

## Posudek na diplomovou práci Petry Vaničkové

### Analysis of meteorological images for storm detection in short-term weather forecast

Předložená diplomová práce se zabývá digitálním zpracováním tepelných snímků oblačnosti z geostacionárních družic za účelem vyhledávání útvarů označovaných v meteorologii jako tzv. „studené U“, resp. „studené V“. Ty často předznamenávají extrémní projevy počasí a jejich včasné rozpoznání by bylo možné využít ke zmírnění ničivých následků některých bouří. Metody pro automatické rozpoznávání těchto struktur však zatím nebyly popsány. Cílem práce byl tedy návrh algoritmu na detekování „studených U/V“ ze satelitních snímků pomocí metod zpracování obrazu a rozpoznávání a jeho implementace.

Práce obsahuje přehledný popis vzniku a vývinu „studených U/V“, jejich fyzikálních vlastností (tvar, rozložení teplot, persistence) v závislosti na oblasti výskytu a jevů vznikajících při družicovém snímkování. Tato kapitola je napsána přístupně i pro čtenáře, kteří se meteorologii nezabývají.

Při převodu na úlohu zpracování obrazu se autorka musela vyrovnat s nepřesnostmi meteorologické definice „studených U/V“, které se někdy projevují rozdíly v jejich detekci experty. Vzhledem k významu „studených U/V“ autorka správně upřednostňuje klasifikaci jiných struktur jako „studené U/V“ (statistická chyba 2. druhu – „false positives“) před nedetekováním „studených-U/V“ (statistická chyba 1. druhu – „false negatives“). Prvním krokem navrženého algoritmu je segmentace kandidátů „studených U/V“ – vyprahování ramen „studených U/V“ na základě teploty a úprava jejich tvaru pomocí binární morfologie. Důvodem pro rozšíření intervalu teplot [medián(TMIN), medián(TMAX)] v neadaptivní části prahování však zřejmě není pokles teploty mraků směrem k rovníku, neboť autorka se zabývá pouze výskyty U/V nad Evropou. Daný interval by totiž způsobil vynechání částí U/V, pro které je  $TMIN < \text{medián}(TMIN)$ , resp.  $TMAX > \text{medián}(TMAX)$ . Druhým krokem je stanovení parametrů vhodných pro klasifikaci vysegmentovaných kandidátů – výběr poměru hloubky prohlubně a šířky ramen a teplotního rozdílu  $TDIFF = TMAX - TMIN$  a zamítnutí orientace kvůli vytváření false negatives. Metody pro měření těchto parametrů i stanovení intervalu pro klasifikaci považuji za vhodně zvolené. Chybí zde však popis dosavadních výzkumů a aplikací na podobných úlohách a také přehled metod zpracování obrazu a rozpoznávání často používaných v meteorologii, který by mohl sloužit k porovnání s navrženými metodami.

Další část se zabývá testováním navrženého algoritmu. Vzhledem k malému množství dat na začátku vývoje algoritmu byla k učení použita pouze velmi malá trénovací množina (10 vzorů), což je vzhledem k variabilitě „studených U/V“ nedostatečné a vede k vysokému podílu špatných klasifikací. Otázkou je, proč nebyly na základě později získaných dat alespoň upraveny parametry algoritmu. V diskuzi, proč byl algoritmus zaměřen pouze na snímky Evropy, chybí údaje o „false negatives“, takže závěry nejsou dostatečně opodstatněné. Při pohledu na výsledky testů je zarážející nejen podíl „false positives“ (kolem 80%), ale i „false negatives“ (kolem 30%) při klasifikaci. Analýza příčin je správná, namísto přidávání dalších parametrů by asi bylo vhodnější zvětšení trénovací množiny a klasifikace pomocí statistického klasifikátoru.

Práce obsahuje řadu námětů na diskuzi a další výzkum, mj. předzpracování družicových snímků, použití sofistikovanější metody klasifikace, analýzu časových řad a výběr vhodnější trénovací množiny. Pro odstranění projektivní transformace by se ale místo přímého mapování a následné interpolace mělo použít zpětné mapování. Aplikace bayesovského klasifikátoru místo jednoduchého rozhodovacího stromu je velmi dobrá myšlenka, v textu by měl být jeho podrobnější popis a zdůvodnění pro prahování namísto úpravy apriorních pravděpodobností tříd. Analýza časových řad zřejmě povede ke zpřesnění klasifikace, otázkou je, zda by nebylo lepší spíše „uvolnit“ parametry (snížení podílu „false negatives“) a „zesílit“ podmínku na persistenci (snížení podílu „false positives“). Jinak vzhledem k četnosti snímkování (15 min) a průměrné době persistence „studených U/V“ (30 min) by posloupnost jednoho výskytu obsahovala obvykle jen 2-3 snímky. Pro účely velmi krátkodobé předpovědi počasí je navíc důležité detekovat „studené U/V“ co nejdříve. Tento fakt jen podtrhuje významnost jejich rozpoznávání na jednotlivých snímcích.

Součástí práce je i implementace navrženého algoritmu v C++ a příslušná aplikace pro detekci „studených U/V“ na reálných datech.

Po formální stránce je práce velmi dobrá, vyskytuje se zde jen několik menších chyb či nepřesností (např. u obr. 2.1 chybí odkaz na původní publikaci; text v obr. 2.4 je špatně čitelný; na str. 14 nemají být střední hodnoty TMIN a TMAX, ale jejich mediány; v popisku k obr. 3.5 má být místo mediánu střední hodnota). Grafická úprava je téměř dokonalá, stejně jako volba a popis obrázků. Dále oceňuji především souvislost výkladu, přehledné členění textu a srozumitelný jazyk. Práce je psána anglicky, a to i přes pár překlepů či drobných gramatických chyb (např. na str. 12 a 40 má místo „than“ být „then“) na velmi dobré úrovni.

Výše uvedené výtky kvalitu práce nijak podstatně nezpochybňují. Volba tématu je aktuální a adekvátní diplomové práci. Práce působí dobrým dojmem a je na ní vidět úsilí, které mu autorka věnovala. Oceňuji zvláště podrobnou diskusi výsledků a slibné návrhy na řešení problémů navrženého algoritmu.

Práci považuji za velmi dobrou, a protože její cíle byly splněny, doporučuji ji k obhájení.