

Detekce kartografického zobrazení: techniky založené na srovnání množin identických bodů

Tomáš Bayer

Přírodovědecká fakulta Univerzity Karlovy v Praze, Alebertov 6, Praha 2
bayertom@natur.cuni.cz

Abstrakt. Článek se zabývá problematikou detekce kartografického zobrazení z množiny bodů o známých souřadnicích v mapě. Úlohu detekce kartografického zobrazení převádí na problém opakovaného porovnávání dvojice bodových množin (ve známé a v analyzované mapě) s cílem nalezení dvou nejvíce "podobných" bodových množin. Popisuje čtyři různá testovací kritéria, dle kterých lze usuzovat na vzájemnou podobnost množin identických bodů v obou kartografických dílech. Metodické postupy detekce pro první dvě kritéria vycházejí z aplikace podobnostní transformace, využívají ověření parametrů lokálního a globálního transformačního klíče. Třetí a čtvrté kritérium jsou založena na testování parametrů planárních geometrických struktur, Voroného diagramů, zkonstruovaných nad oběma testovanými množinami. Pro ilustraci vlastností a chování detekčních algoritmů nad různými množinami bodů byly výše uvedené techniky podrobeny čtyřem srovnávacím testům.

Klíčová slova: digitální kartografie, kartografické zobrazení, detekce, Voroného diagram

1 Úvod

Detekce kartografického zobrazení v mapě představuje problém, který lze zařadit do třídy kartometrických analýz. Využití nalezne při analýze starých, historických map či takových kartografických děl, u kterých není informace o kartografickém zobrazení uvedena. Většina starých map však nevznikala na geodetickém podkladu ani za použití geometricko-konstrukčních postupů, v takovém případě nelze hovořit o existenci projekce. Staré mapy vykazují u zobrazovaného území řadu lokálních deformací délkového i úhlového charakteru. Některé části mapového pole mohou být nepřírozně stočeny vůči skutečnosti, mapové prvky si v takovém případě nezachovávají vzájemnou geografickou polohu ve smyslu orientace k světovým stranám. Tyto veličiny mají náhodný charakter, jejich proměnlivost v závislosti na geografické poloze bodu bývá výrazná. Při kartometrické analýze starých map proto vyslovujeme pouze předpoklad, že analyzovaná mapa vykazuje největší podobnost vzhledem k nějakému kartografickému zobrazení.

2 Vstupní parametry

Základní krok procesu detekce kartografického zobrazení představuje nalezení takových geometrických charakteristik prvků analyzované mapy, podle kterých by bylo možno s co nejvyšší mírou pravděpodobnosti rozhodnout o použitém kartografickém zobrazení. Výsledky analýz jsou výrazně ovlivněny charakteristikami analyzované množiny bodů (velikost území, distribuční charakteristika, počet bodů, náhodné odchylky v poloze bodů) a vlastnostmi kartografického zobrazení (poloha zobrazení, konstanty zobrazení, tvar geografické sítě) představující vstupní parametry detekce. Výstupem detekčního algoritmu je číselná hodnota udávající procentuální míru podobnosti analyzované mapy s referenčními mapami téhož území vyhotovenými ve známých kartografických zobrazeních. Na základě procentuální hodnoty podobnosti je vysloven předpoklad, že analyzovaná mapa používá stejné kartografické zobrazení jako referenční mapa s nejvyšší nalezenou mírou podobnosti.

2.1 Testovací množina bodů

Mapové dílo nelze ve většině případů hodnotit z časového či finančního hlediska jako celek, analýzu je možné provádět pouze nad podmnožinou obsahu mapy. Pokud mapa obsahuje zakres geografické sítě, jako perspektivní metoda se jeví možnost detekce kartografického zobrazení dle tvaru poledníků či rovnoběžek. Problematika poskytuje poměrně volné pole působnosti, v kartografické literatuře nebyla doposud detailněji řešena.

Pro detekci kartografického zobrazení lze využít také množinu uzlových bodů geografické sítě. Tímto krokem však dochází k výraznější diskretizaci informace, zanedbáváme veškeré geometrické informace mezi dvěma uzlovými body poledníku či rovnoběžky. Přesto lze pro dostatečně husté rastry tuto metodiku považovat za poměrně spolehlivou. V případech, kdy mapa neobsahuje zakres geografické sítě, je vhodné testování provést nad výběrovým souborem obsahových prvků mapy představovaným významnými a snadno identifikovatelnými prvky. Jedná se zejména o sídla, soutoky řek, hrady, zámky či sakrální stavby.

Parametry testovací množiny. Na testovací množinu klademe řadu požadavků zvyšujících úspěšnost prováděných analýz. Důležitou roli hraje zejména rovnoměrné rozložení testovacích bodů po celé ploše mapy, popisované techniky jsou vhodné pro zpracování množin s přibližně stejnou hustotou bodů. Nepravidelně rozmístěné shluky bodů či místa bez bodů negativně ovlivňují dosažené výsledky. Tento požadavek však v řadě případů není možné u většiny starých map zaručit, hustota mapové kresby nebývá v celé ploše mapového pole rovnoměrná. Typický příklad představuje Fabriciova mapa Moravy, tato vlastnost je zmíněna např. v [7]. Analyzované území by mělo mít podobné rozměry ve směru zeměpisné šířky i délky (tj. bez výraznější dominance v jednom ze směrů), mělo by být dostatečně "velké" (alespoň 200 x 200 km), aby polohové rozdíly bodů testovacích množin v různých kartografických zobrazeních nebyly menší než grafická přesnost analyzované mapy.

Negativní roli hraje umístění analyzovaného území v blízkosti rovníku (podobný tvar geografické sítě u většiny zobrazení) či zeměpisného/kartografického pólu (existence singulárních bodů a výpočetní problémy s tím spojené snižují úspěšnost procesu detekce).

Techniky zaměřené na analýzu podobnosti dvojice množin bodů hrají významnou roli v řadě vědních oborů, např. v astronomii (detekce nových hvězd), biometrii (srovnávání shody otisků prstů) či dálkovém průzkumu Země (automatické hledání identických bodů na snímcích). Příspěvek ilustruje možnosti využití metodických postupů založených na srovnání analyzované mapy s referenční mapou ve známém kartografickém zobrazení (technika je vhodná zejména pro mapy středních měřítek) a seznamuje s předběžnými výsledky.

2.2 Parametry kartografického zobrazení

V současné době je známo několik desítek kartografických zobrazení, jejichž kartografické parametry mohou v závislosti na charakteristice zobrazovaného území variovat. Mezi základní kartografické parametry patří: poloha zobrazení, volba nezkreslených rovnoběžek, volba adičních či multiplikačních konstant, způsob zobrazení kartografického pólu.

Detekce kartografického zobrazení v obecné poloze s neznámými hodnotami dalších kartografických parametrů představuje značně náročný výpočetní proces. V jeho průběhu dochází k přibližnému určení kartografických parametrů (exaktní určení hodnot těchto parametrů není možné) s takovou přesností, aby bylo možno rozhodnout s co nejvyšší spolehlivostí o použitém kartografickém zobrazení.

Hodnoty kartografických parametrů se v průběhu výpočtů nemění spojitě, nabývají pouze diskrétních hodnot v závislosti na zadaných přírůstcích zeměpisné šířky $\Delta\varphi$ a zeměpisné délky $\Delta\lambda$. Mezi určované parametry patří poloha kartografického pólu (tj. současně je prováděna také detekce polohy kartografického zobrazení) a hodnoty konstant zobrazení.

Poloha kartografického pólu i ostatních parametrů kartografického zobrazení se pro každý výpočetní krok mění o zadané přírůstky $\Delta\varphi, \Delta\lambda$. Uvedme ilustrativní příklad: při provádění analýzy s neznámou polohou kartografického pólu mění se v krocích $\Delta\varphi = \Delta\lambda = 10^\circ$ doprovázenou změnou parametrů nezkreslené rovnoběžky s krokem $\Delta\varphi = 10^\circ$ vznikne pro jedno zobrazení cca 10 000 testovacích množin². Jejich počet by bylo možné snížit s využitím symetrie geografické sítě dle základního poledníku, či rovníku, vynecháním polárních oblastí popř. využitím heuristiky, přesto by objemy prováděných výpočtů byly značné. Počet testovacích množin navíc roste vzhledem k velikosti kroku pro výše uvedené parametry kubicky.

Tento příspěvek se omezí pouze na problematiku detekce kartografického zobrazení v normální poloze. Pro zjednodušení dále předpokládejme, že u kuželových a válcových zobrazení bude nezkreslená rovnoběžka procházet těžištěm analyzovaného zobrazení. Ačkoliv je v teoretické rovině k dispozici několik desítek

² Kartografický pól se může nacházet na jednom z 35 poledníků a 16 rovnoběžek, pro každou polohu existuje 16 variant nezkreslených rovnoběžek (při vynechání pólů).

kartografických zobrazení, pro praktickou konstrukci mapových výstupů je používáno pouze malé procento z nich. Do srovnávací množiny bylo zařazeno dvacet nejčastěji používaných kartografických zobrazení ze skupiny jednoduchých, nepravých, modifikovaných či neklasifikovaných zobrazení.

3 Metodika detekce kartografického zobrazení

Kapitola popisuje vybrané metodické postupy detekce kartografického zobrazení, na jejichž základě byla vyvozena kritéria indukující míru podobnosti mezi mapou v analyzovaném kartografickém zobrazení a mapou v referenčním kartografickém zobrazení. Úlohu detekce kartografického zobrazení z mapy lze převést na problém opakovaného porovnávání dvojice bodových množin (referenční a analyzovaná mapa) s cílem nalezení dvou nejvíce "podobných" bodových množin. Problematika analýzy bodových vzorů³ představuje dlouhodobě řešený problém, metodické přístupy lze rozdělit do tří skupin:

- Metody založené na aplikaci lineární transformace. Detekce je založena na analýze parametrů afinní či podobnostní transformace (Chang et al, 1998, Wamlen et al, 2000) popř. vychází z analýzy nn vzdáleností (Upton and Fingleton, 1985, Cressie, 1991). Jejich výhodou je poměrně vysoká relevance výsledků při relativní výpočetní nenáročnosti.
- Metody založené na analýze geometrických parametrů množin. Tato skupina technik je poměrně nová, využívá při porovnávání podobnosti množin pomocné planární geometrické struktury. Existují postupy založené na analýze parametrů jednotlivých trojúhelníků Delaunay triangulací (Ogawa, 1986, Bebis et al, 2000) či Voroného buněk vytvořených nad oběma množinami (Okabe, 1987, Yoshikawa, 1989, Marcelpoil et Usson, 1992).
- Metody založené na využití neuronových sítí (Vinod and Ghose, 1993). Tyto techniky jsou využívány zejména v biometrii, často jsou však předmětem patentových práv popř. nejsou algoritmy detailně publikovány.

Množina testovacích a referenčních bodů. Necht' P a Q představují dvě množiny v rovině se stejným počtem prvků n , pro jejichž prvky p_i a q_i , $i = 1, \dots, n$, platí: $P = \{p_1, p_2, \dots, p_n\}$ a $Q = \{q_1, q_2, \dots, q_n\}$. Prvky $p_i = [x'_i, y'_i]$ a $q_i = [x_i, y_i]$ pak tvoří jednotlivé body v \mathbb{R}^2 . Množina P představuje množinu testovacích bodů, množina Q představuje množinu obrazů referenčních bodů v příslušném kartografickém zobrazení K .

Kartografické zobrazení K . Zobrazuje referenční plochu R_1 na referenční plochu R_2 , jedná se o regulární zobrazení, platí $K : R_1 \rightarrow R_2$. Kartografické zobrazení K lze vzhledem k zavedené symbolice popsat prostřednictvím zobrazovacích rovnic v explicitním vyjádření

³ tzv. points pattern matching problem.

$$\begin{aligned}x &= f(\varphi, \lambda) \\y &= g(\varphi, \lambda)\end{aligned}$$

kde f, g představují zobrazovací funkce. Proměnné φ, λ vyjadřují zeměpisné souřadnice bodu na referenční ploše (rotační elipsoid, koule), x, y pravouhlé souřadnice téhož bodu v rovině kartografického zobrazení.

Míra podobnosti obou množin. Předpokládáme-li existenci podobnosti mezi množinami P a Q , lze vztah mezi těmito množinami vyjádřit s využitím podobnostní transformace $\mathcal{T}(m, a, dx, dy)$, kde m představuje měřítkový koeficient, a úhel stočení, dx, dy translace ve směru souřadnicových os x, y

$$\begin{pmatrix} x'' \\ y'' \end{pmatrix} = m \begin{pmatrix} \cos(\alpha) & -\sin(\alpha) \\ \sin(\alpha) & \cos(\alpha) \end{pmatrix} \begin{pmatrix} x' \\ y' \end{pmatrix} + \begin{pmatrix} dx \\ dy \end{pmatrix}. \quad (1)$$

Míru podobnosti μ , $\mu \in (0, 1)$, lze symbolicky definovat jako poměr velikosti množin R a P

$$\mu = \frac{\|R\|}{\|P\|}, \quad (2)$$

kde $R \subset P$. Množina $R = \{r_1, r_2, \dots, r_m\}$, kde $r_j = [x_j'', y_j'']$ představuje takovou

podmnožinu P tvořenou m prvky, $m \leq n$, pro které platí

$$|T(P) - Q| < \varepsilon. \quad (3)$$

Vztah (2) formálně vyjadřuje poměr počtu obrazů bodů $\mathcal{T}(P)$ ležících do vzdálenosti ε od bodů Q . Hodnotu ε lze vyjádřit jako poloměr dostatečně malé kružnice zkonstruované nad každým bodem množiny Q , viz obr. 1. Předpokládáme, že čím silnější podobnost mezi množinami P a Q existuje, tím více obrazů bodů P leží uvnitř kružnic s poloměrem ε . Vztah (2) vyjadřuje míru podobnosti obou množin.

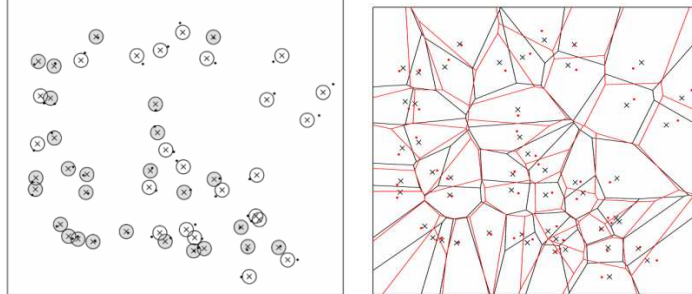
Volba hodnoty ε . Volba hodnoty ε výrazně ovlivňuje určenou míru podobnosti μ obou množin. Zvolíme-li poloměr ε příliš velký, bude podmínku (3) splňovat značné množství bodů ($R \rightarrow P$), v opačném případě může být R prázdnou množinou (pro $\varepsilon \rightarrow \infty : m \rightarrow n$, pro $\varepsilon \rightarrow 0 : m \rightarrow 0$). Hodnota ε není pro množiny P, Q konstantou, je závislá na velikosti území pokrytého prvky obou množin a počtu bodů n obou množin. Rozměr území lze vyjádřit s využitím planárních geometrických struktur, např. poloměrem r nejmenší kružnice opsané množině Q či delší z dvojice hran min-max boxu d zkonstruovaného nad Q . Obě varianty budou dávat podobné výsledky. Míra ztotožnění množin P, Q vyjádřená směrodatnou odchylkou $\sigma_p = \sqrt{\sigma_x^2 + \sigma_y^2}$ je funkcí počtu identických bodů n . Poloměr ε proto zohledňuje velikost obou množin, jeho hodnota se zmenšuje s odmocninou počtu bodů n . Testovací kritérium založené na hodnotě ε lze zapsat ve tvaru (Wamlen et al, 2000)

$$\varepsilon < \frac{d}{\sqrt{n}} \doteq \frac{r}{2\sqrt{n}}. \quad (4)$$

Změna citlivosti detekce. Pro změnu citlivosti detekčního algoritmu lze využít upravenou variantu kritéria (4) ve tvaru

$$\varepsilon < \beta \frac{d}{\sqrt{n}}. \quad (5)$$

Koeficient β , $\beta \in (0, 1)$, upravuje citlivost detekčního procesu vůči polohovým odchylkám prvků množin P , Q . U množin zatížených chybami náhodného charakteru (staré mapy) je nutné hodnoty β volit větší, $\beta \doteq 0.1$, v ostatních případech postačují hodnoty o jeden či dva řády nižší.



Obr. 1 Vlevo: Symbolické znázornění detekce kartografického zobrazení s využitím hodnoty ε představující poloměr kružnice zkonstruované nad body množiny Q , u bodů splňujících podmínku (4) kružnice zvýrazněny. Vpravo: Voroného diagramy podobných množin P a Q jsou podobné.

4 Detekce kartografického zobrazení s využitím podobnostní transformace

Technika využívá globální transformační klíč nad množinami P a Q daný parametry m , a , dx , dy podobnostní transformace T . Parametry transformačního klíče jsou určeny z podmínky metody nejmenších čtverců. Z důvodu jednoduchosti implementace byla použita upravená varianta Helmertovy transformace spočívající v zavedení redukovaných souřadnic $[x^{(r)}, y^{(r)}]$ a $[\xi, \eta]$ k těžišti identických bodů obou množin. Pro redukované souřadnice platí:

$$\begin{aligned} x_i^{(r)} &= x_i - \sum_{i=1}^n \frac{x_i}{n} & y_i^{(r)} &= y_i - \sum_{i=1}^n \frac{y_i}{n}, \\ \xi_i &= x_i' - \sum_{i=1}^n \frac{x_i'}{n} & \eta_i &= y_i' - \sum_{i=1}^n \frac{y_i'}{n}. \end{aligned}$$

Hodnoty transformačních koeficientů λ_1, λ_2 určíme z

$$\lambda_1 = \frac{\sum_{i=1}^n (x_i^{(r)} \xi_i + y_i^{(r)} \eta_i)}{\sum_{i=1}^n (\xi_i \xi_i + \eta_i \eta_i)}, \quad \lambda_2 = \frac{\sum_{i=1}^n (y_i^{(r)} \xi_i - x_i^{(r)} \eta_i)}{\sum_{i=1}^n (\xi_i \xi_i + \eta_i \eta_i)},$$

Kde $\lambda_1 = m \cos(\alpha)$, $\lambda_2 = m \sin(\alpha)$ a $dx = 0$, $dy = 0$. S využitím (1) spočteme přetransformované souřadnice identických bodů $[x'', y'']$ a určíme hodnoty oprav v_x, v_y na identických bodech

$$v_i^x = x_i - x_i'' \quad v_i^y = y_i - y_i''.$$

Následně porovnáme hodnotu opravy v_i^{xy} na každém identickém bodě s hodnotou testovacího kritéria ε . Je-li splněna podmínka

$$v_i^{xy} = \sqrt{(v_i^x)^2 + (v_i^y)^2} \leq \varepsilon, \quad (6)$$

inkrementujeme počet bodů náležících do množiny R . Po porovnání všech identických bodů určíme dle (2) hodnotu kritéria μ . Postup aplikujeme na všechny dvojice analyzovaná mapa-referenční mapa. Na základě hodnot míry podobnosti μ je k analyzované mapě nalezena referenční mapa s hodnotou μ_{max} která je označena jako mapa k analyzované mapě nejvíce podobná. Kartografické zobrazení referenční mapy je přisouzeno analyzované mapě.

Odstranění nepřesně zakreslených obsahových prvků. Vzhledem k faktu, že staré mapy nebývají vyhotoveny na geodetickém podkladu ani za použití geometricko-konstrukčních postupů, dochází u nich při zákresu jednotlivých obsahových prvků k výrazným polohovým chybám náhodného charakteru. Tyto prvky zařazené do transformačního klíče mohou významně ovlivnit hodnoty transformačních koeficientů a zkreslit výsledek kartometrické analýzy.

První krok předcházející vlastnímu procesu detekce kartografického zobrazení představuje odstranění nepřesně zakreslených obsahových prvků mapy z transformačního klíče. Pro detekci nepřesně zakreslených prvků v mapě lze využít (6). Z transformačního klíče odstraníme takové prvky, u kterých bude oprava v_i^{xy} překračovat 2.5 násobek směrodatné odchylky σ_{xy} tj. σ_{xy} , tj. $v_i^{xy} > 2.5\sigma_{xy}$ kde

$$\sigma_{xy} = \sqrt{\frac{\sum_{i=1}^n v_i^{xy} v_i^{xy}}{n}}$$

Alternativu k tomuto postupu může představovat analýza poklesu oprav, který je však z výpočetního hlediska výrazně náročnější.

5 Detekce kartografického zobrazení s využitím metody k-nejbližších sousedů

Tato technika je modifikací předchozího postupu, pro detekci podobnosti množin P a Q , opakovaně využívá lokální transformační klíč t definovaný parametry m , a , dx , dy , (Wamelen et al, 2000). Lokální transformační klíč je určen ze vzájemně odpovídajících si bodů p_i , q_i a jejich k nejbližších sousedů, je tedy tvořen $k + 1$ identickými body. Hodnota k je dalším vstupním parametrem detekčního algoritmu. V souladu s výše uvedenou symbolikou označme k -tý nejbližší bod k bodu p_i jako $p_i^{(k)}$ a k -tý nejbližší bod k bodu q_i jako $q_i^{(k)}$.

Princip metody. Princip metody detekce kartografického zobrazení metodou k-nejbližších sousedů lze formálně vyjádřit takto: Každému bodu p_i , q_i nalezneme k nejbližších sousedů. Z $k + 1$ bodů (doplňných body p_i , q_i) sestavíme s využitím heuristiky nejlepší možný lokální transformační klíč. Tento lokální transformační klíč bude použit k ověření podmínky (1) pro všechny body množin P , Q . Pro každou dvojici bodů p_i , q_i následně obdržíme hodnotu μ_i představující lokální míru

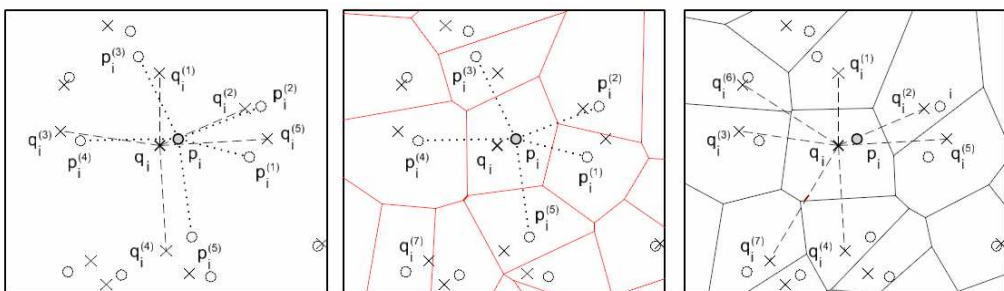
podobnosti obou množin. Globální míru podobnosti množin μ určíme jako aritmetický průměr

$$\mu^{(2)} = \frac{\sum_{i=1}^n \mu_i}{n}. \quad (7)$$

Lze využít i váženou variantu kritéria zohledňující vzdálenosti k nejbližších sousedů od bodů q_i . Váhu w_i průměrné vzdálenosti k nejbližších bodů od bodu q_i určíme ze vztahu

$$\mu^{(2)} = \frac{\sum_{i=1}^n w_i \mu_i}{\sum_{i=1}^n w_i} \quad w_i = \frac{\sum_{j=1}^k d(q_i, q_i^{(j)})}{k}. \quad (8)$$

Tímto krokem přisoudíme lokálním transformačním klíčům vzniklých ze vzdálenějších bodů větší váhu.



Obr. 2 Vlevo: Nalezení 5 nejbližších sousedů $p_i^{(1)}$ až $p_i^{(5)}$ k bodu p_i a 5 nejbližších sousedů $q_i^{(1)}$ až $q_i^{(5)}$ k bodu q_i . Uprostřed: nalezení 5 nejbližších sousedů $p_i^{(1)}$ až $p_i^{(5)}$ k bodu p_i . Vpravo nalezení 7 nejbližších sousedů $q_i^{(1)}$ až $q_i^{(7)}$ k bodu q_i s využitím Voroného diagramu.

Metodika nalezení k nejbližších sousedů. Problém nalezení k nejbližších sousedů lze řešit dvěma přístupy. První přístup vychází z předpokladu, že hodnota k je pevně dána, druhý přístup určuje hodnotu k průběžně, a to na základě geometrických parametrů Voroného buněk.

První varianta využívá pevně daný počet k nejbližších sousedů. Postup jejich nalezení je jednoduchý, k nejbližších sousedů představuje prvních k prvků $p_i^{(1,k)}$ množiny $P - \{p_i\}$ resp. k prvků $q_i^{(1,k)}$ množiny $Q - \{q_i\}$ vzestupně seřazených dle vzdálenosti všech bodů k bodu p_i resp. k bodu q_i . Alternativou k tomuto postupu by mohlo být k opakování Hoarova algoritmu hledajícího k -tý nejmenší prvek množiny přístupem divide and conquer nad neseřazenou posloupností. Hodnoty k volíme zpravidla v intervalu $\langle 3, 6 \rangle$. Pro větší hodnoty se postup stává výpočetně náročným a bude poskytovat podobné výsledky jako globální klíč. Lokální transformační klíč je tvořen $k+1$ identickými body.

Druhý přístup vychází z faktu, že se počet nejbližších sousedů lokálně mění v závislosti na parametrech obou množin. Metoda využívá konstrukci Voroného diagramu nad množinami P, Q . Za nejbližší sousedy jsou označeny takové body,

jejichž Voroného buňka sdílí alespoň jednu hranu s Voroného buňkou generátoru p_i pro množinu testovacích bodů či alespoň jednu hranu s Voroného buňkou generátoru q_i pro množinu testovacích bodů. Počet nejbližších sousedů k bodu p_i resp. q_i může být různý, viz obr. 2, a označíme ho k_p resp. k_q . Hodnotu k pak určíme jako $\min(k_p, k_q)$. Lokální transformační klíč je tvořen k identickými body vybranými z množiny tvořené $\max(k_p, k_q)$ prvky, některé identické body s nevhodnými parametry nebudou do klíče zařazeny. Tato varianta je však výpočetně náročnější.

5.1 Konstrukce lokálního transformačního klíče

Cílem následujícího kroku detekčního algoritmu je nalezení takového optimálního lokálního transformačního klíče $t(m, a, dx, dy)$, který co nejlépe splňujícího podmínku (Wamlen et al, 2000)

$$\left| t(p_i, p_i^{(1,k)}) - (q_i, q_i^{(1,k)}) \right| < \varepsilon. \quad (9)$$

Pro výpočet lokálního transformačního klíče nelze přímo použít uspořádané dvojice nejbližších sousedů $(p_i^{(j)}, q_i^{(j)})$. Nalezení nejbližší sousedé jsou seřazeni pouze podle vzdálenosti k bodům p_i resp. q_i z této informace však nelze získat detailnější informaci o jejich vzájemné poloze. Mezním případem může být situace, kdy pro všechny nejbližší sousedy platí $d(q_i, q_i^{(1,k)}) = konst$ resp. $d(q_i, q_i^{(1,k)}) = konst$, tj. tyto body leží na kružnici. Lokální transformační klíč sestavený metodikou vzdálenostně odpovídajících si dvojic nejbližších sousedů by byl v takovém případě nepoužitelný. Při detekci kartografického zobrazení se s podobným problémem můžeme setkat poměrně často, uzlové body geografické sítě a jejich obrazy představují pravidelný či semipravilný rastr. V průběhu procesu detekce kartografického zobrazení je nalezeno celkem n lokálně optimálních transformačních klíčů.

Proces konstrukce lokálního transformačního klíče. Postup nalezení optimálního transformačního klíče je tvořen několika kroky, které se pravidelně opakují. Využívá heuristiku založenou na předpokladu, že lokálně optimální dvojice identických bodů nejméně "zhorší" míru ztotožnění obou množin bodů. V každém kroku algoritmu je hledán lokálně optimální klíč t_{loc} (vzniklý přidáním lokálně optimálních bodů vzhledem k t) s cílem nalezení optimálního klíče t po k krocích. První dvojici identických bodů lokálního transformačního klíče budou představovat body p_i a q_i . V dalších krocích se k nějakému nejbližšímu sousedovi $p_i^{(j_1)}$ snažíme nalézt nejbližšího souseda $q_i^{(j_2)}$, $j_1, j_2 \in \langle 1, k \rangle$ takového, aby přidání páru $(p_i^{(j_1)}, q_i^{(j_2)})$ do stávajícího lokálního klíče t splnilo podmínku pro body $\sigma_{xy} = \min(\sigma_{xy})$. Po nalezení odpovídajícího páru jsou oba body $(p_i^{(j_1)}, q_i^{(j_2)})$ odstraněny seznamu nejbližších sousedů a postup hledání další vhodné dvojice se opakuje (celkem k -krát). V prvním kroku je provedeno k^2 porovnání, v druhém kroku $(k-1)^2$ porovnání, atd. Existuje celkem $\frac{1}{2}k(k+1)(2k+1)$ takových párů, tento krok

vykazuje kubickou složitost, což pro větší množiny bodů může představovat omezující faktor. V software ilustrujícím metody detekce kartografického zobrazení byla implementována tato varianta hledání lokálního klíče.

Upravený postup konstrukce lokálního transformačního klíče. Předchozí postup lze modifikovat zavedením upravené heuristiky s cílem dosažení kvadratické složitosti. V každém kroku nebudeme hledat nejlepší možný pár, ale k aktuálnímu nejbližšímu sousedovi $p_i^{(j)}$ (vybraného ze seznamu nejbližších sousedů bodu p_i) se snažíme nalézt nejbližšího souseda $q_i^{(l)}$, $l \in \langle 1, k \rangle$ takového, aby přidání páru $(p_i^{(j)}, q_i^{(l)})$ do stávajícího lokálního klíče t splnilo podmínku $\sigma_{xy} = \min(\sigma_{xy})$. Na rozdíl od předchozího případu je bod $p_i^{(j)}$ dán, nemůžeme ho volit. Po nalezení takového páru jsou oba body $p_i^{(j)}, q_i^{(l)}$ odstraněny ze seznamu nejbližších sousedů a postup se opakuje pro následujícího nejbližšího souseda $p_i^{(j+1)}$. Existuje celkem $\frac{1}{2}k(k-1)$ takových párů, postup vykazuje kvadratickou složitost. Popsaná úprava sníží výpočetní dobu tohoto kroku za cenu mírného zhoršení parametrů nalezeného transformačního klíče.

5.2 Aplikace lokálního transformačního klíče jako globálního

Proces detekce kartografického zobrazení je založen na nalezení celkem n lokálně optimálních transformačních klíčů t_i . Každý lokální transformační klíč t_i bude aplikován na všechny body množiny P a bude ověřena podmínka

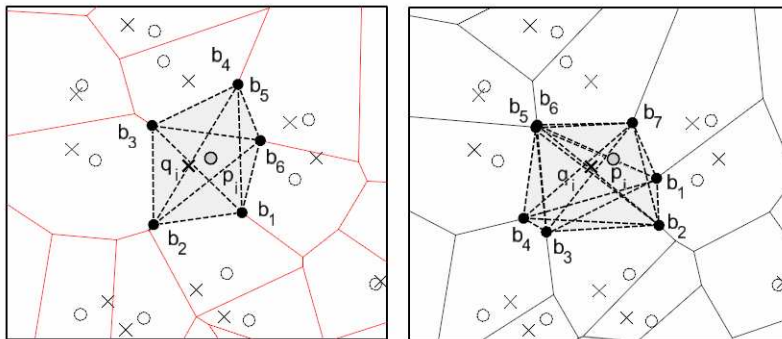
$$|t_i(P) - Q| < \varepsilon. \quad (10)$$

Lokální transformační klíč bude tedy použit jako globální klíč, na jehož základě bude určena hodnota kritéria podobnosti množin μ jako aritmetický průměr z hodnot μ_i , viz (7). Hodnoty kritéria μ mohou být vizualizovány s využitím běžných metod tématické kartografie, např. technikou izochar. Lze tak získat informaci o lokální podobnosti mapy v okolí identických bodů množin P a Q .

6 Detekce s využitím pomocných geometrických struktur

Problematiku detekce kartografického zobrazení převádíme na problematiku opakovaného porovnávání dvojice bodových množin s cílem nalezení dvou nejvíce "podobných" bodových množin. Existuje řada technik založená na analýze parametrů planárních geometrických struktur zkonstruovaných nad oběma množinami bodů. Pro tyto účely lze použít jak jednodušší geometrické útvary, např. elipsu (Li Tian, 2008), tak i geometrické struktury složitější, představované Voroného diagramy (Marcelpoil et Usson, 1992, Perry 1995, Yoshikawa, 1989 či Delaunay triangulaci (Ogawa, 1986, Finch et Wilson, 1995, Bebis et Deaconu, 1999).

Metodika detekce kartografického zobrazení je pak založena na srovnání parametrů pomocných geometrických struktur zkonstruovaných nad množinou bodů v analyzované mapě s parametry geometrických struktur zkonstruovaných nad odpovídající množinou bodů ve známém kartografickém zobrazení. Za podobné jsou prohlášeny takové množiny bodů, u kterých si tyto parametry odpovídají. Metody založené na analýze parametrů geometrických struktur mají řadu výhod, např. invarianci vůči vzájemnému natočení množin. Za nevýhodu lze považovat nižší citlivost detekce a menší odolnost při zanesení náhodné chyby do vstupních množin (tvary pomocných geometrických struktur se mohou významně změnit při přidání polohově "nevhodného" bodu).



Obr. 3 Převod Voroného buňky $\partial V(p_i)$ resp. $\partial V(q_i)$ na strukturu reprezentovanou maticí Wp_i resp. maticí Wq_i , vzdálenosti $d(b_j, b_k)$ znázorněny čárkovaně.

6.1 Detekce kartografického zobrazení založená na analýze Voroného diagramů

Voroného teselace VT přiřazuje každému bodu množiny $P = \{p_1, p_2, \dots, p_n\}$ resp. $Q = \{q_1, q_2, \dots, q_n\}$ uzavřenou či otevřenou oblast $V(P) = \{V(p_1), V(p_2), \dots, V(p_n)\}$ resp. $V(Q) = \{V(q_1), V(q_2), \dots, V(q_n)\}$, takovou, že libovolný bod $A \in V(p_i)$ resp. $A \in V(q_i)$ je blíže k bodu p_i resp. q_i než k jakémukoliv bodu p_j resp. q_j . Uzavřenou oblast $V(p_i)$ resp. $V(q_i)$ nazýváme Voroného buňkou množiny P resp. Q . Označme $\partial V(p_i) = \{b_1, b_2, \dots, b_m\}$ hranici Voroného buňky $V(p_i)$ tvořenou m vrcholy b_j a analogicky hranici Voroného buňky $V(q_i)$ jako $\partial V(q_i)$.

Podobnost Voroného diagramů. Posouzení podobnosti $V(P)$ a $V(Q)$ je klíčovým krokem detekce. Vzhledem k faktu, že Voroného diagram je značně citlivý k poloze generátorů, změna polohy několika z nich může vyvolat výraznou změnu jeho tvaru. Této vlastnosti využijeme při detekci tvaru kartografického zobrazení. Lze tvrdit: $V(P)$ je podobné $V(Q)$ právě když jsou podobné i jejich geometrické parametry. Proces detekce omezíme na takové Voroného buňky $\partial V(p_i)$ resp. $\partial V(q_i)$, jejichž vrcholy leží *uvnitř* min-max boxu vygenerovaného nad množinou P resp. Q . Tímto

krokem vynecháme Voroného buňky nevhodných tvarů ležící na okrajích analyzovaného území.

Rada kritérií založená na analýze Voroného diagramu je výpočetně značně náročná, zejména techniky využívající nn -vzdálenosti (Okabe, 1987, Yoshikawa, 1989). Uvedeme dvojici kritérií, první je založeno na analýze vlastních čísel nad Voroného buňkami (Bayer, 2006), druhé na analýze tvaru Voroného buňky (Marcelpoil et Usson, 1992). Výhodou druhého kritéria je jeho relativní snadnost výpočtu.

Lokální testovací kritérium založené na analýze vlastních čísel γ_1 . Zprostředkující veličinou je poměr sumy kvadrátů vlastních čísel λ_P resp. λ_Q regulární matice W nad uzavřenou oblastí $\partial V(p)$ resp. $\partial V(q)$ a plochy A oblastí $\partial V(p)$ resp. $\partial V(q)$.

$$\gamma_1^{(Q)} = \sum_{i=1}^n \frac{\sum_{j=1}^m (\lambda_{Q_i}^j)^2}{A(\partial V(q_i))} \quad \gamma_1^{(P)} = \frac{\sum_{j=1}^m (\lambda_{P_i}^j)^2}{A(\partial V(p_i))}. \quad (11)$$

Symetrickou regulární matici W řádu m , $j, k \in \langle 1, m \rangle$ nad hranicí oblasti $\partial V(p_i)$ označíme W_{P_i} , analogicky označíme matici nad hranicí oblastí $\partial V(q_i)$ jako W_{Q_i} . Prvky w_{jk} matice W jsou nad uzavřenou oblastí $\partial V(p_i)$ resp. $\partial V(q_i)$ definovány následujícím vztahem, viz obr. 3:

$$w_{ij} = \begin{cases} 0, & \text{pro } j = k, \\ d(b_j, b_k), & \text{pro } j \neq k. \end{cases} \quad (12)$$

Tato matice uchovává tvarové charakteristiky Voroného buňky, je navíc invariantní vůči úhlu vzájemnému natočení množin P a Q . Lokální kritérium je určováno nad všemi dvojicemi Voroného buněk, jejichž generátory p_i a q_i neleží na konvexní obálce (tj. buňky jsou uzavřené). Kritérium vychází z předpokladu, že pokud jsou množiny P a Q podobné (tj. generují podobné Voroného buňky), je libovolná matice W_{Q_i} α -násobkem odpovídající matice W_{P_i} . Zřejmě platí

$$W_{Q_i} = \alpha W_{P_i}$$

Pro vlastní čísla λ matic W_P a W_Q platí

$$\sum_{j=1}^m \lambda_{Q_i}^j = \alpha \sum_{j=1}^m \lambda_{P_i}^j. \quad (13)$$

Pro sumy kvadrátů vlastních čísel matic W_P a W_Q lze psát rovnost, která je základem testovacího kritéria

$$\sum_{j=1}^m (\lambda_{Q_i}^j)^2 = \alpha^2 \sum_{j=1}^m (\lambda_{P_i}^j)^2. \quad (14)$$

Hodnota α zohledňuje tvar Voroného buněk, poměr $\frac{\sum_{j=1}^m (\lambda^j)^2}{A}$ je pro tvarově podobné buňky konstantní. Globální testovací kritérium $\eta^{(1)}$ lze zapsat ve tvaru

$$\eta^{(1)} = \frac{1}{1 + |\gamma_1^{(Q)} - \gamma_1^{(P)}|}. \quad (15)$$

Za nevýhodu kritéria založeného na analýze vlastních čísel lze, kromě výpočetní náročnosti, považovat jeho citlivost na tvar Voroného buněk. Tato vlastnost se projevuje zejména u semipravidelných rastrů s centrálním bodem (např. geografická síť azimutálních zobrazení). Malé chyby v určení polohy bodů (popř. zaokrouhlovací

chyby) ležících kolem centrálního bodu způsobí změny tvaru Voroného buněk spočívající v přidání polohově "nevýrazných" vrcholů (a jim příslušejících velmi krátkých hran), které však vyvolají značné změny hodnot kořenů charakteristických rovnic.

Lokální testovací kritérium γ_2 . Zprostředkující veličinou je poměr plochy $A(\partial V(q_i))$ resp. $A(\partial V(p_i))$ Voroného buňky ke čtverci obvodu $L(\partial V(q_i))$ resp. $L(\partial V(p_i))$ Voroného buňky. Upravené *RF* kritérium (Marcelpoil et Usson, 1992) lze přepsat do tvaru

$$\gamma_2^{(Q)} = \sum_{i=1}^n \frac{A(\partial V(q_i))}{L^2(\partial V(q_i))} \quad \gamma_2^{(P)} = \frac{A(\partial V(p_i))}{L^2(\partial V(p_i))}. \quad (16)$$

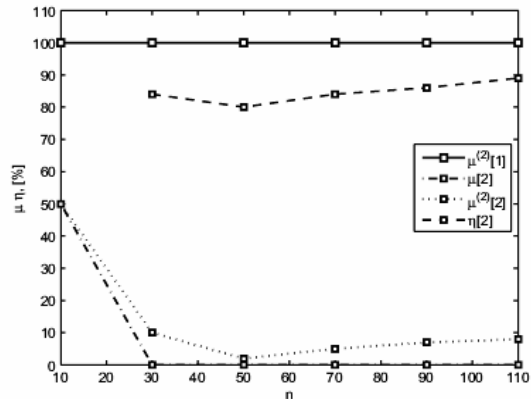
Hodnota γ zohledňuje nejen plochu Voroného buněk, ale částečně také jejich tvar, poměr $\frac{\sum_{j=1}^m (\lambda^j)^2}{A}$ je pro tvarově podobné buňky konstantní. Globální testovací kritérium $\eta^{(2)}$ lze zapsat ve tvaru

$$\eta^{(2)} = \frac{1}{1 + |\gamma_2^{(Q)} - \gamma_2^{(P)}|}. \quad (17)$$

Pro pravidelné rastry lze toto kritérium považovat za spolehlivější než kritérium založené na výpočtu vlastních čísel, hodnota γ_2 je méně citlivá na přidání polohově nevýrazných vrcholů do Voroného buňky (poměr se změní méně než hodnota kořenů charakteristických rovnic). K této situaci dochází zejména u degenerovaných Voroného diagramů, malá změna polohy některého z generátorů způsobí změnu počtu vrcholů Voroného buněk, tvar Voroného buňky se však významně nezmění.

7 Praktické výsledky detekčních metod

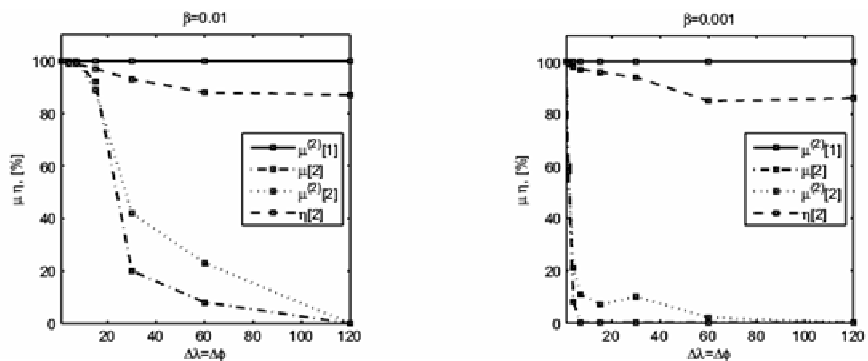
Nalezení srovnávací metodiky ověřující účinnost detekčních kritérií představuje poměrně složitý problém přesahující rozsah tohoto příspěvku. Pro ilustraci vlastností a chování detekčních algoritmů nad různými množinami bodů byly výše uvedené techniky podrobeny čtyřem srovnávacím testům. Testy byly provedeny nad třemi typově různými kartografickými zobrazeními: válcové konformní, sinusoidální a azimutální ekvidistantní zobrazení vzhledem k množině 20 referenčních kartografických zobrazení.



Obr. 4 Závislost druhé nejvyšší hodnoty detekčních kritérií μ , $\mu(2)$, $\eta(2)$ na počtu bodů a její srovnání s nejvyšší hodnotou těchto kritérií.

Do testu nebylo zařazeno kritérium $\eta^{(1)}$, a to z důvodu pomalé konvergence iteračního algoritmu pro výpočet vlastních čísel založeného na QR rozkladu implementovaného autorem. Tento krok algoritmu bude nutné dále optimalizovat tak, aby byl algoritmus použitelný v reálném čase.

U prvních tří testů budeme analyzovat dvě nejvyšší hodnoty kritérií μ , $\mu^{(2)}$, $\nu^{(2)}$ ilustrující nejen úspěšnost/neúspěšnost detekce, ale i vlastní efektivitu detekčního procesu. Velikost druhé nejvyšší hodnoty testovacího kritéria $\mu[2]$, $\mu^{(2)}[2]$, $\nu^{(2)}[2]$ ve vztahu k nejvyšší hodnotě testovacího kritéria $\mu[1]$, $\mu^{(2)}[1]$, $\nu^{(2)}[1]$ naznačuje jednoznačnost provedené detekce. Pokud se hodnoty $\mu[2]$, $\mu^{(2)}[2]$, $\nu^{(2)}[2]$ blíží hodnotám $\mu[1]$, $\mu^{(2)}[1]$, $\nu^{(2)}[1]$, výsledky detekce nejsou jednoznačné.



Obr. 5. Závislost druhé nejvyšší hodnoty detekčních kritérií μ , $\mu(2)$, $\eta(2)$ na velikosti území a její srovnání s nejvyšší hodnotou těchto kritérií pro $\beta = 0.01$ a $\beta = 0.0001$.

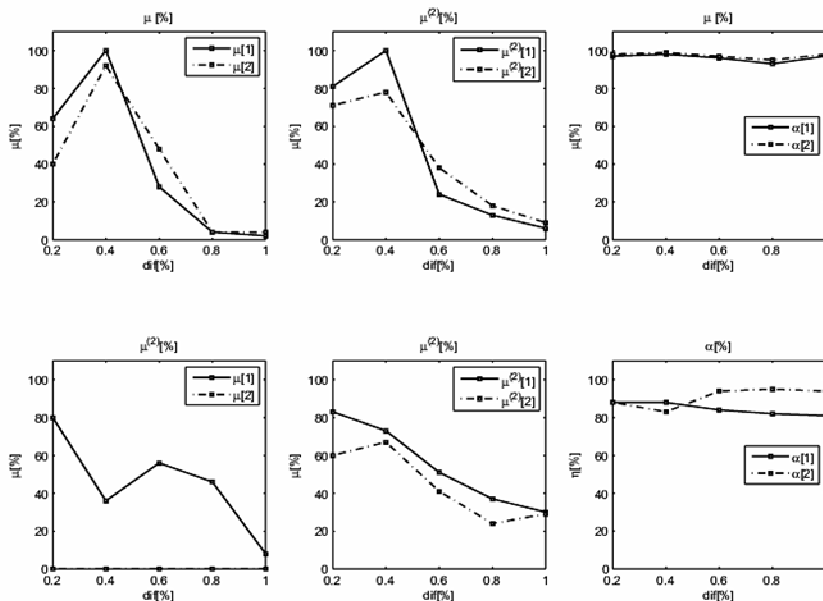
Závislost na velikosti vstupních množin. Cílem testu bylo ověření hodnot testovacích kritérií v závislosti na počtu bodů n vstupních množin. Byly testovány množiny bodů v intervalu $n \in \langle 10, 150 \rangle$ bodů, území bylo vymezeno okrajovými rovnoběžkami $\langle -60^\circ, 60^\circ \rangle$ a poledníky $\langle -120^\circ, 120^\circ \rangle$. Množiny vstupních dat mají různé distribuční charakteristiky: pravidelný rastr, semipravdělný rastr, náhodně vygenerované množiny bodů a množiny bodů představované agregovanými shluky (každý shluk byl tvořen 10 body). Pro množiny $n = 10$ nebylo možné určit z důvodu malého počtu Voroného buněk kritérium $\eta^{(2)}$. Efektivita detekce roste pro kritéria μ , $\mu^{(2)}$ s počtem bodů vstupních množin, pro kritérium $\eta^{(2)}$ je tato závislost méně výrazná, viz obr. 4. Výsledky detekce významněji nezávisí na distribuční charakteristice množiny. Za dostačující počet bodů lze pro účely detekce zobrazení považovat hodnoty kolem $n = 50$, přidávání dalších identických bodů již nepřináší výraznější efekt.

Závislost na velikosti analyzovaného území. Test analyzuje hodnoty detekčních kritérií v závislosti na velikosti analyzovaného území ve tvaru sférického čtyřúhelníku rozloženého symetricky kolem rovníku, viz obr. 5. Množiny P , Q jsou tvořeny $n = 50$ prvky. V každém výpočetním kroku se rozměry území zmenšily s využitím rekurze na polovinu, výchozí rozměry území jsou stejné jako v předchozím testu. Z množiny 50 bodů lze: pro hodnotu $\beta = 0.01$ poměrně spolehlivě detekovat kartografické zobrazení na území cca $10^\circ \times 10^\circ$, pro hodnotu $\beta = 0.0001$ detekovat kartografické zobrazení na území cca $1.5^\circ \times 1.5^\circ$.

Závislost na poloze analyzovaného území. Test ukazuje závislost úspěšnosti detekce kartografického zobrazení na poloze analyzovaného území na sféře. Poloha území ve tvaru sférického čtyřúhelníku se stranami 10° se měnila s krokem $\Delta\varphi = \Delta\lambda = 10^\circ$. Nižší úspěšnost detekce pro kritéria μ , $\mu^{(2)}$ byla dosažena v místech, kde mají kartografická zobrazení "podobné" tvary, tj. kolem obrazu rovníku a základního poledníku (nejhorší výsledky byly dosaženy v okolí průsečíku jejich obrazů), ve středních zeměpisných šířkách se efektivita detekce zvýšila, viz obr. 6. Kritérium $\eta^{(2)}$ vykazuje malou závislost na poloze analyzovaného území. Úspěšnost detekce v okrajových zeměpisných šířkách také závisela na tvaru obrazu pólu. Pokud obraz pólu představuje úsečku, efektivita detekce je vyšší (takových zobrazení je méně) než v případech, kdy je obrazem pólu bod či křivka.

Obr. 6 Závislost druhé nejvyšší hodnoty detekčních kritérií μ , $\mu^{(2)}$, $\eta^{(2)}$ na poloze analyzovaného území.

Vliv náhodné změny polohy prvků vstupní množiny. Tento test se snaží simulovat podmínky, které nastávají při reálné analýze starých map. Vstupní body jsou vzhledem k absenci geometrických i konstrukčních základů zatíženy polohovou chybou náhodného charakteru. Poloha bodů množiny P byla upravena zavedením náhodných posunů v intervalu 0.2% až 1% z hodnoty souřadnice x , y . Testované území mělo tvar sférického čtyřúhelníku se stranami 10°. Pro území ležící blízko rovníku bylo možné provádět spolehlivou detekci do 0.4 - 0.6% procentního náhodného posunu, pro území ve středních zeměpisných šířkách v intervalu 0.8 - 1% procentního náhodného posunu, viz obr. 7. Kritéria μ , $\mu^{(2)}$ dosáhla výrazně lepších výsledků než kritérium $\eta^{(2)}$ (které se pro tento účel ukázalo jako nespolehlivé), kritérium $\mu^{(2)}$ je poněkud citlivější na náhodné posuny bodů. Se zvětšující se hodnotou náhodné chyby klesá spolehlivost detekce.



Obrázek 7: Vliv náhodní změny polohy bodů na úspěšnost detekce, znázorněny první dvě nejvyšší hodnot testovacích kritérií μ , $\mu^{(2)}$, $\eta^{(2)}$. Nahore území rozložené symetricky podél rovníku, dole území nacházející se ve středních zeměpisných šířkách.

Detekce kartografického zobrazení u map bez geometricko-konstrukčního základu představuje poměrně obtížný problém, při jehož řešení řada detekčních kritérií nedosáhne uspokojivých výsledků. Nejnižší efektivitu detekce dosáhneme při kombinaci několika nevhodných parametrů vstupních množin: množiny pokrývající malá území s body zatíženými náhodnými chybami. Tato situace však velmi často nastává při analýze starých map. Celkově lze výše uvedená kritéria označit jako vhodná pro detekci kartografického zobrazení u map středních měřítek.

8 Závěr

Tento článek seznamuje s několika technikami automatizované detekce kartografického zobrazení z množiny identických bodů v mapě, využívá srovnání analyzované mapy s referenční mapou ve známém kartografickém zobrazení. Popsané techniky závisí na řadě parametrů, důležitou roli hraje zejména velikost analyzované množiny, její poloha, rozloha analyzovaného území či existence geometricko-konstrukčních základů v analyzované mapě. Ve srovnávacích testech dosáhla lepšího výsledku kritéria založená na aplikaci podobnostní transformace, ukázala se být citlivější pro malé množiny bodů a zároveň odolnější vůči náhodným chybám ve vstupních množinách. U technik založených na Voroného diagramech není nutné předem nastavovat citlivost detekce β , což lze považovat za výhodu, nevýhodou je menší odolnost vůči náhodným chybám.

Malá velikost analyzovaného území spojená s existencí náhodných polohových odchylek bodů vstupních množin ztěžují detekci kartografického zobrazení nad mapami vzniklými bez geometrického a konstrukčního základu. V takových případech se mohou výsledky analýz s využitím jednotlivých kritérií výrazněji lišit.

Na základě dosažených výsledků lze navrhnout metodiku detekce kartografického zobrazení založenou na dvou krocích. První krok představuje kombinovanou analýzu území s využitím více testovacích kritérií s následným nalezením potenciálně vhodných kandidátů (tj. zobrazení). Nad vybranými kartografickými zobrazeními je následně provedena detailnější detekce zahrnující i obecné polohy zobrazení. Jako nejvhodnější se jeví techniky detekce využívající aplikace podobnostní transformace. S využitím výše uvedených postupů se pokusíme u starých map českých z období 1518-1720 provést kartometrické analýzy zahrnující také detekci kartografického zobrazení. Výsledky přineseme v některém z dalších článků.

9 Poděkování

Článek vznikl za podpory grantu GAČR 205/07/0385 s názvem "Kartometrická a semiotická analýza a vizualizace starých map českých zemí z období 1518-1720".

Reference

1. Tian L. Kamata S. (2008): A Two-Stage Point Pattern Matching Algorithm Using Ellipse Fitting and Dual Hilbert Scans, *IEEE Trans.*
2. Ogawa H. (1986): Labeled point pattern matching by Delaunay Triangulation and Maximal Cliques, *Pattern Recognition*. Vol. 19. No. 1, pp. 35-40, 1986
3. Andrew M. Finch A., Wilson R., Hancock E. (1995): *Matching Delaunay Triangulations by Probabilistic Relaxation*, New York.
4. Bebis G., Deaconu T., Georgiopoulos M. (2000): *Fingerprint Identification Using Delaunay Triangulation*, University of Nevada.
5. Wamelen P., Li Z., Iyengar S. (2000): *A Fast Expected Time Algorithm for the Point Pattern Matching Problem*, Louisiana State University.
6. Okabe A., Boots B., Sugihara K., Cui S. N. (2000): *Spatial tessellations*, John Wiley & Sons.
7. Kuchař K.(1953): *Vývoj mapového zobrazení*, Praha.