує гомографія (4). Тому почнемо з показу того, що гомографія взагалі однозначно визначає площину якою вона утворена й навпаки. Взагалі означає, що ця площина не містить центри камер. Інакше гомографія буде виродженою.

Нехай положення площини в просторі визначається рівнянням $\pi^T X = 0$ де $\pi = (v^T, 1)^T$. А проєктивні матриці камер $P = [I \ 0]$, $P' = B[I \ -C_b]$.

Для обчислення гомографії H спроекуємо обернено точку x на першому знімку і визначимо точку перетину X цього променя з площиною π . Для першого знімку $x = PX = [I \ 0] X$ й тому будь-яка точка на промені $X = (x^T, \rho)^T$ проєкується в x , де ρ параметризує точку на промені. Просторова точка X належить площині π , а тому $\pi^T X = 0$. Це визначає ρ і $X = (x^T, -v^T x)^T$. Спроекуємо точку X на другий знімок

$$x' = P'X = B[I \ -C_b] X = B(I + C_b v^T)x$$

й одержимо шукану гомографію

$$H = B(I + C_b v^T). \quad (7)$$

В загальному випадку для нетривіальної матриці першої камери

$$P = A[I \ -C_a], \quad P' = B[I \ -C_b]. \quad (8)$$

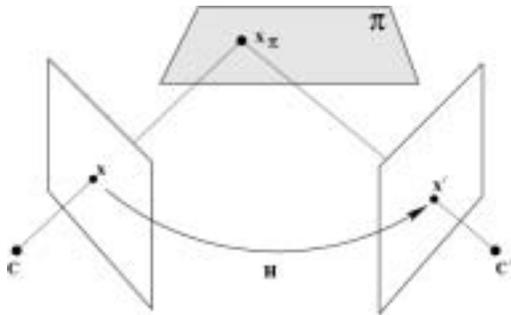


Рис. 3: Гомографія утворена площиною.

виконаємо проєктивне перетворення

$$M = \begin{bmatrix} A & -AC_a \\ 0 & 1 \end{bmatrix}, \quad M^{-1} = \begin{bmatrix} A^{-1} & C_a \\ 0 & 1 \end{bmatrix}. \quad (9)$$

Тоді проєкційні матриці й площина $\pi = (\tilde{\pi}^T, \pi_4)^T$ перетворюються наступним чином

$$\begin{aligned} P_{new} &= PM^{-1} = [I \ 0], \\ P'_{new} &= P'M^{-1} = BA^{-1} [I \ A(C_a - C_b)], \\ \pi_{new}^T &= \pi^T M^{-1} = [\tilde{\pi}^T A^{-1} \ \tilde{\pi}^T C_a + \pi_4]. \end{aligned}$$

А отже за (7) отримаємо

$$H_{new} = B \left(I + \frac{(C_b - C_a) \tilde{\pi}^T}{\tilde{\pi}^T C_a + \pi_4} \right) A^{-1}. \quad (10)$$

Зауважимо, що в разі співпадання центрів камер $C_a = C_b$ гомографія спрощується $H = BA^{-1}$ і вже не залежить від площини. Тобто отримуємо гомографію між повними знінками (пояснює (2)), а не лише між зображеннями площини.

Припустимо за допомогою методів з 2.2 отримано набір гомографій між зображеннями площин. Нехай H_{ij}^k – відома гомографія між зображеннями площини k на знімках i та j , а проєкційна матриця i -ї камери $P_i = M_i [I \ -C_i]$. Тоді помилка метричної реконструкції

$$E_{MR}^2 = \sum_{k,i,j} d \left(H_{ij}^k, M_j \left(I + \frac{(C_j - C_i) \tilde{\pi}_k^T}{\tilde{\pi}_k^T C_i + \pi_{4k}} \right) M_i^{-1} \right)^2, \quad (11)$$

де $d(A, B)$ – норма Фробеніуса. Мінімізація цієї помилки методом Левенберга-Марквардта дозволяє виконати метричну реконструкцію сцени.

Зрозуміло, що не за кожним набором гомографій H_{ij}^k можна виконати реконструкцію сцени. Ці обмеження випливають з кількості шуканих параметрів моделі та кількості зв'язків між ними. По-перше для кожної просторової площини π_k , положення якої визначається, повинна бути обчислена хоча б одна гомографія, тобто її зображення має бути хоча б на двох знімках. По-друге на кожному знімку з параметрами P_i повинно бути принаймні дві відповідні площини. Це випливає з того, що P^j містить 11 параметрів, а кожна гомографія H_{ij}^k дає 8 зв'язків. По-третє подібний підрахунок параметрів повинен діяти і для всієї моделі в цілому. Тобто якщо в моделі N_C знімків, N_π шуканих площин та N_H утворених ними гомографій, то повинно виконуватися:

$$8N_x \geq 11N_C + 3N_\pi. \quad (12)$$

Тобто кількість зв'язків повинна бути не меншою за кількість шуканих параметрів. У разі виконання цих умов повинен існувати єдиний мінімум функції (11).

3.2. Сегментація площин сцени та ректифікація їх текстур

Для отримання текстур необхідна сегментація площин на знімках. Отримані векторизовані контури зображень площин необхідно розділити на трикутники, тобто виконати триангуляцію. Наступним важливим кроком є ректифікація текстур на отриманих трикутниках, тобто виправлення проективного викривлення зображення, що є важливим для реалістичності моделі.

Припустимо є зображення знімку зробленого камерою з проективною матрицею P . На цьому зображенні є трикутник з вершинами x_A , x_B , x_C , про який відомо, що він є частиною деякої площини у просторі. Просторові координати вершин трикутника $X_A = (\tilde{X}_A^T \ 1)^T$, $X_B = (\tilde{X}_B^T \ 1)^T$, $X_C = (\tilde{X}_C^T \ 1)^T$, що належать площині π можна визначити з систем

$$\begin{cases} x_i = PX_i \\ \pi^T X_i = 0. \end{cases} \quad (13)$$

Введемо систему координат в площині якій належить трикутник. Справді нехай X_A – точка відліку, $X_B - X_A$, $X_C - X_A$ – базисні вектори. Тоді будь-яка точка трикутника

$$\begin{aligned} X &= X_A + \beta(X_B - X_A) + \gamma(X_C - X_A) = \\ &= [X_B - X_A \quad X_C - X_A \quad X_A] \begin{pmatrix} \beta \\ \gamma \\ 1 \end{pmatrix}. \end{aligned} \quad (14)$$

Для матриці камери $P = M [I \quad -C]$ маємо $x = PX = M [\tilde{X}_B - \tilde{X}_A \quad \tilde{X}_C - \tilde{X}_A \quad \tilde{X}_A - C] x_\pi$.

Отже для отримання фронтальної текстури трикутника необхідно використати проективне перетворення H^{-1} , де

$$H = M [\tilde{X}_B - \tilde{X}_A \quad \tilde{X}_C - \tilde{X}_A \quad \tilde{X}_A - C]. \quad (15)$$

Тобто для кожного пікселя x_π з фронтальної текстури знайти відповідний піксель на знімку $x = Hx_\pi$ і його колір записати в текстуру. Звичайно можна використати деяке згладжування, наприклад фільтром Ланцоша.

4. Висновки

Запропонований спосіб відновлення моделі сцени за знімками дозволяє створювати точні моделі з площин для приміщень, будівель, тощо. Її головними ознакою є незалежність точності побудови моделі від оператора. Складність такої моделі визначається кількістю площин з якої вона



Рис. 4: Приклад створеної моделі.

складається. Вона може збільшуватися послідовно під наглядом оператора. Але обчислювальна складність запропонованого методу метричної реконструкції є кубічною від кількості знімків задіяних у методі. А тому збільшуючи кількість відновлених площин у моделі, часові витрати на її відбудову не збільшуються.

Недоліком технології є необхідність сегментації зображень площин, що відбудовуються. Також на текстурах відновленої моделі присутні переходи яскравості. Вони пояснюються різними параметрами балансу білого або експозиції на різних знімках. Та навіть за їх фіксації при зйомці, переходи яскравості можуть залишитися через вин'єтування. Тому автоматичному виправленню цього недоліку має бути присвячена окрема робота.

Далеко не всі сцени складаються тільки з площин, тому подальші розробки будуть спрямовані на відбудову поверхонь довільної форми.

5. Література

- [1] Baillard C. and Zisserman A., “Automatic reconstruction of piecewise planar models from multiple views”, IEEE Conf. on Computer Vision and Pattern Recognition, 6:559–565, 1999.
- [2] Lowe D.G., “Object Recognition from Local Scale-Invariant Features”, Int. Conf. Computer Vision(Corfu), Sept. 1999.
- [3] Fischler M.A. and Bolles R.C., “Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography”, Comm. Assoc. Comp. Mach., Vol. 24(6), 1981, p.381–395.
- [4] Hartley R.I. and Zisserman A., “Multiple View Geometry in Computer Vision”, Cambridge University Press, ISBN: 0-521-54051-8, 2004.