

POUŽITIE R-STROMU AKO PRIESTOROVEJ ÚDAJOVEJ ŠTRUKTÚRY PRE SPRACOVANIE MRAČIEN BODOV

Using R-Tree as spatial indexing structure for point cloud processing

Juraj Mužik, Ing. PhD.,

Katedra geotechniky, SvF, Žilinská univerzita v Žiline, Slovensko
juraj.muzik@fstav.uniza.sk

Andrej Villim, Ing.,

Katedra geodézie, SvF, Žilinská univerzita v Žiline, Slovensko
andrej.villim@fstav.uniza.sk

Martin Decký, prof. Dr. Ing.,

Katedra cestného staviteľstva, SvF, Žilinská univerzita v Žiline, Slovensko
martin.decky@fstav.uniza.sk

Matúš Kováč, doc. Ing. PhD.,

Katedra cestného staviteľstva, SvF, Žilinská univerzita v Žiline, Slovensko
matus.kovac@fstav.uniza.sk

Abstract – In this paper, we discussed the usage of spatial access methods, which have been originally developed for organizing spatial data in spatial database systems and GIS, for the persistent storage of point clouds produced by laserscanning. As potential data structures, the BANG file and the buddy tree as representatives of hash trees and the R*-tree as R-trees group have been selected, implemented and experimentally investigated using real laserscanner data. The first results show that both types of index structures have the potential for organizing point clouds originating from laserscanning.

Key words – laser scan, 3D data, GIS, spatial indexing

Úvod

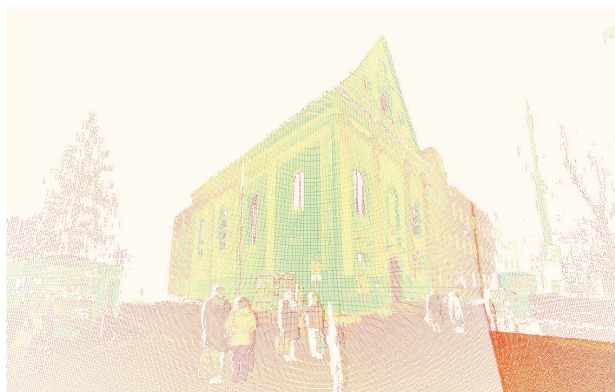
Laserové skenovanie sa v posledných rokoch stáva dôležitým prvkom v diagnostike stavebných objektov. Predstavuje jednoduché a finančne nenáročné meranie priestorových objektov ako fasády alebo interiéry budov. Laserové skenovanie vytvára mračno trojrozmerných bodov, ktoré vyžadujú efektívne a efektne usporiadanie a umiestnenie. Mračno bodov ako výsledok laserového skenovania predstavuje skupinu trojrozmerných bodov (x, y, z) v karteziánskej súradnicovej sústave.

Kvôli veľkému objemu údajov – niekoľko miliónov bodov s narastajúcou tendenciou – nie je vhodné ukladať body

ako konvenčné body v CAD programe. Vo všeobecnosti, prístup ukladania mračna bodov v pamäti RAM má niekoľko nevýhod:

- Takýto prístup si vyžaduje dlhý čas na načítanie údajov z pevného úložiska (ako hard disk).
- Pamäť RAM ukazuje zlú škálovateľnosť z dôvodu swapovania (po dosiahnutí určitej prahovej hodnoty) časti pamäte do sekundárneho úložiska

Alternatívou je použitie trvalých údajových štruktúr. Takéto údajové štruktúry ukladajú údaje na sekundárnom úložisku a dovoľujú načítanie len určitej časti údajov na základe dopytu. Takéto údajové štruktúry boli vyvinuté na indexovanie údajov v databázových systémoch. Ale, konvenčné indexové štruktúry sú optimalizované pre jednorozmerné údajové typy ako čísla a znakové reťazce. Nemôžu byť použité, bez modifikácie, pre priestorové údaje. Z tohto boli vyvinuté údajové štruktúry pre priestorové údaje použiteľné v priestorových databázových systémoch a geografických informačných systémoch (GIS). Tieto priestorové údajové štruktúry sa podľa detailu uchovávaných údajov a spôsobu prístupu k nim delia na bodové a ohraničené. V prípade bodových štruktúr sa každý jeden bod v databáze ukladá ako samostatná entita, na rozdiel od ohraničených, kde atomickú entitu tvorí určitá oblasť ohraničená jednoduchým objektom (kváder, obdĺžnik) v ktorej sa môže nachádzať viacero bodov.



Obrázok 1- Príklad mračna bodov

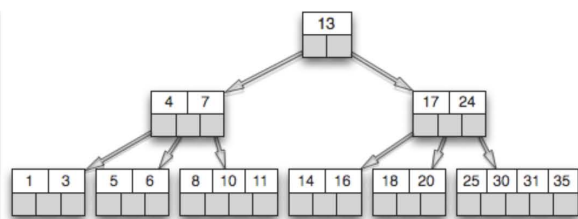
Príspevok sa zaoberá vhodnosťou priestorových vyhľadávacích metód na ukladanie mračien bodov získaných laserovým skenovaním. V kapitole dva sú predstavené rôzne priestorové údajové štruktúry. Hlavné zameranie bude na tzv. Hašovacie stromy (Hash-tree) a R-stromy (R-tree). V kapitole 3 je posúdenie použitia týchto údajových štruktúr pre mračna bodov.

PRIESTOROVÉ ÚDAJOVÉ ŠTRUKTÚRY

II.1 INDEXOVANIE ÚDAJOV V DATABÁZOVOM SYSTÉME

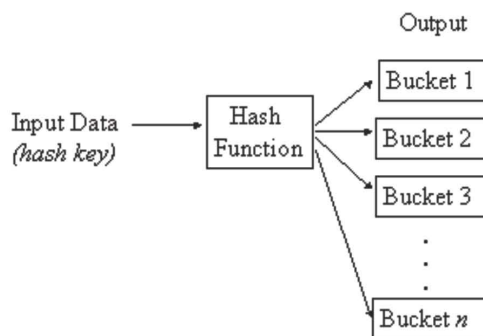
Hlavnou úlohou databázového systému (DBS) je trvalé uloženie veľkého množstva údajov. Systém riadenia bázy údajov (DBMS) musí podporovať vloženie, zmenu a vymazanie ľubovoľných údajov v ľubovoľných sekvenciách. Z tohto dôvodu, DBMS usporadúva údaje do databázových blokov. Vyhľadávanie v rámci sekundárneho úložiska (t.j. zvyčajne hardisk) je vykonávané sekvenčne a často je potrebné preskúmať väčšie množstvo údajov, ktoré nespĺňajú podmienku dopytu. Z dôvodu zrýchlenia vyhľadávania a ukladania informácií je potrebné použiť indexačné údajové štruktúry, ktoré ukazujú na bloky údajov v databáze s minimálnym (v rámci možnosti konkrétnej údajovej štruktúry) prístupom k údajovým blokom nevyhovujúcim dopytu. Medzi najpoužívanejšie indexové štruktúry patria hašovacie tabuľky a stromové štruktúry (B-strom, R-strom).

B-strom je dynamicky vyvážený strom. Každý z jeho uzlov zodpovedá databázovému bloku. B-strom ukladá údaje zoradené podľa indexových atribútov. Spracovanie dopytu začína na najnižšom bode (koreň), prehľadávajú sa len podstromy pokiaľ sa nenájdu požadované údaje. Obrázok 2 zobrazuje takýto B-strom.



Obrázok 2- Príklad B-stromu pre jednoduché hodnoty celých čísel

Hašovanie počíta umiestnenie bloku na sekundárnom úložisku (t.j. blok adres) použitím jedného alebo dvoch vybraných atribútov nahratia. Tento výpočet je urobený pomocou hašovacej funkcie. Obrázok číslo 3 opisuje systém hašovania. Hašovanie podporuje efektívne presné priradenie dopytov, t.j. hľadá záznamy s hodnotami atribútov, ktoré môže presne spojiť s požadovanými požiadavkami. Hašovanie má problém s nerovnomerným rozmiestnením údajov.

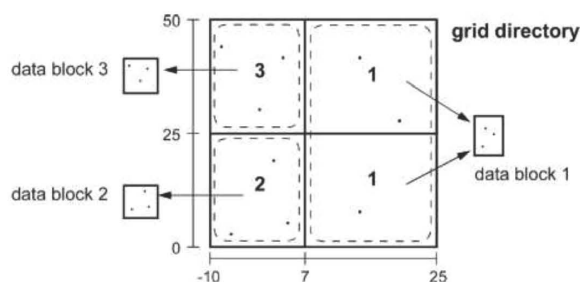


Obrázok 3- Príklad hašovania - vstupné údaje sú na základe hašovacej funkcie prekódované na príslušnosť k jednotlivým kontajnerom (bucket)

II.2 INDEXOVANIE PRIESTOROVÝCH A BODOVÝCH ÚDAJOV

B-stromy vyžadujú lineárne usporiadanie údajov a hašovanie ako bolo spomenuté predtým má problémy s nerovnomernou distribúciou údajov. Z tohto dôvodu nemôžu byť konvenčné indexové štruktúry použité bez vylepšenia na usporiadanie priestorových údajov. Preto boli vyvinuté nové údajové štruktúry umožňujúce vyhľadávanie a dopyty nad priestorovými údajmi.

GRID súbor (GRID file) je príklad pre viacrozmernú vyhľadávaciu metódu, založenú na hašingu. Hašovacia funkcia je nahradená mriežkovým adresárom. Tento adresár ukladá blokove adresy v jeho bunkách (pozri obrázok 4).



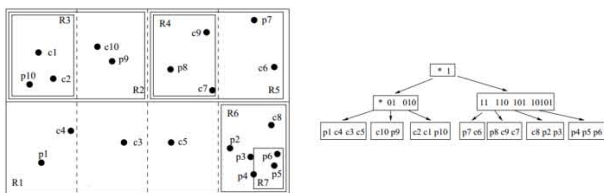
Obrázok 4- Príklad pre mriežkový súbor [3]

Rozdelenie údajového miesta na časti mriežkovým súborom má nasledujúce vlastnosti:

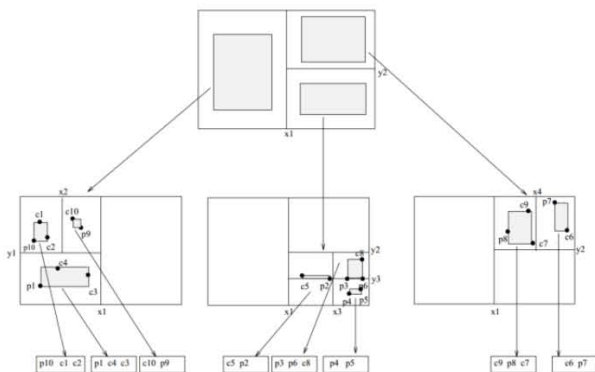
- oblasť je popisovaná údajovým blokom (tzv. oblastný blok), ktorý má tvar obdĺžnika,
- celkový údajový priestor je kompletne pokrytý blokovými regiónmi,
- údajové regióny sa neprekrývajú.

II.3 HAŠOVACIE STROMY

Hašovacie stromy sú viacrozmerné bodové stromy, ktoré kombinujú hašovanie so stromovými údajovými štruktúrami. Typickým príkladom hašovacieho stromu je Bang súbor (Balanced and Nested Grid File-vyvážený a vnorený GRID súbor) [1]. BANG súbor je hierarchická stromová štruktúra. Najnižšia časť stromu je adresár a uzly listov ukladajú reálne údaje ("údajové uzly"). Blokové oblasti adresárových uzlov sú založené na mriežkovej štruktúre a reprezentované (viacrozmerným) obdĺžnikom. Oproti konvenčnému GRID súboru, bloková oblasť nereprezentuje celkovú oblasť tohto obdĺžnika. Ale zahnuté obdĺžniky menších blokových oblastí v tom istom uzle sú premiestnené do obdĺžnika. Následne tvar blokovej oblasti je nepravidelný a môže pozostávať z rôznych nespojených oblastí. Obrázok číslo 5 ukazuje skupinu bodov organizovaných BANG súborom. Obrázok popisuje rozdelenie na časti BANG súboru všetkých uzlov majúcich rovnakú výšku v strome.



Obrázok 5- Príklad rozdelenia na časti BANG súboru [3]



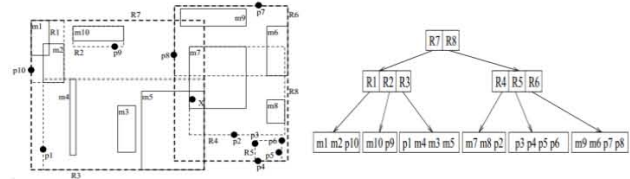
Obrázok 6 - BUDDY strom -rozdelenie údajovej oblasti [3]

Ďalším príkladom hašovacieho stromu je BUDDY strom [1]. BUDDY strom je tiež hierarchický strom s adresárom uzlov obsahujúcim obdĺžnikové blokové oblasti. Oproti mriežkovému súboru a BANG súboru, preto oblasti nepotrebujú pokrytie kompletného údajového priestoru. Obrázok číslo 6 ilustruje rozčlenenie údajovej oblasti použitím BUDDY stromu.

II.4 R-STROMY

R-stromy [2] je priestorová vyhľadávacia metóda, ktorá usporadúva viacrozmerné body tak ako obdĺžniky. R-strom má rovnaké vlastnosti ako B strom ale nevyžaduje si lineárne usporiadanie. Blokové oblasti sú minimálne ohraničujúce obdĺžniky všetkých oblastí alebo údajov v korešpondujúcom podstrome. Tieto blokové oblasti môžu byť prekryté a nevyžadujú pokrytie celého údajového priestoru.

Existujú rôzne varianty R-stromov. Lišia sa v počiatočnej fáze (to znamená že každý podstrom je vybratý pre uloženie nového objektu) a kritéria použité pre rozdelenie uzla ak dôjde k pretečeniu? Obrázok 7 zobrazuje R-strom: zobrazuje blokovú oblasť koreňového uzla a (časť) rozčlenenia uzla zameraného na údajový uzol, ktorý zobrazuje prekrytie medzi blokovými oblasťami.



Obrázok 7- R-strom, rozdelenie údajovej oblasti [1]

II.5 PRAKTICKÉ VLASTNOSTI PREZENTOVANÝCH ÚDAJOVÝCH ŠTRUKTÚR

Prezentované priestorové indexové štruktúry umožňujú vloženie, zmenu a mazanie bodov. Podporujú trvalé umiestnenie údajov na sekundárnom úložisku ako napr. hardisky. Všetky prezentované priestorové vyhľadávacie metódy sú vhodné pre dvoj-, troj- alebo viac rozmerné body. Podporujú efektívne spracovanie základných priestorových požiadaviek. Takéto priestorové požiadavky sú:

- priestorový dopyt na výber údajov,
- určenie najbližšieho k suseda (požiadavka najbližšieho suseda),
- priestorové spojenie.

Ohraničené priestorové štruktúry zachovávajú princíp lokality a v prípade priestorového dopytu sú blízke objekty uložené s veľkou pravdepodobnosťou v tom istom údajovom bloku. Toto je podstatné pre spracovanie priestorových požiadaviek pretože susedné objekty sa často nachádzajú v tom istom bloku, ktorého sa požiadavka týka, prípadne načítanie iného databázového bloku zo sekundárneho úložiska je veľmi nákladná (t.j. pomalá) operácia v porovnaní s operáciami priamo v pamäti RAM. Ukladanie blízkych objektov v tom istom údajovom objekte redukuje počet blokov a zvyšuje pravdepodobnosť nájdenia bloku v medzipamäti databázového systému alebo v pamäti RAM.

Iné techniky optimalizačné prístupy sa snažia, aby priestorová dekompozícia údajov korešpondovala s fyzickými údajovými oblasťami na sekundárnom úložisku. Úlohou je redukovať množstvo prehľadávaných fyzických blokov pri priestorovom dopyte. Uvedené priestorové vyhľadávacie metódy nezachovávajú globálne usporiadanie.

POUŽITIE PRE ÚDAJE Z LASEROVÉHO SKENOVANIA

III.1 PRÍPRAVA PRE ZÍSKAVANIE A SPRACOVANIE ÚDAJOV

Pre vyšetrenie údajov získaných skenovaním pomocou metód popísaných v kapitolách 2.3 a 2.4 boli vytvorené nasledovné údajové štruktúry v prostredí ArcGIS 10:

- [6] IŽVOLTOVÁ, J., KOŤKA, V.: *Tvorba a využitie bázy údajov vo výučbe odboru Geodézia a kartografia*. In: Zborník V. vedecko-odbornej konferencie s medzinárodnou účasťou Geodézia, kartografia a geografické informačné systémy 2008, Stará Lesná, Vysoké Tatry, 16. až 19.9. 2008, ISBN 978-80-553-0079-5
- [7] MASAROVÍČOVÁ, S., SITÁNYIOVÁ, D.: GIS application for solution of the problems of public transportation system in Žilina. In: *Communications - scientific letters of the University of Žilina*. Vol. 6, Nr. 3 (2004), pp. 47-50, ISSN 1335-4205.
- [8] SARDADI, M., DAMAN, D., SHAFRY, M., JUPRI, Z.: *QuadR-Tree Indexing Selection Engine for Tuning Spatial Database System using Mobile Geographical Information System Technology*. Prístupné na: http://cdn.intechopen.com/pdfs/10739/InTech-Quad_r_tree_indexing_selection_engine_for_tuning_spatial_database_system_using_mobile_geographical_information_system_technology.pdf.