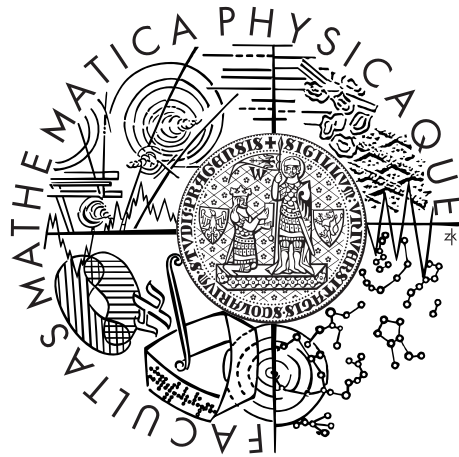


Univerzita Karlova v Praze
Matematicko-fyzikální fakulta

DIPLOMOVÁ PRÁCE



Jan Kotera

Analýza variačních metod pro segmentaci digitálního obrazu

Ústav teoretické fyziky

Vedoucí diplomové práce: RNDr. Barbara Zitová, Ph.D.

Studijní program: Fyzika

Studijní obor: Matematické a počítačové modelování ve fyzice a v technice

Praha 2011

Rád bych poděkoval vedoucí práce RNDr. Barbaře Zitové, Ph.D., za konzultace, zapůjčení literatury a veškerou pomoc při tvorbě práce, Ing. Filipu Šroubkovi, Ph.D., za konzultace a zapůjčení literatury a Mgr. Jiřímu Boldyšovi, Ph.D., za konzultace a četné připomínky k práci.

Prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně a výhradně s použitím citovaných pramenů, literatury a dalších odborných zdrojů.

Beru na vědomí, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 Sb., autorského zákona v platném znění, zejména skutečnost, že Univerzita Karlova v Praze má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle §60 odst. 1 autorského zákona.

V Praze dne 15.4.2011

Jan Kotera

Název práce: Analýza variačních metod pro segmentaci digitálního obrazu

Autor: Jan Kotera

Ústav: Ústav teoretické fyziky

Vedoucí diplomové práce: RNDr. Barbara Zitová, Ph.D., Ústav teorie informace a automatizace AV ČR

Abstrakt: Práce se zabývá teoretickým popisem Mumford–Shahova segmentačního modelu digitálního obrazu. Je probrána silná i slabá formulace problému, otázka existence a jednoznačnosti řešení a vlastnosti tohoto řešení. Dále je součástí práce numerická implementace konkrétní varianty tohoto modelu nazvané active contours without edges. Tato implementace je prakticky testována na několika snímcích, výsledky testování jsou podrobně popsány a teoreticky vysvětleny.

Klíčová slova: variační segmentace digitálního obrazu, Mumford–shahův segmentační model, active contours without edges.

Title: Variational methods for segmentation of digital images

Author: Jan Kotera

Department: Institute of Theoretical Physics

Supervisor: RNDr. Barbara Zitová, Ph.D., The Institute of Information Theory and Automation of the ASCR

Abstract: The text covers the theory of the Mumford and Shah model for digital image segmentation. The strong and weak formulation is presented and the questions of existence, uniqueness, and solution regularity is answered. Then, a particular variant of the model called ‘active contours without edges’ is numerically implemented. This implementation is tested on several images, the results are presented in detail and theoretically explained.

Keywords: variational image segmentation, Mumford and Shah image segmentation, active contours without edges.

Obsah

Úvod	1
1 Segmentace digitálního obrazu	3
Digitální obraz	3
Definice problému	3
2 Mumford–Shahův segmentační model	5
Formulace problému	5
Existence řešení	8
Diferenciální charakterizace minima	16
Regularita řešení	20
Jednoznačnost řešení	23
Asymptotické verze MS funkcionálu	25
3 Active Contours Without Edges	35
Level set metoda	35
Formulace problému	38
Existence řešení	40
Numerická implementace	41
4 Experimentální výsledky	47
Dílní testy	47
Shrnutí praktických poznatků	55
Závěr	58
Literatura	59
A Zdrojový kód funkce acwe	61

Úvod

Tato práce má dva hlavní cíle. V první řadě je to důkladné teoretické studium Mumford–Shahova segmentačního modelu, který lze považovat za fundamentální mezi variačními segmentačními modely, přičemž toto postavení je zároveň v práci obhájeno. Druhým cílem je vlastní implementace konkrétní varianty tohoto modelu a praktické testování této implementace. Po úvodní kapitole, ve které jsou ujasněny základní pojmy, následují dvě hlavní kapitoly a každá z nich je věnována splnění jednoho z vytyčených cílů.

Kapitola 2 má rešeršní charakter, snažil jsem se v ní důkladně avšak přehledně popsat většinu teorie týkající se Mumford–Shahova segmentačního modelu. Jsou v ní uvedeny a následně aplikovány teoretické poznatky jednak z článků, které se bezprostředně týkají Mumford–Shahova modelu a jiných problémů podobného typu (tzv. *free discontinuity problems*), tak z obecnějších disciplín, jako je funkcionální analýza a teorie parciálních diferenciálních rovnic. U tvrzení, která nejsou v práci dokázána, je vždy odkázáno na literaturu, kde je možné příslušný důkaz nalézt, a často jsem se v tom případě snažil alespoň popsat princip důkazu, aby se v něm čtenář lépe zorientoval. Tato kapitola tak poskytuje kvalitní souhrn teorie Mumford–Shahovy variační segmentace, aniž by byla potřebná další literatura.

Kapitola 3 je naopak orientována prakticky. Začíná opět výkladem potřebné teorie, přičemž jsem se snažil vše zasadit do kontextu předchozí kapitoly. Následuje vlastní numerická implementace konkrétního segmentačního modelu (*active contours without edges*, jedna z variant Mumford–Shahovy obecné segmentace). Způsob numerického řešení jsem podrobně vysvětlil. Tam, kde existuje více možností jak daný dílčí problém řešit, jsem svou volbu vždy odůvodnil a teoreticky vysvětlil. Domnívám se tak, že nejen konkrétní implementace, ale hlavně její důkladný popis tak může ušetřit čas komukoliv, kdo by ji chtěl modifikovat na svůj problém. Poté jsem implementaci testoval na několika zkušebních snímcích, zde jsem se snažil poukázat na správné i nesprávné chování metody a obojí opět teoreticky zdůvodnit. Chování metody je také několikrát porovnáno s jinými variačními segmentačními metodami založenými na odlišném principu. Zcela na závěr práce jsem shrnul obecné zkušenosti, které jsem při praktických testech s metodou získal, z této části je možné si rychle udělat přehled, na jaký typ problémů metoda je či není vhodná.

Použité značení

\mathbb{R}	Množina reálných čísel.
\mathbb{N}	Množina přirozených čísel.
$L^p(\Omega)$	Lebesgueův prostor funkcí integrovatelných v p -té mocnině.
$L^\infty(\Omega)$	Lebesgueův prostor funkcí esenciálně omezených.
$W^{k,p}(\Omega)$	Sobolevův prostor funkcí, jejichž k -tá derivace je v $L^p(\Omega)$.
$C^{0,s}(\Omega)$	Prostor funkcí hölderovsky spojitých, tzn. $ f(x) - f(y) \leq C x - y ^s$ pro nějaké C .
$C^{k,s}(\Omega)$	Prostor funkcí, jejichž k -tá derivace je hölderovsky spojitá s exponentem s .
$C^k(\Omega)$	Prostor funkcí k krát spojitě diferencovatelných.
$C_0^k(\Omega)$	Prostor funkcí k krát spojitě diferencovatelných s kompaktním nosičem uvnitř Ω .
$BV(\Omega)$	Prostor funkcí s konečnou variací.
$SBV(\Omega)$	Prostor speciálních funkcí s konečnou variací.
dx	Lebesgueova míra.
\mathcal{H}^k	Hausdorffova míra dimenze k .
$ \mu $	Totální variace míry μ .
$\nu \ll \mu$	Míra ν je absolutně spojitá vzhledem k μ .
$\nu \perp \mu$	Míra ν je singulární k μ .
$x_n \rightarrow x$	Silná konvergence (v normě).
$x_n \xrightarrow{w} x$	Slabá konvergence.
$\mathbf{1}_A$	Charakteristická funkce množiny A .
\overline{A}	Uzávěr množiny A .
$\overline{A^\Omega}$	Uzávěr množiny A vzhledem k Ω .
∇f	Gradient funkce f , $\nabla f = (\partial_x f, \partial_y f)$.
Δf	Laplaceův operátor funkce f , $\Delta f = \partial_{xx} f + \partial_{yy} f$.
$\operatorname{div} \mathbf{v}$	Divergence, $\operatorname{div} \mathbf{v} = \partial_x v_x + \partial_y v_y$.

Kapitola 1

Segmentace digitálního obrazu

Práce se zabývá studiem variačních metod pro segmentaci digitálního obrazu, v této úvodní kapitole si nejprve uvedeme tyto dva pojmy.

Digitální obraz

V praxi je digitální obraz obvykle diskrétní mřížka konečného počtu pixelů (matice), jejichž hodnoty světelné intenzity jsou také diskrétně rozděleny, například v intervalu $[0, 1]$ nebo jako celočíselné hodnoty 0 až 255. Pro teoretické úvahy s užitím matematické analýzy je však výhodné uvažovat definiční obor i obor hodnot obrázku jako kontinuum. Pod pojmem (šedotónový) digitální obrázek tak budeme rozumět jednoduše souvislou omezenou oblast $\Omega \subset \mathbb{R}^2$ a funkci $u : \Omega \rightarrow \mathbb{R}$ představující intenzitu obrazu (stupeň šedi) v daném bodě. V případě barevných obrazů se namísto skalární funkce uvažuje vektorová funkce $u : \Omega \rightarrow \mathbb{R}^3$, případně s jiným odpovídajícím počtem komponent.

Další požadavky na regularitu oblasti Ω a funkce u se stanovují podle konkrétní aplikace a modelu, stejně jako v jiných oblastech analýzy tvoří důležitý prvek pro korektnost úlohy. Obecně lze říci, že funkci u většinou nelze uvažovat spojitou, neboť právě nespojitosti funkce mohou představovat hranice objektů na obrázku, a tedy důležité prvky pro jeho analýzu. Naopak lze předpokládat, že hodnoty světelné intenzity jsou omezeny minimální a maximální naměřitelnou (či zobrazitelnou) hodnotou, a budeme tak pro vstupní obraz u_0 vždy předpokládat, že $u_0 \in L^\infty(\Omega)$, aniž by to představovalo omezení pro praktické aplikace.

Stojí za to zmínit, že výše popsaná reprezentace obrazu není jediná možná, například ve stochastických modelech se obrázek uvažuje jako pole náhodných veličin. Možností je mnoho a je třeba zvolit vhodnou reprezentaci podle matematického aparátu, jakým danou metodu studujeme.

Definice problému

Jednoduše řečeno, cílem segmentace je rozdělit vstupní obraz na jednotlivé logické celky, typicky objekty zájmu a pozadí. Každý pixel vstupního obrazu tak potřebujeme klasifikovat, zda náleží pozadí, nebo některému z objektů zájmu, případně různými hodnotami odlišit různé druhy objektů.

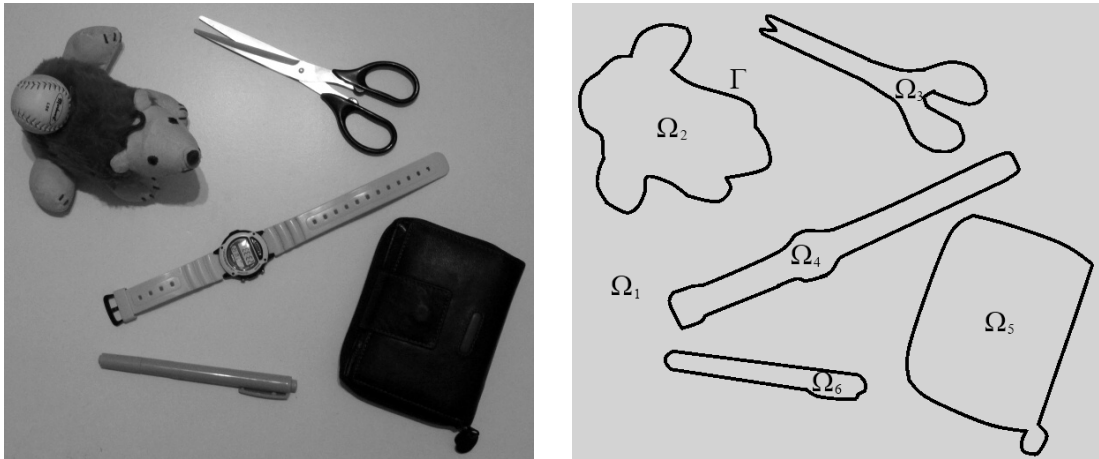
V kontextu reprezentace obrazu popsané výše tak můžeme segmentaci definovat jako proces

$$\Omega \xrightarrow{\text{segmentace}} \Omega = \bigcup_i^n \Omega_i \cup \Gamma$$

s vlastnostmi

1. Ω_i je neprázdná otevřená a souvislá $\forall i$, Γ je uzavřená v Ω s $\mathcal{H}^1(\Gamma) < \infty$,
2. $\Omega_i \cap \Omega_j = \emptyset \forall i \neq j$,
3. $P(\Omega_i) = \text{true} \forall i$,
4. $P(\Omega_i \cup \Omega_j) = \text{false} \forall i \neq j$,

kde množiny Ω_i tvoří jednotlivé objekty resp. pozadí, množina Γ představuje hranice objektů a funkce P je logický predikát, který stanovuje, zda daná část obrázku tvoří jeden vizuální celek (tedy buď jen objekt, nebo jen pozadí).



Obrázek 1.1: Příklad vstupního obrazu a očekávaného výstupu segmentace

Predikát P zde používáme jen ve vágním smyslu — jde o to nějak rozhodnout, zda je segmentace z lidského pohledu správná a přenést tuto informaci co nejvěrněji do matematického popisu nebo počítačového algoritmu. Například ve variačních metodách se namísto logického predikátu používá funkce určující míru nekvality segmentace (energie), jejíž minimum se snažíme nalézt. Konkrétní kritéria určující rozdělení obrazu nelze dostatečně dobře určit obecně, každá segmentační metoda tak typicky dobře funguje jen na určité třídě vstupních obrazů a je třeba tak přizpůsobit výběr metody pro danou aplikaci.

Klasické segmentační metody lze přibližně rozdělit na tzv. *edge-based* a *region-based* metody. Do první skupiny spadají metody, které se snaží detekovat objekty podle jejich okrajů, tzn. z výše popsaného procesu se zaměřují na detekci množiny Γ . Druhou skupinu tvoří k nim komplementární metody, které na základě přibližné homogenity intenzity (či jiného deskriptoru) seskupují pixely vstupního obrazu do jednotlivých objektů, snaží se tedy naopak určit jednotlivé Ω_i . Pro seznámení s klasickými segmentačními metodami lze doporučit například [19].

Kapitola 2

Mumford–Shahův segmentační model

V této kapitole se budeme podrobně zabývat jedním ze nejdůležitějších variačních segmentačních modelů, tzv. Mumford–Shahův model¹. Podle hrubého rozdělení z předchozí kapitoly bychom mohli MS model charakterizovat spíše jako *region-based* model, objekty jsou tak rozpoznány podle přibližné homogenity jejich vnitřku, nikoliv podle okrajů, uvidíme však v poslední části této kapitoly, že hrany objektů také hrají svou roli. Tento model je historicky významný a přesto stále aktuální, klade zajímavé matematické otázky a jeho modifikacemi lze odvodit modely vhodné pro širokou třídu aplikací. Někdy bývá dokonce označován za základní segmentační model v tom smyslu, že jiné segmentační metody lze formulovat jako speciální případy MS segmentace.

V první části se podíváme na výchozí představu, ze které model vychází, a uvedeme si přesnou formulaci segmentace jako minimalizační problém. Dále se budeme zabývat existencí řešení tohoto problému, zde bude potřeba nejprve uvést některé matematické pojmy. Ve třetí části se podíváme na diferenciální rovnice, které taková ideální segmentace (řešení minimalizace) musí splňovat, z nich lze vyčíst některé vlastnosti o segmentaci. Poté se podíváme na regularitu vzniklé segmentace a geometrii hran vysegmentovaných objektů, uvidíme například, že model je postaven tak, že z principu nemůže vysegmentovat hrany objektů zcela přesně. S dokázanou existencí řešení je přirozené se ptát, zda je řešení jednoznačné, na to se podíváme v další části (a uvidíme, že není). Nakonec se budeme věnovat jakýmsi limitním případům představeného segmentačního modelu, které jsou zajímavé jednak z toho důvodu, že dávají více nahlédnout „pod pokličku“ a ukazují, proč je tento model pro variační teorii segmentaci natolik důležitý, dále pak inspirují k modifikaci MS segmentačního modelu na jiné segmentační metody.

Text této kapitoly se klade za cíl čitelně a komplexně představit Mumford–Shahův segmentační model a ukázat, jakým teoretickým aparátem lze tento model zkoumat. V žádném případě si naopak neklade za cíl zde všechna tvrzení a uvedené poznatky precizně dokázat, neboť to by znamenalo shrnout stovky stránek jiných publikací či článků. U těžších tvrzení tak pouze odkážeme na literaturu, kde je odpovídající důkaz možné nalézt, případně naznačíme, jaký je jeho princip.

Formulace problému

Po částech hladká reprezentace obrazu

Model pochází z článku [25] z roku 1989 a vychází z představy obrázku jako zachycení 3D scény kamerou na 2D plochu. Jsou-li povrch a textura na povrchu jednotlivých

¹Dále bude převážně označován zkráceně jako MS model.

objektů hladké, jsou pak tyto objekty kamerou zobrazeny jako podmnožiny 2D obrazu, na kterých je funkce intenzity také hladká. V místě hranice objektů nebo překryvu jednoho objektu druhým jsou naopak v intenzitě prudké skoky, nespojitosti. Intenzitu obrazu tak můžeme očekávat jako po částech regulární funkci, jejíž definiční obor je možné (v souladu s první kapitolou) rozdělit na otevřené množiny objektů a uzavřenou množinu hran. Jak je ve zpracování signálu obvyklé, u vstupního obrazu u_0 je potřeba počítat s nejrůznějšími vadami, jako je například šum či poškození vzniklé při expozici, nemůžeme tak dopředu chtít, aby u_0 byl po částech regulární. Myšlenku MS segmentace lze tedy shrnout následovně: Nalézt optimální aproximaci vstupní funkce u_0 po částech hladkou funkcí u . Získáme tak segmentaci u_0 na jednotlivé souvislé části a rekonstrukci hladkého (nepoškozeného) signálu na těchto částech.

Vstupní představa fotografování po částech hladkých objektů není zdaleka dokonalá (a je na to poukázáno i v citovaném článku). Na obraze vznikají další nespojitosti vlivem neregularity povrchu objektů (roh nebo hrana krychle), neregularity textury na objektech (šachovnice), nerovnoměrného osvětlení (hrany stínů) a další. Mnoho segmentovaných obrazů naopak vůbec nepředstavují fotografie 3D makro scény, třeba snímky z mikroskopů, lékařské snímky, družicové snímky terénu, nebo dokonce audio signály, které sice nejsou „obrazy“ v běžném chápání, ale potřeba a princip segmentace je u nich podobný. Je tak třeba si ujasnit, jaké typy obrazů jsou vhodné pro MS segmentaci. Lze očekávat, že model může dobře fungovat na obrazech, které se skládají z rozumně mnoha objektů s přibližně homogenní intenzitou, mezi které patří i množství snímků z průmyslových či vědeckých aplikací, ačkoliv se vymykají původní představě fotografie makro objektů. Pro segmentaci na základě textury lze model snadno modifikovat a namísto intenzity pracovat s jiným vhodným deskriptorem (např. vybrané koeficienty waveletového rozkladu), snímky které jsou pak po částech homogenní v rámci tohoto deskriptoru lze segmentovat dobře. Naopak skutečné fotografie složité 3D scény (běžný snímek z dovolené) jsou typicky natolik komplikované, že „signál“ ploch objektů je přebitý „šumem“ v podobě obrovského množství hran mezi objekty a v rámci vnitřní struktury (textury) jednotlivých objektů.

Uveďme si nyní přesněji výchozí pojmy a formulaci Mamford–Shahovy segmentace.

Definice 2.1 *Nechť $\Omega \subset \mathbb{R}^2$ je oblast s lipschitzovskou hranicí. Řekneme, že*

$$\Omega = \bigcup_{i=1}^n \Omega_i \cup \Gamma, \quad n \in \mathbb{N} \quad (2.1)$$

je dělení obrazu (též segmentace obrazu) Ω , jestliže splňuje

- $\Omega_i \forall i$ a Γ jsou po dvou disjunktní,
- $\Omega_i \forall i$ je otevřená souvislá množina a,
- Γ je uzavřená v Ω s $\mathcal{H}^1(\Gamma) < \infty$.

Lze si velice snadno rozmyslet (viz [12]), že takové dělení splňuje (v topologii vzhledem k Ω)

- $\partial\Gamma = \Gamma$,
- $\partial\Omega_i \subset \Gamma \forall i$,
- $\Gamma = \bigcup_{i=1}^n \partial\Omega_i$,

tedy Γ tvoří úzkou hranici mezi objekty, nemá žádný vlastní vnitřek. Dělení obrazu tak splňuje představu o výsledku segmentace — obraz je beze zbytku rozdělen na konečně mnoho souvislých objektů, mezi kterými je ostrá hranice. Protože dělení je (až na očíslování) určené množinou Γ , budeme užívat pojmu „dělení určené (indukované) Γ “ bez upřesňování Ω_i .

Definice 2.2 Řekneme, že funkce (obraz) $u : \Omega \rightarrow \mathbb{R}$ je po částech hladká (konstantní) na Ω , jestliže existuje dělení Ω dané (2.1) tak, že $\forall i$ je $u|_{\Omega_i}$ hladká (konstantní).

Segmentace obrazu podle MS modelu tak znamená ze vstupního obrazu u_0 na Ω nalézt jeho aproximaci u (anglicky trefně nazvanou *cartoon representation*) a dělení (2.1) tak, že u je vzhledem k tomuto dělení po částech hladký. Takových aproximací a dělení je libovolně mnoho, potřebujeme tak definovat enegetický funkcional aproximace a hledat tu nejlepší (čím nižší energie, tím lepší kvalita aproximace a dělení). Je žádoucí zahrnout alespoň tyto požadavky:

- podobnost se vstupním obrázkem (přesnost aproximace),
- přibližná konstantnost na jednotlivých Ω_i (korektnost segmentace),
- míra dělení (omezení shora na počet vysegmentovaných částí obrazu).

Každý z těchto požadavků odpovídá jednomu členu v energetickém funkcionalu, který vzápětí uvedeme. Přesnost aproximace měříme jako L^2 normu rozdílu, regularitu aproximace jako $W^{1,2}$ seminormu na jednotlivých Ω_i a míru dělení jako délku hranice Γ , což zahrnuje požadavek na malý počet dělení i na malou klikatost hranic.

Silná formulace problému

Formulujme nyní (podle [25]) Mumford–Shahův segmentační problém.

Definice 2.3 (Silná formulace MS segmentace) Pro danou $\Omega \subset \mathbb{R}^2$ omezenou otevřenou s lipschitzovskou hranicí a $u_0 \in L^\infty(\Omega)$ najděme $(u, \Gamma) \in \mathcal{D}$ minimalizující $E(u, \Gamma)$ na této množině, kde

$$E(u, \Gamma) = \int_{\Omega} (u - u_0)^2 dx + \beta \int_{\Omega \setminus \Gamma} |\nabla u|^2 dx + \alpha \mathcal{H}^1(\Gamma), \quad (\text{MS})$$

$$\mathcal{D} = \left\{ (u, \Gamma); u \in W^{1,2}(\Omega \setminus \Gamma) \cap L^\infty(\Omega), \Gamma \subset \Omega \text{ uzavřená vzhledem k } \Omega, \mathcal{H}^1(\Gamma) < \infty \right\}.$$

Řekneme pak, že (u, Γ) je řešením MS segmentace.

Poznámka: Řekneme-li, že u resp. Γ je řešením minimalizace E , máme tím pak na mysli odpovídající prvek z (u, Γ) ve smyslu definice výše.

Ptáme se nyní, zda tento problém má řešení a jaké jsou jeho vlastnosti. Autoři původního článku vznesli následující domněnku:

Problém 2.3 má řešení pro každou u_0 spojitou. Navíc je-li (u, Γ) řešením, pak Γ je tvořeno konečně mnoha body spojenými konečně mnoha C^1 křivkami, které se potkávají nejvýše ve svých koncových bodech. Koncové body křivek (tedy jediné singularity Γ) mohou mít jen jednu z následujících podob: bod, ve kterém končí jediná křivka a žádná další z něho nepokračuje, nebo bod, ve kterém se potkávají právě tři křivky pod úhlem $2\pi/3$. Křivka může také vést až na $\partial\Omega^2$ a to jen kolmo k jejímu hladkému bodu.

²Ve smyslu $\text{dist}(\Gamma, \partial\Omega) = 0$.

Mnoho další práce na MS modelu bylo věnováno právě této domněnce, jednak existenci řešení a jednak regularitě řešení a geometrii Γ .

Podívejme se na skladbu funkcionálu E . První člen je „rekonstrukční“, který říká, že funkce u má dobře aproximovat u_0 až na případný šum či jiné poškození u_0 . Druhý člen je „regularizační“, který říká, že u (a tím i u_0) má být na jednotlivých oblastech Ω_i , které tvoří $\Omega \setminus \Gamma$, přibližně konstantní. Poslední člen je „škálový“, určuje míru segmentace ve smyslu počtu komponent či délky jejich hranice. Vynecháním kteréhokoliv členu dostaneme triviální řešení nebo posloupnost nezajímavých řešení, pro kterou $\inf E = 0$. Bez prvního členu je řešením konstanta, $\Gamma = \emptyset$, bez druhého členu je řešením $u = u_0$, $\Gamma = \emptyset$, bez posledního členu dosáhneme zhušťováním rovnoměrné sítě Γ_ε a vhodným po částech konstantním u_ε hodnoty $E(u_\varepsilon, \Gamma_\varepsilon) < \varepsilon$ pro libovolné $\varepsilon > 0$. Pouze se všemi třemi členy je tak funkcionál zajímavý a má smysl zkoumat jeho vlastnosti. Navzdory tomu se jednotlivé zmíněné verze (bez některého ze členů) kupodivu užívají a při vhodném omezení uvažovaných funkcí (abychom vyloučili triviální řešení) mají dobrý smysl. Budeme se jimi zabývat v poslední části této kapitoly.

Existence řešení

V této části se podíváme na problém existence řešení (u, Γ) splňující

$$E(u, \Gamma) \leq E(u', \Gamma') \quad \forall (u', \Gamma') \in \mathcal{D}.$$

Tato otázka nebyla adresována v původním článku [25], kde byl MS model navržen, ale získala mnoho pozornosti v dalším vývoji. Pokusíme se dobře osvětlit kostru a princip celého důkazu, v jednotlivých krocích odkážeme na literaturu pro konkrétní důkazy.

Existence je dokázána následujícím standardním postupem, který je znám jako přímá metoda variačního počtu. Shrňme ho pro názornost na obecném funkcionálu F . Mějme $F : V \rightarrow \mathbb{R}$ dáno, chceme nalézt $u \in V$ takové, že

$$F(u) = \inf_V F(v).$$

Existence takového u je dokázána v těchto krocích (viz např. [14]):

1. Uvažujme libovolnou minimalizující posloupnost $\{u_n\} \subset V$, tedy takovou, že $F(u_n) \rightarrow \inf F$. Potřebujeme pro další krok ukázat, že $\{u_n\}$ je omezená. Pokud by tomu tak nebylo, těžko bychom z prvků posloupnosti hledali kandidáty na řešení (lze si představit například minimalizaci $x \rightarrow \exp(x)$ v \mathbb{R}). Jednou z možností, jak zaručit omezenost $\{u_n\}$ je koercivita F , tedy vlastnost, že $F(u) \rightarrow \infty$ pro $\|u\| \rightarrow \infty$, pak je nutně $\|u_n\| < C \forall n$.
2. Chceme říct, že v $\{u_n\}$ existuje vybraná $\{u_{n_k}\}$ konvergující k nějakému $u \in V$. Takovou posloupnost většinou v silné topologii nenajdeme, je třeba si vystačit s tím, že $\{u_{n_k}\}$ bude konvergovat například pouze slabě. Proto pro další krok potřebujeme (jako jednu z možností), aby byl prostor V reflexivní, pak je známé, že z omezené $\{u_n\}$ lze vybrat $\{u_{n_k}\}$ (dále píšeme jen $\{u_n\}$) slabě konvergující k u ve V , píšeme $u_n \xrightarrow{w} u$. Jinou možností je kompaktní vnoření prostoru V do jiného prostoru W , pak opět z posloupnosti omezené ve V lze vybrat posloupnost konvergující v nějaké topologii ve W .
3. Chceme ukázat, že u je minimum F . Protože však F není ve většině případů slabě spojitý, nelze usoudit, že platí

$$u_n \xrightarrow{w} u \implies F(u_n) \rightarrow F(u).$$

Postačí však splnění slabší podmínky, funkcionál F musí být tzv. slabě zdola polospojité. To znamená, že pro všechny $u_n \xrightarrow{w} u$ je

$$\liminf_n F(u_n) \geq F(u)$$

(analogicky v jiné topologii). Je-li F slabě zdola polospojité, můžeme psát

$$\inf_V F = \lim_n F(u_n) \geq \liminf_n F(u_n) \geq F(u) \geq \inf_V F,$$

neboli $F(u) = \inf_V F$ a existence řešení minimalizace je tak dokázána.

Navržené kroky nejsou jedinou možností a jednotlivé podmínky lze různě nahrazovat, sledujeme však zhruba cestu, která vede k existenci minimalizace E . Shrňme tedy, že základními stavebními kameny jsou omezenost minimalizující posloupnosti, nějaká forma kompaktnosti použitého prostoru, ze které plyne konvergence minimalizující posloupnosti v nějaké topologii, a odpovídající zdola polospojité funkcionálu. Poznamenejme ještě, že nelze bez dalších kritérií říct nic o jednoznačnosti řešení minimalizace, zejména proto, že počáteční volba $\{u_n\}$ byla zcela libovolná, jednoznačností řešení se nyní však vůbec zabývat nebudeme.

Pokusme se tento postup aplikovat na E . Funkcionál má dva argumenty, u a Γ . Zobecnit výše uvedený postup z jednoho argumentu u na systém $\mathbf{u} = (u, \Gamma)$ je přímočaré, problém je však v rozdílném druhu jednotlivých argumentů, konkrétně tedy v charakteru Γ . Ukazuje se totiž, že je obtížné zvolit pro Γ vhodný prostor množin konečné \mathcal{H}^1 délky a topologii s rozumnými kompaktními vlastnostmi, aby přitom bylo možné dokázat odpovídající polospojité E . Namísto toho lze situaci vyřešit tak, jak je ve variačním počtu a oblasti parciálních rovnic běžné — šikovným způsobem oslabíme požadavky na u a Γ tak, aby bylo možné je uzavřít do prostoru s vhodnou strukturou, a namísto přímé (silné) formulace (MS) zavedeme slabou formulaci problému, pro kterou již budeme schopni ukázat existenci řešení. Následně je třeba ukázat, že toto slabé řešení je ve skutečnosti možné převést na silné řešení, čímž získáme silnou existenci. Jaký je tedy vhodný nový prostor pro u a Γ ? Protože je to právě Γ , které dělá problémy, je vhodné ho jako volný argument E zcela vyloučit a množinu nespojitostí u uvažovat přímo jakou součástí funkce samotné, kompaktnost a polospojité E , kterou je problematické ukázat pro Γ , bychom pak získali na vhodném prostoru funkcí. Máme $u \in W^{1,2}(\Omega \setminus \Gamma)$, očekáváme však, že funkce u bude přes Γ dělat skoky a nebude tedy $u \in W^{1,2}(\Omega)$. Je tedy potřeba místo $W^{1,2}(\Omega)$ uvažovat prostor, na kterém u může být nespojitá na nadrovině — množina těchto nespojitostí pak nahradí Γ . Vhodným kandidátem je prostor funkcí s konečnou variací (BV), konkrétně jeho podprostor prostor speciálních funkcí s konečnou variací (SBV). Provedme krátké shrnutí těchto prostorů, které mají pro MS segmentaci (a obecně ve zpracování obrazu) důležitý význam. Pro podrobný výklad teorie BV prostoru a důkazy uvedených tvrzení lze doporučit například [13].

Prostory BV a SBV

Uvažujme funkce, které jsou rozumné v tom smyslu, že „neoscilují moc“, neboli jejich totální variace

$$\int_{\Omega} |\nabla u|$$

je konečná. Pro jaké funkce dává tento integrál smysl? Zjevně stačí $u \in W^{1,1}(\Omega)$, to je však příliš restriktivní — chceme zahrnout i funkce u nespojitě na nadrovině nenulové délky, gradient Du takové funkce je třeba uvažovat jako míru, nikoliv jako funkci. Učiňme podobné zobecnění, jako v případě zavedení právě slabých derivací. Mějme

prozatím $u \in C^1$ a vektorovou testovací funkci $\varphi : \Omega \rightarrow \mathbb{R}^2$, $\varphi \in C_0^1(\Omega)$ (tzn. s nosičem uvnitř Ω) takovou, že pro její složky platí $|\varphi_{1,2}| \leq 1$. Pak je

$$\left| \int_{\Omega} u \operatorname{div} \varphi \right| = \left| \int_{\Omega} \varphi \cdot \nabla u \right| \leq \int_{\Omega} |\varphi| |\nabla u| \leq \int_{\Omega} |\nabla u|.$$

Na druhou stranu, vhodnou volbou φ pro dané u se lze levou stranou k hodnotě totální variace na pravé straně libovolně přiblížit. Výraz na levé straně má však nyní smysl pro $u \in L^1$, představuje tak možnost pro rozšíření pojmu totální variace.

Definice 2.4 *Nechť $u \in L^1(\Omega)$. Pak totální variaci funkce u definujeme*

$$\int_{\Omega} |Du| \stackrel{\text{def}}{=} \sup \left\{ \int_{\Omega} u \operatorname{div} \varphi; \varphi : \Omega \rightarrow \mathbb{R}^2 \in C_0^1(\Omega), |\varphi_{1,2}| \leq 1 \right\}. \quad (2.2)$$

Jako prostor $BV(\Omega)$ budeme uvažovat funkce, jejichž totální variace podle 2.4 je konečná.

Definice 2.5 *Nechť je $\Omega \subset \mathbb{R}^2$ omezená, otevřená. Prostor $BV(\Omega)$ funkcí s konečnou variací a normu na tomto prostoru definujeme*

$$BV(\Omega) = \left\{ u \in L^1(\Omega); \int_{\Omega} |Du| \leq \infty \right\}$$

$$\|u\|_{BV} = \|u\|_{L^1} + \int_{\Omega} |Du|.$$

Jak si lze totální variaci funkce u představit? Pro 1D funkce je to poměrně jasné, graf funkce narovnáme tak, aby vedl vždy „do kopce“ a změříme celkový přírůstek funkce. V případě 2D funkcí uvažujme graf funkce u jako krajinu, pak pro každé $\lambda \in \mathbb{R}$ nalezneme vrstevnice γ_{λ} s výškou λ , tedy $u(x) = \lambda \iff x \in \gamma_{\lambda}$ (pomiňme, že pro L^1 funkce nemá tato úvaha smysl, jde nám o intuitivní znázornění). Pak totální variace funkce u je „součet“ (integrace) délek všech vrstevnic vážený výškovým přírůstkem $d\lambda$

$$\int_{\Omega} |Du| = \int_{\mathbb{R}} \text{délka} \gamma_{\lambda} d\lambda,$$

tedy počítáme plochy, po kterých lze jít „do kopce“, vážené tím, jak moc jdou do kopce.

Vraťme se k našemu problému. Označme tedy Du derivaci funkce $u \in BV(\Omega)$, ukažme si stručně, že Du je skutečně míra. Nechť $u \in BV(\Omega)$, definujme funkcionál $L : C_0^1(\Omega)^2 \rightarrow \mathbb{R}$ jako

$$L(\varphi) = \int_{\Omega} u \operatorname{div} \varphi.$$

Pak lze snadno vidět, že L je lineární a spojitý. Aproximací hladkými funkcemi lze L rozšířit na spojitý lineární funkcionál $L : C_0(\Omega)^2 \rightarrow \mathbb{R}$. Z Riezsovy charakterizace duálu k prostoru spojitých funkcí je známo, že existuje vektorová Radonova míra ν taková, že

$$L(\varphi) = \int_{\Omega} \varphi d\nu$$

Můžeme rozepsat $\nu = -\sigma\mu$, kde μ je nezáporná míra a σ je vektorová funkce $|\sigma| \leq 1$ μ -skoro všude, pak máme

$$\begin{aligned} L(\varphi) &= \int_{\Omega} u \operatorname{div} \varphi = - \int_{\Omega} \nabla u \cdot \varphi \\ &= - \int_{\Omega} \varphi \cdot \sigma d\mu \end{aligned}$$

a tedy $Du = -\nu = \sigma\mu$ je hledaná vektorová Radonova míra. Lze navíc dále ukázat, že $|Du|$ jako totální variace míry je rovna $\int_{\Omega}|Du|$, totální variaci u , takže i názvosloví dává dobrý smysl.

Z rozkladu míry Du (později) uvidíme, že pro $u \in W^{1,1}$ je $Du = \nabla u \, dx$, tedy Du je zobecněním sobolevovské seminormy a tak $W^{1,1}(\Omega) \subset BV(\Omega)$. Prostor BV je tedy hledaným rozšířením $W^{1,1}$ na nespojité, leč stále diferencovatelné funkce. Podívejme se stručně na základní vlastnosti $BV(\Omega)$, částečně pro získání lepší představy o BV funkcích, částečně pro další použití. Veškeré podstatné informace a zde uvedená tvrzení lze nalézt v např. v [13].

Věta 2.6 *Nechť $\Omega \subset \mathbb{R}^2$ omezená, otevřená s lipschitzovskou hranicí. Pak platí následující tvrzení.*

1. (Úplnost) $BV(\Omega)$ s normou $\|\cdot\|_{BV}$ je Banachův.
2. (L^1 polospojitosť) Nechť $u_n \subset BV(\Omega)$ a $u_n \xrightarrow{L^1} u$, pak

$$\int_{\Omega}|Du| \leq \liminf_n \int_{\Omega}|Du_n|. \quad (2.3)$$

Důsledek: Je-li navíc $\{u_n\}$ omezená v $BV(\Omega)$, pak $u \in BV(\Omega)$.

3. (Stopa) Existuje lineární spojitý operátor stopy $\text{Tr} : BV(\Omega) \rightarrow L^1(\partial\Omega, \mathcal{H}^1)$.
4. (Aproximace) Pro každou $u \in BV(\Omega)$ existuje posloupnost $\{u_n\} \subset C^\infty(\Omega)$, že
 - $u_n \xrightarrow{L^1} u$,
 - $\int_{\Omega}|Du_n| \rightarrow \int_{\Omega}|Du|$ a
 - $u_n|_{\partial\Omega} = \text{Tr}(u) \forall n$.
5. (Kompaktnost) Nechť $\{u_n\} \subset BV(\Omega)$ je omezená v $BV(\Omega)$, pak existuje $u \in BV(\Omega)$ a vybraná posloupnost $u_{n_k} \xrightarrow{L^1} u$ (konvergující v L^1) a $Du_{n_k} \xrightarrow{M}^w Du$ (konvergující slabě* v prostoru měr, tzn. $\int_{\Omega} \varphi Du_{n_k} \rightarrow \int_{\Omega} \varphi Du \forall \varphi \in C_0(\Omega)^2$).

Důkaz (Částečný) Ukažme polospojitosť. Je

$$\int_{\Omega} u \operatorname{div} \varphi = \lim_n \underbrace{\int_{\Omega} u_n \operatorname{div} \varphi}_{\leq \int_{\Omega}|Du_n|} \leq \liminf_n \int_{\Omega}|Du_n|,$$

pro všechna φ z definice totální variace, stačí vzít na levé straně supremum. Poznamenejme, že lze snadno sestavit posloupnost s $\int|Du_n| \equiv \text{konst.} > 0$ přičemž $\int|Du| = 0$ (po částech konstantní funkce, u kterých roste počet skoků stejně, jako klesá jejich výška), takže nerovnost je nutná.

Dále ukažme úplnost, ta plyne z právě ukázané polospojivosti. Je-li $\{u_n\}$ cauchyovská v $BV(\Omega)$, je cauchyovská i v $L^1(\Omega)$, kde konverguje k $u \in L^1(\Omega)$. Z (2.3) plyne, že $u \in BV(\Omega)$. Zvolme pevné m a použijme opět polospojitosť (2.3) na posloupnost $\{u_m - u_n\}$ (konvergující v L^1 k $u_m - u$)

$$\int_{\Omega}|Du_m - Du| \leq \liminf \int_{\Omega}|Du_m - Du_n|.$$

Protože $\{u_n\}$ je cauchyovská, jde pravá strana do nuly a tedy $u_n \rightarrow u$ v $BV(\Omega)$.

Tvrzení o stopě ukazuje, že funkce $u \in BV(\Omega)$ nemůže oscilovat až k $\partial\Omega$, zvolíme-li systém kompaktních pomnožin $\Omega_1 \subset \Omega_2 \subset \dots$ takových, že $\Omega = \bigcup_i \Omega_i$, pak je $\int_{\Omega \setminus \Omega_i} |Du| \rightarrow 0$ a tedy totální variace u na menším a menším okolí hranice jde do nuly. Aproximace hladkými funkcemi umožňuje snadno přenášet známá tvrzení ze Sobolevových prostorů nebo prostorů spojitých funkcí na BV . Ukažme tímto způsobem poslední tvrzení (jen část o konvergenci v L^1).

Pro danou $\{u_n\}$ zvolme díky předchozímu tvrzení o aproximaci posloupnost $\{v_n\} \subset W^{1,1}(\Omega)$ tak, že

$$\int_{\Omega} |u_n - v_n| \rightarrow 0, \quad \int_{\Omega} |Dv_n| \leq \int_{\Omega} |Du_n| + 1.$$

Pak $\{v_n\}$ je omezená ve $W^{1,1}(\Omega)$ a protože $W^{1,1}(\Omega)$ je kompaktně vnořeno do $L^1(\Omega)$, máme (vybráním) $v_n \xrightarrow{L^1} u \in L^1(\Omega)$. Pak

$$\|u_n - u\|_{L^1} \leq \|u_n - v_n\|_{L^1} + \|v_n - u\|_{L^1} \rightarrow 0$$

a tedy původní posloupnost (vybraná se stejnými indexy) konverguje také. \square

Zbývá uvést poslední, ale velice důležitou věc týkající se BV funkcí — potřebujeme v gradientu funkce u rozpoznat nepojitosti této funkce, abychom je mohli ve slabé formulaci MS funkcionalu použít jako náhradu za množinu hran Γ . Věta o Lebesgueově rozkladu míry říká, že každou vektorovou míru ν na borelovských podmnožinách \mathbb{R}^2 lze napsat jako³

$$\nu = \nu_{ac} + \nu_s, \quad \text{kde } \frac{d\nu}{d\mu} = \frac{d\nu_{ac}}{d\mu}, \quad \nu_{ac} \ll \mu, \quad \nu_s \perp \mu,$$

kde μ je zvolená nezáporná míra. Použijme tuto větu na $\nu = Du$ a $\mu = dx$, dostaneme

$$Du = \nabla u \, dx + D_s u.$$

Je $\nabla u = \frac{dDu}{dx} \in L^1$ a $D_s u \perp dx$, derivace ∇u tak bývá označována jako přibližná derivace u , neboť se dx -skoro všude shoduje s derivací u ve smyslu $W^{1,1}$. O rozkladu Du lze ukázat ještě více ([2]), míru $D_s u$ lze dále rozložit na tzv. „skokovou“ část J_u a „cantorovskou“ část C_u , vysvětleme. Chceme popsat body, kde funkce u dělá podstatný skok, tedy kde je dx -nezanedbatelné množství bodů skokově větších než dx -nezanedbatelné množství jiných bodů. Zavedme z toho důvodu horní a dolní esenciální limitu funkce u .

Definice 2.7 *Nechť $u \in L^1_{\text{loc}}(\Omega)$. Řekneme, že u je v bodě $x \in \Omega$ esenciálně nejvýše λ (resp. alespoň λ), $\lambda \in \mathbb{R}$ (značíme $u(x) \stackrel{\text{ess}}{\leq} \lambda$ resp. $u(x) \stackrel{\text{ess}}{\geq} \lambda$), jestliže*

$$\lim_{r \rightarrow 0^+} \frac{|\{x \in \Omega \cap B(x, r); u(x) > \lambda\}|}{|B(x, r)|} = 0$$

resp.

$$\lim_{r \rightarrow 0^+} \frac{|\{x \in \Omega \cap B(x, r); u(x) < \lambda\}|}{|B(x, r)|} = 0,$$

kde $|\cdot|$ značí Lebesgueovu míru. Dále definujeme esenciální horní resp. dolní limitu (obálku) funkce u v bodě x značenou $u^+(x)$ resp. $u^-(x)$ vztahy

$$u^+(x) = \inf_{\lambda} \left\{ u(x) \stackrel{\text{ess}}{\leq} \lambda \right\},$$

$$u^-(x) = \sup_{\lambda} \left\{ u(x) \stackrel{\text{ess}}{\geq} \lambda \right\}.$$

³ $\nu_{ac} \ll \mu$ značí, že ν_{ac} je absolutně spojitá vzhledem k μ , a $\nu_s \perp \mu$, že ν_s je singulární k μ , $d\nu/d\mu$ je Radon–Nikodymova derivace míry.

První definice říká, že množství bodů, které danou podmínku porušují, musí být blízkém okolí dx -zanedbatelné. Druhá definice představuje nejmenší (resp. největší) funkci, která je větší (resp. menší) než u , představa je poměrně jasná. Poznamejme ještě, že zjevně $u^+(x) \geq u^-(x)$. Nyní, je-li $u \in L^1(\Omega)$, je skoro každý bod x Lebesgueův, ve kterém mj.

$$u(x) = \lim_{r \rightarrow 0^+} \int_{B(x,r)} u(y) dy,$$

pak je $u^+(x) = u^-(x) = u(x)$. Může však existovat dx -nulová množina bodů

$$S_u = \{x \in \Omega, u^-(x) < u^+(x)\}.$$

Je pozoruhodné, že pro $u \in BV(\Omega)$ je S_u (až na \mathcal{H}^1 -nulovou množinu) složena z částí spočetně mnoha hladkých křivek γ , jejichž Hasdorffova dimenze je 1, je tak možné \mathcal{H}^1 -skoro všude definovat normálu n k S_u . Na takové křivce γ je pak derivaci (skok) funkce u možné napsat jako

$$Du|_\gamma = (u^+ - u^-)n \mathcal{H}^1.$$

Celkově lze ukázat ([2]), že pro $u \in BV(\Omega)$ je možné derivaci Du je rozložit na

$$Du = \nabla u dx + \underbrace{(u^+ - u^-)n \mathcal{H}^1|_{S_u}}_{J_u} + C_u, \quad (2.4)$$

pro totální variaci $|Du|$ pak platí

$$|Du|(\Omega) = \int_\Omega |Du| = \int_\Omega |\nabla u| dx + \int_{S_u} (u^+ - u^-) n d\mathcal{H}^1 + |C_u|(\Omega). \quad (2.5)$$

Máme tak identifikované skoky funkce u (což jsou kandidáti na hrany Γ v MS segmentaci) jako množinu S_u . Zbývá nám podívat se na míru C_u . To je tzv. cantorovská část derivace Du , je $C_u \perp dx$ a lze ukázat, že $C_u(U) = 0$ pro všechny U takové, že $\mathcal{H}^1(U) < \infty$. Nosič míry C_u má tak Hausdorffovu dimenzi striktně mezi 1 a 2. Známým příkladem je v 1D tzv. Cantorova funkce, která je spojitá s dx -skoro všude nulovou derivací, přesto má totální variaci 1 na intervalu $(0, 1)$. Derivace takové funkce je právě C_u s nosičem na dx -nulové cantorovské množině.

Vraťme se nyní k původnímu otázce — zkonstruovat slabou formulaci MS problému. Je zřejmé, že prostor BV je (téměř) přesně to, co potřebujeme, množinu hran Γ nahradíme množinou skoků S_u , získáme tak funkcionál $F(u)$, u kterého již existence řešení vypadá slibně. Jediný problém, který však musíme odstranit, je právě C_u . Prostor BV obsahuje patologické funkce (více informací v [3]), jejichž derivace $Du = C_u$, je tedy $\nabla u \equiv 0$ a $S_u = \emptyset$. Pro takovou funkci je MS funkcionál

$$E(u, \Gamma = \overline{S_u^\Omega}) = \int_\Omega (u - u_0)^2 \geq \inf E(u, \Gamma).$$

Protože však takové funkce jsou v $L^2(\Omega)$ husté, můžeme jít s integrálem do nuly a dostaneme $\inf E(u, \Gamma) = 0$, přičemž žádné z těchto případných triviálních řešení nás nezajímá, neboť je dx -skoro všude konstantní a neobsahuje žádné hrany. Jinými slovy lze říci, že E není na BV koercivní, nekontroluje plně normu u , a tak hodnota $E(u, S_u)$ může být libovolně nízká, zatímco $\|u\|_{BV} \rightarrow \infty$. Nezbyvá, než po vzoru [16] takové funkce z definičního oboru MS funkcionálu vyloučit.

Definice 2.8 *Prostor speciálních funkcí s konečnou variací, SBV, definujeme*

$$SBV(\Omega) = \{u \in BV(\Omega); C_u \equiv 0\}.$$

Tento prostor (podprostor BV) je konečně ten správný, můžeme se tak vrátit k otázce existence řešení MS segmentace.

Slabá formulace

Idea je nyní zřejmá, místo $u \in W^{1,2}(\Omega)$ zvolíme $u \in SBV(\Omega)$ a množinu hran Γ nahradíme množinou skoků S_u , kompaktnost a polospojitosť, kterou se nepovedlo ukázat u množin, se pokusíme ukázat pro funkce. Napišme tedy slabou formulaci.

Definice 2.9 (Slabá formulace MS segmentace) Pro dané $u_0 \in L^\infty(\Omega)$ najděme $u \in SBV(\Omega)$ minimalizující $F(u)$ na této množině, kde

$$F(u) = \int_{\Omega} (u - u_0)^2 dx + \beta \int_{\Omega} |\nabla u|^2 dx + \alpha \mathcal{H}^1(S_u). \quad (\text{WMS})$$

Existenci slabého řešení problémů MS-typu se zabýval L. Ambrosio v roce 1989, syntézou jeho prací [2, 4] a aplikací výsledků relevantních pro MS funkcionál (WMS) můžeme zformulovat následující větu.

Věta 2.10 Necht' $\{u_n\} \subset SBV(\Omega)$ s $\|u_n\|_\infty \leq C$, která splňuje

$$\int_{\Omega} |\nabla u_n|^2 + \mathcal{H}^1(S_{u_n}) \leq C. \quad (2.6)$$

Pak existuje $\{u_{n_k}\}$ která ve smyslu míry (tzn. mj. pro s.v. x) konverguje k $u \in SBV(\Omega)$ s $\|u\|_\infty \leq C$, dále $\nabla u_{n_k} \xrightarrow[L^1]{w} \nabla u$. Navíc $\liminf_k \mathcal{H}^1(S_{u_{n_k}}) \geq \mathcal{H}^1(S_u)$.

Důkaz (Odkazem) Konvergence $u_{n_k} \rightarrow u$ (tzv. kompaktnost SBV) ve smyslu míry, ze které můžeme vybrat posloupnost konvergující dx -skoro všude, a slabá konvergence přibližné derivace je ukázána v [2], věta 2.1. Předpoklady jsou splněny díky (2.6), připomeňme, že pracujeme s Ω omezenou. Jiný důkaz je možné nalézt např. v [1], kde je mj. formulováno, že také skoková část derivace $J_{u_{n_k}}$ slabě* konverguje k J_u v prostoru měr.

Polospojitosť pro odpovídající konvergenci je v obecnější podobě ukázána v [4], věta 3.3. Předpoklady jsou pro Ω omezenou splněny díky jednoduchému tvaru integrálu přes S_u . \square

Pro doplnění polospojitosťi dále uvedme výsledek z [20], podle kterého je

$$\int_{\Omega} (u - u_0)^2 + |\nabla u|^2 dx \leq \liminf_k \int_{\Omega} (u_{n_k} - u_0)^2 + |\nabla u_{n_k}|^2 dx, \quad (2.7)$$

neboť integrand je spojitý, navíc konvexní v ∇u . Přístupme tak ke slabé existenční větě.

Věta 2.11 Necht' $\Omega \subset \mathbb{R}^2$ otevřená a omezená, $u_0 \in L^\infty(\Omega)$ dáno. Pak existuje aspoň jedno $u \in SBV(\Omega)$ s $\|u\|_\infty \leq \|u_0\|_\infty$ takové, že

$$F(u) = \inf_{SBV(\Omega)} F(v),$$

tedy u je řešení problému 2.9.

Důkaz Pro $u \equiv 0$ je $F(u) < \infty$ a tedy problém má smysl. Necht' $\{u_n\} \subset SBV(\Omega)$ ji minimalizující posloupnost F . Můžeme předpokládat, že $\|u_n\|_\infty \leq \|u_0\|_\infty$, protože položíme-li

$$\bar{u}_n(x) = \begin{cases} \|u_0\|_\infty & u_n(x) > \|u_0\|_\infty, \\ -\|u_0\|_\infty & u_n(x) < -\|u_0\|_\infty, \\ u_n(x) & \text{jinak,} \end{cases}$$

je $\bar{u}_n \in SBV(\Omega) \cap L^\infty(\Omega)$ a $F(\bar{u}_n) \leq F(u_n)$. Dále, protože $F(u_n) \leq C$, musí být nutně splněna podmínka (2.6). Věta 2.10 nám dává existenci limity $u \in SBV(\Omega) \cap L^\infty(\Omega)$, navíc spolu s (2.7) dostáváme polospojitosť $F(u) \leq \liminf F(u_n)$. Přímá metoda variačního počtu nastíněná na začátku této části nám dává existenci řešení. \square

Jinou možností jak dosáhnout kompaktnosti minimalizující posloupnosti je využít kompaktnost $BV(\Omega)$ (věta 2.6, část 5) pro $\{u_n\}$ omezenou v $BV(\Omega)$ a ukázat, že ve skutečnosti $u \in SBV(\Omega)$. Tento postup použil De Giorgi v [17].

Silná existence

Podobným způsobem jako v předchozí části zmapujeme důkaz silné existence MS problému 2.3. Označme hodnoty infim silné resp. slabé verze MS funkcionálu M_{strong} resp. M_{weak} . V [4] a [17] jsou jiným způsobem ukázána tvrzení, která lze shrnout následující větou.

Věta 2.12 *Nechť $\Gamma \subset \Omega$ uzavřená v Ω s $\mathcal{H}^1(\Gamma) < \infty$, nechť dále $u \in W^{1,2}(\Omega \setminus \Gamma) \cap L^\infty(\Omega)$. Pak $u \in SBV(\Omega)$ a $S_u \subset \Gamma \cup K$ kde $\mathcal{H}^1(K) = 0$.*

Každé případné silné řešení je tak kandidátem na slabé řešení, takže zjevně platí

$$M_{\text{weak}} \leq M_{\text{strong}}.$$

Mějme nyní \bar{u} minimalizující F . Pro minimalizaci $E(u, \Gamma)$ se nabízí vzít $u = \bar{u}|_{\Omega \setminus \Gamma}$ (pak zřejmě $u \in W^{1,2}(\Omega \setminus \Gamma) \cap L^\infty(\Omega)$) a $\Gamma = \overline{S_{\bar{u}}^\Omega}$ (uzávěr v Ω), čímž bychom dokázali opačnou nerovnost. Problém je v tom, že ačkoliv $\mathcal{H}^1(S_{\bar{u}}) < \infty$, není zaručeno, že $\mathcal{H}^1(\overline{S_{\bar{u}}^\Omega}) < \infty$ a je tak možné použít tuto množinu jako Γ . Podívejme se na protipříklad.

Mějme $\Omega = (0, 1)^2$, označme $\{x_n\} \subset \Omega$ posloupnost všech bodů s racionálními souřadnicemi v libovolném pořadí. Dále mějme klesající posloupnost čísel $\{r_n\} \subset (0, \infty)$, $n \in \mathbb{N}$ takovou, že $\sum_n r_n < \infty$. Označme C_n kružnici v bodě x_n o poloměru r_n a B_n kumulované sjednocení těchto kružnic

$$C_n = \{x \in \Omega; |x - x_n| = r_n\}$$

$$B_n = \bigcup_{i=1}^n C_i, \quad B = \lim_{n \rightarrow \infty} B_n.$$

Pak ze σ -aditivity míry máme

$$\mathcal{H}^1(B) = \lim_n \mathcal{H}^1(B_n) = \sum_n 2\pi r_i < \infty$$

(každé dvě kružnice se mohou protnout nejvýše ve dvou bodech, což je \mathcal{H}^1 -nulová množina). Protože B je zjevně v Ω hustá, máme $\overline{B^\Omega} = \Omega$ a tedy $\mathcal{H}^1(\overline{B^\Omega}) = \infty$.

Pro přechod od slabého řešení k silnému je tak třeba ještě ukázat regularitu $S_{\bar{u}}$ v tom smyslu, že hustota \mathcal{H}^1 míry musí být dostatečně lokalizovaná a rovnoměrně rozložená podél $S_{\bar{u}}$ a nemůže tak být příliš roztroušená v 2D oblasti. Můžeme uvést výsledek z [17], odkud tak pochází první důkaz silné existence.

Věta 2.13 ([17]) *Nechť $\bar{u} \in SBV(\Omega)$ je slabé řešení 2.9. Pak*

$$\mathcal{H}^1(\overline{S_{\bar{u}}^\Omega} \setminus S_{\bar{u}}) = 0.$$

Je tedy i každé slabé řešení kandidát na silné, dostáváme $M_{\text{strong}} = M_{\text{weak}}$ a existence silného řešení je dokázána.

Diferenciální charakterizace minima

Podívejme se na rovnice, jaké splňuje řešení (u, Γ) minimalizace E (Euler–Lagrangeovu rovnici funkcionálu). Standardně budeme postupovat fixováním jednoho argumentu (u resp. Γ) a výpočtem derivace podle druhého, abychom dostali podmínky pro jednotlivé proměnné. Vyslovme najednou výsledek této části v podobně následující věty. Důkaz podmínky na Γ provedeme pro názornost s předpokladem jednoduché geometrie segmentace, se stejnými argumenty lze však provést důkaz obecného případu.

Věta 2.14 *Nechť (u, Γ) je řešení minimalizace E , nechť Γ indukuje dělení Ω podle (2.1). Předpokládejme navíc, že $u \in C^1(\Omega_i)$ a Γ je složena z $C^{1,1}$ jednoduchých křivek, které se protínají nejvýše v koncových bodech. Pak je splněno (alespoň ve smyslu distribucí)*

$$u - \beta\Delta u = u_0 \text{ na každé } \Omega_i, \quad \frac{\partial u}{\partial \mathbf{n}} = 0 \text{ na } \partial\Omega_i, \quad (2.8)$$

$$e(u^L) - e(u^R) + \alpha \operatorname{curv} \gamma_i = 0 \text{ na každé } \gamma_i, \quad (2.9)$$

kde $e(u) = (u - u_0)^2 + \beta|\nabla u|^2$ je hustota energie segmentace, u^L resp. u^R jsou stopy u na levé resp. pravé straně γ_i vzhledem k libovolné parametrizaci a $\operatorname{curv} \gamma_i$ označuje křivost γ_i počítanou vzhledem ke stejné parametrizaci vztahem

$$\operatorname{curv} \gamma(p) = \frac{x'(p)y''(p) - y'(p)x''(p)}{(x'(p)^2 + y'(p)^2)^{\frac{3}{2}}}, \quad \text{pro } \gamma(p) = (x(p), y(p)). \quad (2.10)$$

Jak uvidíme v další části, předpoklady na regularitu u a Γ nejsou přehnané a ačkoliv nelze přímo říci, že jsou splněny pro každé řešení, lze s nimi do značné míry počítat.

Důkaz Ukažme rovnici (2.8) pro libovolnou Ω_i část dělení Ω . Vezměme libovolné $v \in W^{1,2}(\Omega_i)$ a $h \in \mathbb{R}$, pak dosazením dostáváme

$$E(u + hv) - E(u) = \int_{\Omega_i} 2hv(u - u_0) + 2h\beta\nabla u \cdot \nabla v + h^2v^2 + h^2\beta|\nabla v|^2 dx.$$

Vydělením h a přechodem k limitě $h \rightarrow 0$ dostáváme integrální rovnici

$$\int_{\Omega_i} (u - u_0)v dx + \beta \int_{\Omega_i} \nabla u \cdot \nabla v dx = 0 \quad \forall v \in W^{1,2}(\Omega_i) \quad (2.11)$$

Tato rovnice je již slabou verzí (2.8). Uvažujeme-li nyní řešení dostatečně hladké, dostaneme integrací per-partes ve druhém členu

$$\int_{\Omega_i} (u - u_0 - \beta\Delta u)v dx + \beta \int_{\partial\Omega_i} \frac{\partial u}{\partial \mathbf{n}} v = 0 \quad \forall v.$$

Vhodnou volbou v s nosičem uvnitř Ω_i v malém okolí libovolného bodu a standardní argumentací dostáváme rovnici (2.8), naopak volbou v nenulovou okolo libovolného bodu na $\partial\Omega_i$ dostaneme Neumannovu okrajovou podmínku v (2.8).

Důkaz (2.9) je delší, provedme ho pro názornost s jednoduchou geometrií Γ , ačkoliv stejným způsobem je možné postupovat v případě obecné geometrie. Plný důkaz je možné nalézt v [25].

Předpokládejme, že Γ je složeno z jediné jednoduché uzavřené C^2 křivky⁴ γ , což odpovídá situaci, že na obrázku je jediný objekt s hladkou hranicí. Oblast Ω je tak

⁴Pro $C^{1,1}$ křivku je argumentace stejná s tím, že křivost $\operatorname{curv} \gamma$ je nespojitá, existuje však skoro všude.

ve smyslu definice dělení 2.1 rozdělena na právě dvě podmnožiny — vnitřek křivky a vnějšek křivky γ , označme je Ω_I a Ω_E . Uvažujme nyní pohyb této křivky s (eulorovským) rychlostním polem $\mathbf{v}(x, t)$ (resp. pohyb bodů v Ω , ale zajímavý je pro nás jen na křivce γ), označme $\gamma(t), \Omega_I(t), \Omega_E(t)$ jednotlivé množiny v čase t . Protože řešení u je na jednotlivých Ω_I resp. Ω_E dáno jednoznačně (viz dále), můžeme označit jeho hodnotu $u_I(x, t)$ resp. $u_E(x, t)$. Nelze totiž uvažovat změnu Γ při u pevném, jak se to obvykle dělá, neboť u je závislé na dělení Ω daném Γ , při změně Γ není typicky ani $u \in W^{1,2}(\Omega \setminus \Gamma)$. Nakonec označme $E(t)$ hodnotu funkcionálu v čase t , tedy

$$E(t) = \int_{\Omega \setminus \gamma(t)} \left((u(x, t) - u_0)^2 + \beta |\nabla u(x, t)|^2 \right) dx + \alpha \int_{\gamma(t)} ds.$$

Rozdělením integrace na Ω_I a Ω_E můžeme psát

$$\begin{aligned} E(t) = & \int_{\Omega_I(t)} \left((u_I(x, t) - u_0)^2 + \beta |\nabla u_I(x, t)|^2 \right) dx \\ & + \int_{\Omega_E(t)} \left((u_E(x, t) - u_0)^2 + \beta |\nabla u_E(x, t)|^2 \right) dx + \alpha \int_{\gamma(t)} ds. \end{aligned}$$

Nyní nás zajímá $E'(t)$, což představuje derivaci $E(u, \Gamma)$ podle Γ . Protože integrační oblasti jsou závislé na čase t , udělejme malou vsuvku a uveďme dva výsledky z integrálního počtu.

Na derivaci prvních dvou členů se vztahuje tzv. věta o transportu (důkaz je poměrně technický, viz např. [15]), podle které

$$\frac{d}{dt} \int_{U(t)} F(x, t) dx = \int_{U(t)} \left(\frac{\partial F}{\partial t}(x, t) + \operatorname{div}(F\mathbf{v}) \right) dx = \int_{U(t)} \frac{\partial F}{\partial t}(x, t) dx + \int_{\partial U(t)} F(x, t) \mathbf{v} \cdot \mathbf{n}.$$

Pro derivaci druhého členu (což je změna délky křivky při změně jejího obrazu) platí

$$\frac{d}{dt} \int_{\gamma(t)} ds = - \int_{\gamma(t)} \operatorname{curv} \gamma(t) \mathbf{v} \cdot \mathbf{n} ds. \quad (2.12)$$

Odvození této identity odložíme na konec důkazu.

Pokračujme tak v důkazu úpravami $E'(t)$. Abychom započítali všechny povrchové členy se správným znaménkem, udělejme konvenci, že křivka je parametrizována proti směru hodinových ručiček (vzhledem k tomuto směru je počítána její křivost), její jednotkový normálový vektor \mathbf{n} směřuje dovnitř Ω_I . Aplikací výše uvedených vět dostáváme

$$\begin{aligned} E'(t) = & 2 \int_{\Omega_I(t)} (u_I - u_0) \frac{\partial u_I}{\partial t} dx - \int_{\gamma(t)} (u_I - u_0)^2 \mathbf{v} \cdot \mathbf{n} \\ & + 2\beta \int_{\Omega_I(t)} \nabla u_I \cdot \nabla \left(\frac{\partial u_I}{\partial t} \right) dx - \beta \int_{\gamma(t)} |\nabla u_I|^2 \mathbf{v} \cdot \mathbf{n} \\ & + 2 \int_{\Omega_E(t)} (u_E - u_0) \frac{\partial u_E}{\partial t} dx + \int_{\gamma(t)} (u_E - u_0)^2 \mathbf{v} \cdot \mathbf{n} \\ & + 2\beta \int_{\Omega_E(t)} \nabla u_E \cdot \nabla \left(\frac{\partial u_E}{\partial t} \right) dx + \beta \int_{\gamma(t)} |\nabla u_E|^2 \mathbf{v} \cdot \mathbf{n} \\ & - \alpha \int_{\gamma(t)} \operatorname{curv} \gamma(t) \mathbf{v} \cdot \mathbf{n} ds, \end{aligned}$$

kde jsme využili, že můžeme položit $\mathbf{v} = 0$ v okolí $\partial\Omega$, neboť nás zajímá jen pohyb křivky γ . Greenova věta dále dává

$$\int_{\Omega_I(t)} \nabla u_I \cdot \nabla \left(\frac{\partial u_I}{\partial t} \right) dx = - \int_{\Omega_I(t)} \Delta u_I \frac{\partial u_I}{\partial t} dx + \int_{\partial\Omega_I(t)} \frac{\partial u_I}{\partial t} \frac{\partial u_I}{\partial \mathbf{n}},$$

kde je možné využít rovnice a okrajové podmínky (2.8), čímž dostáváme

$$2\beta \int_{\Omega_I(t)} \nabla u_I \cdot \nabla \left(\frac{\partial u_I}{\partial t} \right) dx = -2 \int_{\Omega_I(t)} (u_I - u_0) \frac{\partial u_I}{\partial t} dx.$$

Totéž lze napsat pro u_E . Tyto členy již ve vztahu pro $E'(t)$ máme s opačným znaménkem, po vyrušení nám zbyde

$$\begin{aligned} E'(t) = & - \int_{\gamma(t)} \left((u_I - u_0)^2 + \beta |\nabla u_I|^2 \right) \mathbf{v} \cdot \mathbf{n} \\ & + \int_{\gamma(t)} \left((u_E - u_0)^2 + \beta |\nabla u_E|^2 \right) \mathbf{v} \cdot \mathbf{n} - \alpha \int_{\gamma(t)} \text{curv } \gamma(t) \mathbf{v} \cdot \mathbf{n}, \end{aligned}$$

s připomenutím definice $e(u)$ ze znění věty můžeme napsat

$$E'(t) = \int_{\gamma(t)} \left(-e(u_I) + e(u_E) - \alpha \text{curv } \gamma(t) \right) \mathbf{v} \cdot \mathbf{n}.$$

Protože $(u, \Gamma = \gamma)$ je řešení minimalizace, je nutně $E'(0) = 0$, a protože rychlostní pole $\mathbf{v}(x, t)$ je libovolné, dostáváme konečně

$$e(u_I) - e(u_E) + \alpha \text{curv } \gamma = 0 \quad \text{na } \gamma.$$

Zbývá ukázat vztah (2.12). Nejprve uveďme pár základních pojmů. Mějme parametrizaci hladké křivky $c(p) = (x_1(p), x_2(p))$, $p \in [a, b]$. Pro délku křivky platí

$$L = \int_a^b |c'(p)| dp,$$

namísto parametrizace p tak můžeme použít parametrizaci délkou od začátku parametrem $s \in [0, L]$, kde

$$s(p) = \int_0^p |c'(p)| dp,$$

je pak

$$\frac{\partial}{\partial s} = \frac{1}{|c'(p)|} \frac{\partial}{\partial p}, \quad \text{formálně } ds = |c'(p)| dp.$$

Dále označme \mathbf{t} jednotkový tečný vektor a \mathbf{n} jednotkový normálový vektor dané vztahy

$$\begin{aligned} \mathbf{t} = c'(s) &= \frac{c'(p)}{|c'(p)|}, \\ \mathbf{N} &= (-x_2'(p), x_1'(p)), \quad \mathbf{n} = \frac{\mathbf{N}}{|\mathbf{N}|}. \end{aligned}$$

Takto definovaný normálový vektor směřuje vlevo vzhledem k obíhání křivky (například při parametrizaci kruhu proti směru hodinových ručiček míří \mathbf{n} do středu kruhu). Nakonec označme \mathbf{c} vektor křivosti, který představuje změnu tečného vektoru, tedy

$$\mathbf{c} = \frac{\partial \mathbf{t}}{\partial s} = \frac{\partial^2 c}{\partial s^2} = \frac{1}{|c'(p)|} \frac{\partial}{\partial p} \left(\frac{c'(p)}{|c'(p)|} \right).$$

Lze ukázat (a intuitivně je docela dobře představitelné), že vektor křivosti je kolineární s normálovým vektorem. Konstantu úměrnosti označme $k(s)$ resp. $k(p)$ a vztah zapišme

$$\mathbf{c} = k\mathbf{n} \quad \sim \quad \frac{1}{|c'(p)|} \frac{\partial}{\partial p} \left(\frac{c'(p)}{|c'(p)|} \right) = k(p)\mathbf{n}(p). \quad (2.13)$$

Konstanta úměrnosti k se nazývá křivost, v důkazu výše pro názornost označována $\text{curv } c$. Takto definovaná křivost je tedy kladná, pokud křivka „zatačí doleva“ a záporná, pokud křivka „zatačí doprava“ (při parametrizaci kruhu pro směr hodinových ručiček je $k(p) > 0$, proti směru hodinových ručiček je $k(p) < 0$).

Nyní můžeme přistoupit k odvození (2.12). Uvažujme v souladu s geometrií použitou v důkazu uzavřenou křivku, tedy $c(a) = c(b)$. Aby se normálový vektor této křivky svou orientací shodoval s konvencí použitou v důkazu, je třeba křivku parametrizovat proti směru hodinových ručiček. Uvažujme časovou změnu délky křivky (pro větší přehlednost budeme skalární součin zapisovat jako $\mathbf{a} \cdot \mathbf{b} = \langle \mathbf{a}, \mathbf{b} \rangle$)

$$\begin{aligned} L'(t) &= \frac{d}{dt} \int_c ds = \frac{d}{dt} \int_a^b |c'(p)| dp = \int_a^b \frac{d}{dt} |c'(p)| dp \\ &= \int_a^b \frac{d}{dt} \sqrt{\langle c'(p), c'(p) \rangle} dp = \int_a^b \frac{\left\langle \frac{\partial c}{\partial p}, \frac{\partial^2 c}{\partial t \partial p} \right\rangle}{|c'(p)|} dp. \end{aligned}$$

Derivaci a integrál je možno bez problému přehodit pro $c \in C^2[a, b]$. Nyní provedeme per-partes pro přehození derivace podle p , dostaneme (připomeňme, že $c(a) = c(b)$, takže vypadne povrchový člen)

$$L'(t) = - \int_a^b \left\langle \frac{\partial}{\partial p} \left(\frac{c'(p)}{|c'(p)|} \right), \frac{\partial c}{\partial t} \right\rangle dp = - \int_a^b |c'(p)| \left\langle \frac{1}{|c'(p)|} \frac{\partial}{\partial p} \left(\frac{c'(p)}{|c'(p)|} \right), \frac{\partial c}{\partial t} \right\rangle dp.$$

Dosadíme za první člen součinu ze vztahu (2.13), dále si uvědomíme, že $\frac{\partial c}{\partial t}$ je rychlostní pole \mathbf{v} , dostáváme tak hledaný vztah

$$\frac{d}{dt} \int_c ds = - \int_a^b \langle k(p) \mathbf{n}, \mathbf{v} \rangle |c'(p)| dp = - \int_c \text{curv } c(s) \mathbf{v} \cdot \mathbf{n} ds.$$

□

Jako důsledek věty 2.14 můžeme uvést následující větu.

Věta 2.15 *Nechť Γ a jím určené dělení oblasti Ω podle (2.1) je pevně dáno. Pak pro každé $u_0 \in L^\infty(\Omega)$ ⁵ existuje právě jedno $u \in W^{1,2}(\Omega_i) \forall i$ řešení minimalizace $E(u, \Gamma)$.*

Důkaz Z první části důkazu předchozí věty plyne (rovnice (2.11)), že na každé Ω_i je u řešením eliptického problému

$$a(u, v) = f(v) \quad \forall v \in W^{1,2}(\Omega_i),$$

kde $a : W^{1,2}(\Omega_i) \times W^{1,2}(\Omega_i) \rightarrow \mathbb{R}$ a $f : W^{1,2}(\Omega_i) \rightarrow \mathbb{R}$ jsou bilineární resp. lineární formy dané vztahy

$$\begin{aligned} a(v, w) &= \int_{\Omega_i} (vw + \beta \nabla v \cdot \nabla w) dx, \\ f(v) &= \int_{\Omega_i} u_0 v dx. \end{aligned}$$

Je zřejmé, že pro $\beta > 0$ je forma a spojitá a eliptická (navíc dokonce symetrická) a f je spojitá na $W^{1,2}(\Omega)$ (je $u_0 \in L^\infty(\Omega)$ a tedy jeho restrikce je z $L^\infty(\Omega_i)$), z Lax–Milgramovy věty (např. [14]) dostáváme ihned existenci právě jediného řešení $u \in W^{1,2}(\Omega_i)$. □

⁵Stačí zjevně $u_0 \in L^2(\Omega)$.

Regularita řešení

V této části se podíváme, jakou hladkost a jiné vlastnosti je možné očekávat od minima MS funkcionálu E , tedy funkce u a množiny hran Γ . Situace je zde pikantní v tom, že pro odhad globální regularity funkce u je potřeba mít zaručenou dostatečnou hladkost hranice, tedy v tomto případě části Γ , naopak pro charakterizace regularity a geometrie Γ je potřeba uvažovat funkci u na okolí Γ a mít zaručenou určitou její regularitu, mj. omezenost funkce a jejích derivací.

Otázka regularity u byla dobře adresována v původním článku [25], kde jsou aplikovány obecné výsledky z [18]. Naopak regularita Γ v původním článku analyzována nebyla, ale s předpokladem dostatečné regularity (přibližně — Γ se musí skládat z konečně mnoha hladkých křivek) byly ukázány zajímavé důsledky pro geometrii hran. V dalších letech se touto otázkou zabývalo několik matematiků (za významnou můžeme považovat například práci A. Bonneta [6], kde je dosaženo podobných výsledků se slabšími předpoklady). Lze říci, že plná otázka regularity a geometrie Γ bez apriorních předpokladů (tzn. čistě z faktu, že $u \in W^{1,2}$ a $\mathcal{H}^1(\Gamma) < \infty$ je minimum E) nebyla dosud plně uspokojivě vyřešena.

Regularita funkce u

Podívejme se nejprve, co lze říci o funkci u restringované na jednotlivé souvislé oblasti Ω_i (v této části budeme psát jen Ω a jsou tím myšleny všechny Ω_i ze vzniklé segmentace, neboť celá oblast Ω v původním chápání zde nehraje roli). Víme (viz první část důkazu věty 2.14 a větu 2.15), že $u \in W^{1,2}(\Omega)$ je řešením rovnice

$$u - \beta \Delta u = u_0 \text{ na } \Omega, \quad \frac{\partial u}{\partial \mathbf{n}} = 0 \text{ na } \partial \Omega, \quad (2.14)$$

resp. slabé verze

$$\int_{\Omega} (uv + \beta \nabla u \cdot \nabla v) dx = \int_{\Omega} u_0 v dx \quad \forall v \in W^{1,2}(\Omega). \quad (2.15)$$

Pro řešení eliptických problémů na hladkých oblastech je znám odhad regularity řešení, uveďme jej ve formě věty (viz např. [18]).

Věta 2.16 *Nechť $\Omega \subset \mathbb{R}^2$ omezená otevřená, $\partial \Omega \in C^{1,1}$ a u je řešení problému (2.15) s $u_0 \in L^\infty(\Omega)$. Pak $u \in W^{2,p}(\Omega)$, $\forall p < \infty$. Důsledek — $u \in C^{1,s}(\Omega)$, $s < 1$ ze Sobolevových nerovností.*

Z tohoto faktu můžeme vyjít pro analýzu regularity funkce u na nehladkých oblastech Ω_i které mohou být řešením minimalizace $E(u, \Gamma)$. Výsledky shrňme přehledně do následující věty.

Věta 2.17 *Nechť je $\Omega \subset \mathbb{R}^2$ otevřená omezená, u je řešením (2.15). Pak platí následující tvrzení.*

1. $u \in W_{loc}^{2,p}(\Omega)$, $\forall p < \infty$ (tzn. $u \in W^{2,p}(U)$ pro $\bar{U} \subset \Omega$), speciálně $u \in C^{1,s}(U)$, $s < 1$. Rovnici (2.14) je splněna skoro všude v Ω .
2. Pro skoro všechna $x \in \Omega$ je $\min_{\Omega} u_0 \leq u(x) \leq \max_{\Omega} u_0$, speciálně $u \in L^\infty(\Omega)$, $\|u\|_{L^\infty} \leq \|u_0\|_{L^\infty}$ (minimum a maximum lze uvažovat esenciálně).
3. Je-li P bod na hladké části $\partial \Omega$, neboli pro nějaké jeho okolí U platí $\partial \Omega \cap U \in C^{1,1}$, pak $u \in C^1(\bar{\Omega} \cap U)$ a $\frac{\partial u}{\partial \mathbf{n}}(P) = 0$.

4. Necht' je P bod konvexního rohu na hranici, neboli $\partial\Omega$ se lokálně skládá ze dvou $C^{1,1}$ křivek které se spojují v bodě P , přičemž pro vnitřní úhel α mezi jejich tečnami v bodě P platí $0 < \alpha < \pi$. Pak $u \in C^1(\overline{\Omega} \cap U)$ pro nějaké okolí U bodu P a $\nabla u(P) = 0$.

5. Necht' P je bod nekonvexního rohu na hranici, neboli pro úhel α z předchozí položky platí $\pi < \alpha \leq 2\pi$. Toto se týká i situace, kdy bod P je na konci jediné $C^{1,1}$ křivky, která v něm končí. Pak u má v polárních souřadnicích (r, θ) se středem v bodě P tvar

$$u(r, \theta) = cr^{\frac{\pi}{\alpha}} \sin\left(\frac{\pi}{\alpha}(\theta - \theta_0)\right) + \hat{u}(r, \theta),$$

kde $\hat{u} \in C^1$ a c, θ_0 jsou vhodné konstanty. Speciálně, $\frac{\partial u}{\partial r} \rightarrow \infty$ pro $r \rightarrow 0+$.

Tvrzení věty není zcela vyčerpávající, nezahrnuje například situaci, kdy křivka horší než $C^{1,1}$ končí v bodu P (což je možné řešení MS segmentace, [25] Appendix A.I částečně adresuje tento případ).

Důkaz (Částečný, pro obecné věty týkající se regularity viz [18], konkrétní aplikace na MS problém je v [25], Appendix A). Ukažme tvrzení 1. Oblast Ω můžeme vložit do větší oblasti Ω_R s hladkou hranicí, na této oblasti řešíme (2.15), označme u_R řešení. Pak podle 2.16 je $u_R \in W^{2,p}(\Omega_R)$, $p < \infty$. Na Ω je však

$$\int_{\Omega} \varphi(u - u_R) + \beta \nabla \varphi \nabla (u - u_R) = 0 \quad \forall \varphi \in W_0^{1,2}(\Omega),$$

(neboli $(u - u_R) - \beta \Delta(u - u_R) = 0$). Je známo, že řešení $u - u_R$ takového problému (mj. vlastní funkce operátoru Δ) je třídy C^∞ na okolí každého bodu z Ω (např. [14], 6.3.1 Theorem 3). Mějme nyní množinu U takovou, že $\overline{U} \subset \Omega$, pak je $u - u_R \in C^\infty(\overline{U})$, $u_R \in W^{2,p}(U)$ a tedy nutně i $u \in W^{2,p}(U)$, z čehož dále plyne $u \in C^{1,s}(U)$ ze Sobolevových nerovností. Z (2.15) a faktu, že $\Delta u \in L^p$ máme Greenovou větou, že (2.14) je splněno skoro všude.

Tvrzení 2 jsme již v jiné podobě využili v části o existenci řešení, jedná se o jednoduché pozorování, pojďme si ho však ukázat. Funkce u , která řeší (2.15), zároveň minimalizuje funkcional

$$G(u) = \int_{\Omega} \left((u - u_0)^2 + \beta |\nabla u|^2 \right) dx,$$

to je vidět z věty 2.14 kde je rovnice (2.15) odvozena. Mějme řešení u , položme

$$\tilde{u}(x) = \begin{cases} u(x) & \min_{\Omega} u_0 \leq u(x) \leq \max_{\Omega} u_0, \\ \min_{\Omega} u_0 & u(x) < \min_{\Omega} u_0, \\ \max_{\Omega} u_0 & u(x) > \max_{\Omega} u_0 \end{cases}$$

(minimum a maximum lze brát esenciální). Funkce \tilde{u} pak splňuje tvrzení věty. Je však skoro všude buď $\tilde{u}(x) = u(x)$ nebo $\nabla \tilde{u}(x) = 0$, takže

$$\int_{\Omega} |\nabla \tilde{u}|^2 \leq \int_{\Omega} |\nabla u|^2.$$

Taktéž je skoro všude $|\tilde{u} - u_0| \leq |u - u_0|$, musí tedy nutně být $G(\tilde{u}) \leq G(u)$, což znamená že $\tilde{u} = u$ z jednoznačnosti minima.

Podívejme se, jak lze ukázat tvrzení 3. Mějme P jak bylo určeno, označme $U_P \subset \Omega$ otevřenou množinu s $C^{1,1}$ hranicí, jejíž hranice je shodná s hranicí Ω na okolí bodu P . Chceme ukázat, že $u \in C^1$ na této množině, k tomu využijeme výsledek z [18],

Proposition 2.5.2.3 (stručně: jde o analogii 2.16, je-li u řešením eliptické rovnice s pravou stranou v L^p a splňuje v nějakém slabším smyslu Neumannovu okrajovou podmínku, pak $u \in W^{2,p}$). Protože však u okrajovou podmínku (na ∂U_P která není společná s $\partial\Omega$) nesplňuje, přenásobíme ji vhodnou funkcí, aby tak byla okrajová podmínka splněna. Mějme pro tento účel hladkou funkci φ takovou, že $\varphi = 1$ na okolí bodu P , $\varphi = 0$ na $\partial U_P \setminus (\partial\Omega \cap \partial U_P)$ a $\partial\varphi/\partial\mathbf{n} = 0$ na ∂U_P . Položme $\bar{u} = \varphi u$, pak platí (rozderivováním)

$$\beta\Delta\bar{u} - \bar{u} = \beta u\Delta\varphi + 2\beta\nabla u \cdot \nabla\varphi - u_0\varphi \in L^2(U_P),$$

neboť $u \in W^{1,2}(U_P)$, tedy $\nabla u \in L^2(U_P)$. Navíc \bar{u} splňuje Neumannovu podmínku na ∂U_P ve smyslu

$$\int_{U_P} \psi\Delta\bar{u} = \int_{U_P} \bar{u}\Delta\psi \quad \forall \psi \in W^{2,2}(U_P), \quad \frac{\partial\psi}{\partial\mathbf{n}} = 0 \text{ na } \partial U_P.$$

Ukažme.

$$\int_{U_P} \bar{u}\Delta\psi = - \int_{U_P} \nabla\bar{u} \cdot \nabla\psi = - \int_{\partial U_P} \psi \frac{\partial\bar{u}}{\partial\mathbf{n}} + \int_{U_P} \psi\Delta\bar{u}.$$

Chceme ukázat, že povrchový integrál na pravé straně je nulový. Rozdělme integraci na společnou a samostatnou část hranice ∂U_P vzhledem k $\partial\Omega$. Pak máme (rozderivováním)

$$\int_{\partial U_P} \psi \frac{\partial\bar{u}}{\partial\mathbf{n}} = \int_{\partial U_P \cap \partial\Omega} \psi \left(\frac{\partial u}{\partial\mathbf{n}}\varphi + u \frac{\partial\varphi}{\partial\mathbf{n}} \right) + \int_{\partial U_P \setminus \partial\Omega} \psi \left(\frac{\partial u}{\partial\mathbf{n}}\varphi + u \frac{\partial\varphi}{\partial\mathbf{n}} \right),$$

kde poslední 3 členy vypadnou, protože vždy φ nebo jeho derivace jsou nulové, a první člen vypadne proto, že u je řešením (2.15). Z [18] pak dostáváme $\bar{u} \in W^{2,2}(U_P)$. Protože v tom případě $\nabla\bar{u} = \nabla u\varphi + u\nabla\varphi \in W^{1,2}(U_P) \subset L^p(U_P)$, $p < \infty$, musí být $\nabla u \in L^p(U_P)$. Celý proces tak můžeme zopakovat s tím, že nyní $\Delta\bar{u} - \bar{u} \in L^p(U_P)$, takže [18] dává $\bar{u} \in W^{2,p}(U_P)$. Protože na okolí P je $\bar{u} = u$, je tedy nutně $u \in W^{2,p}$, ze Sobolevových nerovností dále $u \in C^1(\bar{\Omega})$ (restringováno na okolí P) a Neumannova podmínka je tím splněna silně.

Pro důkazy zbývajících tvrzení odkážeme na [25], Appendix A.F–H. Princip je v obou případech takový, že se zkonstruuje $C^{1,1}$ difeomorfyzmus, který situaci na hranici Ω přenese do jiné vhodné geometrie, ve které se pak řeší jiný eliptický problém s proměnnými koeficienty. Výsledky z [18] pro tento problém se pak přenesou zpět na původní oblast. \square

Regularita a geometrie množiny hran Γ

Se znalostí chování funkce u uvnitř a na hranici oblasti Ω je možné něco říci o geometrii hran Γ . V článku [25] je ukázána následující věta.

Věta 2.18 *Nechť (u, γ) je řešením minimalizace E takové, že Γ se skládá z konečného počtu jednoduchých $C^{1,1}$ křivek γ_i , které se protínají nejvýše ve svých koncových bodech. Pak koncové body těchto křivek (tzn. „vrcholy“ Γ) mají právě jednu z následujících podob:*

- bod, ve kterém γ_i vede na $\partial\Omega$ pod úhlem $\pi/2$,
- bod, ve kterém se schází právě tři křivky γ_i pod úhlem $2\pi/3$ a
- bod, ve kterém jediná γ_i končí a žádnou další křivku neprotíná.

Není tak možné, aby například dvě křivky svíraly pravý úhel nebo aby se protínaly čtyři a více křivek. Je-li tak na obraze u_0 předmět s ostrým rohem tak bude, v malém okolí tohoto rohu, nutně vysegmentován špatně (což však vzhledem k celkové přenosti segmentace nebývá v praktických aplikacích problém). Zdůrazněme, že věta vychází z předpokladu, že Γ se skládá z konečného počtu hladkých křivek, což však nelze apriori předpokládat.

Myšlenka důkazu je taková, že uvažujeme řešení (u, Γ) na malém okolí bodu P , kde nějaká z křivek γ_i končí. Nyní v bodě P postupně uvažujeme jednotlivé situace, které odporují tvrzení, tedy

- z bodu P vychází dvě křivky pod jiným úhlem než π ,
- P je na $\partial\Omega$ a úhel mezi γ_i a $\partial\Omega$ ji jiný než π ,
- tři křivky končí v P a tři takto vzniklé úhly mezi sousedními křivkami nejsou shodné,
- v bodě P začínají čtyři nebo více křivek, v něm dvě křivky tvoří „hrot“, tedy svírají vzájemný úhel 0.

Pro každou z těchto situací lze ukázat, že je možné zkonstruovat (u', Γ') , které se shoduje s (u, Γ) mimo okolí bodu P a takové, že na okolí bodu P je segmentační energie (u', Γ') menší než (u, Γ) , nutně je tak $E(u', \Gamma') < E(u, \Gamma)$, což je spor. Plný důkaz je možné nalézt v [25], podrobněji shrnutý nástin důkazu je uveden v [5].

Dále je tedy třeba ukázat, že množina hran Γ , která minimalizuje E , splňuje předpoklady této věty, aby bylo možné ji užít. Dobrým východiskem je fakt, že S_u jako množina hran slabého řešení, se skládá (až na \mathcal{H}^1 -nulovou množinu) z částí C^1 křivek, pro regularitu Γ v požadavcích věty to však nestačí. Tato otázka je pravděpodobně nejobtížnější částí teorie MS segmentace a není dosud plně zodpovězena. Dobré zmapování postupů a dosažených výsledků je v [24], uveďme zde jeden z hlavních výsledků A.Bonneta [6], který lze vágně formulovat takto: Je-li omezen počet komponent Γ , pak každá izolovaná komponenta splňuje tvrzení věty 2.18, navíc Γ je $C^{1,1}$ \mathcal{H}^1 -skoro všude.

Jednoznačnost řešení

Podívejme se nyní, na otázku jednoznačnosti řešení minimalizace MS funkcionálu. To lze vyřešit poměrně snadno, protože minimalizace bohužel obecně jednoznačná není. Konkrétněji — pro danou množinu hran Γ je regulární aproximace u na jednotlivých Ω_i jednoznačná (viz 2.15) a dána rovnicí (2.8), mohou však existovat dvě různé množiny Γ_1, Γ_2 a jim odpovídající funkce u_1, u_2 tak, že

$$E(u_1, \Gamma_1) = E(u_2, \Gamma_2) = \min_{u, \Gamma} E(u, \Gamma).$$

To je špatná zpráva z hlediska segmentace obrazu, protože ta je dána právě množinou hran Γ a znamená to tedy, že pro jeden obraz může existovat více optimálních segmentací.

Podívejme se na konkrétní příklad nejednoznačného řešení v jednodimenzionálním případě (protipříklad založený na naprosto stejném principu lze ukázat i ve 2D, diskuze možných řešení je však náročnější, dáme tak přednost názornosti). Označme $\mathcal{H}^0(\cdot)$ počítací míru a budeme pracovat s jednodimenzionálním MS modelem

$$E(u, \Gamma) = \int_{\Omega} (u - u_0)^2 + \beta \int_{\Omega \setminus \Gamma} |u'|^2 + \alpha \mathcal{H}^0(\Gamma), \quad u \in W^{1,2}(\Omega \setminus \Gamma), \quad \mathcal{H}^0(\Gamma) < \infty,$$

(množina nespojitostí u je nyní reprezentována konečně mnoha body, jinak je situace naprosto stejná jako ve 2D případě). Nechť $\Omega = (0, 2)$ a mějme vstupní obraz

$$u_0(x, y) = \begin{cases} 0 & x < 1, \\ 1 & x > 1. \end{cases}$$

Podívejme se, jaké mohou být minimalizace E na tomto vstupním obrazu. Nejprve nechť $\mathcal{H}^0(\Gamma_1) > 0$, pak $\mathcal{H}^0(\Gamma_1) \geq 1$ a tedy

$$E(u, \Gamma) \geq \alpha \mathcal{H}^0(\Gamma) \geq \alpha.$$

Pro

$$u_1 = u_0 \text{ na } \Omega \setminus \Gamma_1, \Gamma_1 = \{1\}$$

je $E(u_1, \Gamma_1) = \alpha$, čímž je spodní meze dosaženo.

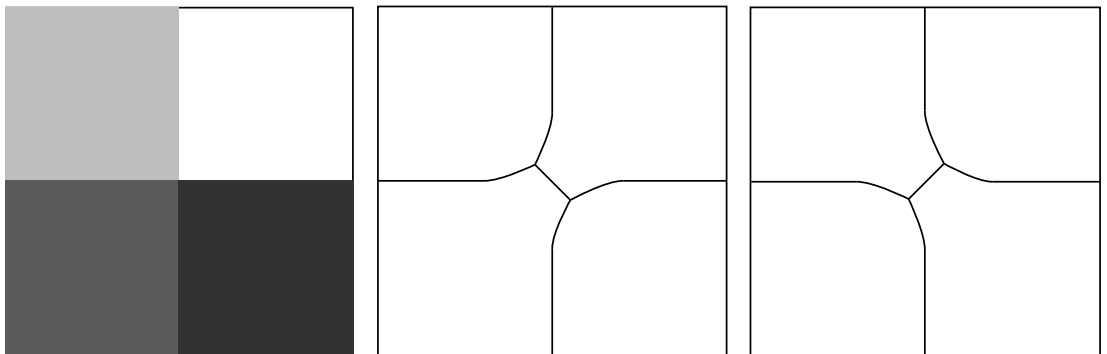
Nechť je nyní $\mathcal{H}^0(\Gamma_2) = 0$, tedy $\Gamma_2 = \emptyset$. Nutně pak existuje (opět viz 2.15) řešení u_2 problému $\min_u E(u, \emptyset) = E(u_2, \emptyset)$. Zvolíme-li nyní $\alpha = E(u_2, \emptyset)$ tak pro u_0 máme

$$E(u_1, \Gamma_1) = \alpha = E(u_2, \Gamma_2) = \min_{u, \Gamma} E(u, \Gamma),$$

neboli dvě různé minimalizace funkcionálu E a tedy dvě různé optimální segmentace obrazu u_0 .

Funkcionál E obsahuje regularizační člen $\int |\nabla u|^2$ a z příkladu je vidět, že „špatné“ rozdělení obrazu, funkcionál z energetického hlediska vykompenzuje dobrou hladkou aproximací. V další části zavedeme zjednodušený funkcionál E^C , který aproximuje po částech konstantními funkcemi, což stačí, pokud nás zajímá pouze segmentace. Mohlo by se zdát, že v tom případě již problém s nejednoznačností nenastane. Na stejném příkladu je však vidět, že zde je situace naprosto stejná. Pro $\Gamma_2 = \emptyset$ je minimalizací $u_2 \equiv 1/2$, pak je $E(u_2, \Gamma_2) = 1/2$ a tedy pro $\alpha = 1/2$ máme stejnou nejednoznačnost řešení.

V obou případech nejednoznačnost sice existuje, ale jen pro konkrétní volbu parametrů, je tak ze statistického hlediska velice nepravděpodobná. Koeficient α představuje škálový parametr, který určuje, zda chceme segmentaci malých (nízké α) nebo velkých (vysoké α) objektů, vzhledem k velikosti obrázku, je tak poměrně přirozené, že pro daný obrázek existuje mezní hodnota, pro kterou jsou si dvě segmentace energeticky ekvivalentní.



Obrázek 2.1: Vstupní obrázek a dvě možné energeticky ekvivalentní segmentace okolí prostředního bodu.

Lze však ukázat i situace, kdy nejednoznačnost spojená s koeficientem α není, konkrétně $\Gamma_1 \neq \Gamma_2$, ale $\mathcal{H}^1(\Gamma_1) = \mathcal{H}^1(\Gamma_2)$. Z části o regularitě řešení víme (ve smyslu „je téměř dokázané“), že jednotlivé křivky, ze kterých se skládá Γ , se mohou protínat

jen ve trojicích a pod úhlem $2\pi/3$. Je-li ve vstupním obrazu prvek jako na obrázku 2.1, je pak v nejlepší případě možná jen některá ze segmentací naznačených na tomto obrázku, Γ nemůže hrany obrazu kopírovat přesně. Obrázek tak ilustruje další zdroj nejednoznačnosti MS segmentace a zároveň již zmíněnou nevýhodu MS modelu — rohy objektů nemohou být nikdy vysegmentovány zcela přesně. Nejednoznačnost je tak MS modelu vlastní a je třeba ji mít na paměti.

Asymptotické verze MS funkcionálu

V poslední části této kapitoly se budeme věnovat třem různým limitním případům MS funkcionálu E . Prakticky půjde o tři nové funkcionály, které vzniknou vždy vynecháním jednoho ze členů v E a vhodnou restrikcí uvažovaného oboru funkcí. Tři takto vzniklé funkcionály, které jsou vlastně speciální případy E , mají ve zpracování obrazu svůj smysl, což ukazuje na důležitost MS modelu.

Rekonstrukce obrazu

Nejprve se podíváme na MS funkcionál bez posledního (délkového) členu, tedy

$$E(u, \Gamma) = \int_{\Omega} (u - u_0)^2 dx + \beta \int_{\Omega \setminus \Gamma} |\nabla u|^2 dx.$$

Protože délka hran Γ není nijak penalizována, je potřeba ji pevně stanovit (omezit tím definiční obor funkcionálu), jinak by neměl smysl. Přírozenou volbou je $\Gamma = \emptyset$, neboli aproximovat u_0 funkcí u , která je regulární v celé oblasti Ω . Zabýváme se tedy minimalizací funkcionálu

$$E^R(u) = \int_{\Omega} \left((u - u_0)^2 + \beta |\nabla u|^2 \right) dx, \quad u_0 \in L^\infty(\Omega), u \in W^{1,2}(\Omega) \quad (2.16)$$

Jedná se tedy o MS funkcionál zcela bez volné množiny nespojitostí Γ . Z věty 2.14 resp. jejího důsledku 2.15 víme, že pro každé $u_0 \in L^\infty(\Omega)$ existuje právě jedno řešení $u \in W^{1,2}(\Omega)$, které splňuje (alespoň jako distribuce a striktně uvnitř Ω skoro všude) rovnici a okrajovou podmínku (2.8), připomeneme ji

$$u - \beta \Delta u = u_0 \text{ na } \Omega, \quad \frac{\partial u}{\partial \mathbf{n}} = 0 \text{ na } \partial\Omega. \quad (2.8)$$

Funkcionál E^R nemá přímé uplatnění v segmentaci obrazu, ale ve zpracování obrazu je znám jako jeden z nejjednodušších modelů rekonstrukce obrazu u z pozorování u_0 (odstranění šumu, inverze rozmazání apod.) v podobě

$$F(u) = \int_{\Omega} (R[u] - u_0)^2 + \beta \varphi(|\nabla u|) dx,$$

kde operátor R představuje poškození obrazu (v našem případě identita) a φ je vhodná regularizační funkce (v našem případě $t \rightarrow t^2$). Řešení minimalizace tohoto funkcionálu (pro identický operátor) je rozmazaná (vyhlazená) verze u_0 (samozřejmě v závislosti na koeficientu β a funkci φ), lze tedy říci, že funkcionál zjednodušuje původní obraz, což je v souladu s MS modelem. Diskuze o tomto funkcionálu v kontextu MS segmentace má smysl ze dvou důvodů. Za prvé si uvědomme, že na každé souvislé Ω_i části dělení Ω podle (2.1) se MS funkcionál E chová jako E^R , segmentovaný obraz tak bude na jednotlivých Ω_i původní obraz vyhlazený rovnicí (2.8). Dále si můžeme uvědomit, že

lze spojit rekonstrukční i segmentační model do jednoho a napsat tak zobecněný MS funkcionál jako

$$E(u, \Gamma) = \int_{\Omega} (R[u] - u_0)^2 dx + \beta \int_{\Omega \setminus \Gamma} \varphi(|\nabla u|) dx + \alpha \mathcal{H}^1(\Gamma).$$

Řekli jsme, že zjednodušené verze MS funkcionálu jsou v určitém smyslu limitami jeho plné verze, podívejme se na toto tvrzení v případě E^R . Pokusíme se intuitivně ukázat, že funkcionál E (resp. jeho řešení) se blíží ke zjednodušené verzi E^R (resp. jeho řešení) pro parametr $\alpha \rightarrow \infty$. Uvažujme plný MS funkcionál a jeho zjednodušenou verzi ve tvarech

$$E_{\alpha}(u, \Gamma) = \int_{\Omega} (u - u_0)^2 dx + \int_{\Omega \setminus \Gamma} |\nabla u|^2 dx + \alpha \mathcal{H}^1(\Gamma),$$

$$E^R(u) = \int_{\Omega} (u - u_0)^2 dx + \int_{\Omega} |\nabla u|^2 dx.$$

Nechť $(u_{\alpha}, \Gamma_{\alpha})$ je řešení $E_{\alpha}(u, \Gamma)$ s daným koeficientem α a u je řešení zjednodušené verze $E^R(u)$. Označme dále hodnoty minim

$$E_{\alpha} = E_{\alpha}(u_{\alpha}, \Gamma_{\alpha}), \quad E^R = E^R(u)$$

a energii délkového členu

$$h(\alpha) = \mathcal{H}^1(\Gamma_{\alpha}).$$

Věta 2.19 *Tyto veličiny pak mají následující vlastnosti*

1. $E_{\alpha} \leq E, \quad \forall \alpha > 0,$
2. E_{α} je neklesající a $h(\alpha)$ je nerostoucí v α ,
3. $h(\alpha) \leq \frac{E^R}{\alpha}$, mj. tak $h(\alpha) \rightarrow 0$ pro $\alpha \rightarrow \infty$.

Důkaz Platí, že

$$E_{\alpha} = E_{\alpha}(u_{\alpha}, \Gamma_{\alpha}) \leq E_{\alpha}(u, \emptyset) = E^R(u) = E^R,$$

z čehož plyne 1. Dále je $\alpha h(\alpha) \leq E_{\alpha} \leq E^R$, takže

$$h(\alpha) \leq \frac{E_{\alpha}}{\alpha} \leq \frac{E^R}{\alpha},$$

což dokazuje 3. Zbývá ukázat monotonii E_{α} a $h(\alpha)$. Mějme $\alpha_2 > \alpha_1$, pak dosazením dostaneme

$$E_{\alpha_1} = E_{\alpha_1}(u_{\alpha_1}, \Gamma_{\alpha_1}) \leq E_{\alpha_1}(u_{\alpha_2}, \Gamma_{\alpha_2}) = E_{\alpha_2}(u_{\alpha_2}, \Gamma_{\alpha_2}) + (\alpha_1 - \alpha_2)h(\alpha_2),$$

kde poslední člen je záporný. Tím je dokázána monotonie E_{α} . Stejným způsobem můžeme napsat

$$E_{\alpha_2} \leq E_{\alpha_1} + (\alpha_2 - \alpha_1)h(\alpha_1).$$

Porovnáním $h(\alpha_1)$ a $h(\alpha_2)$ dostaneme z posledních dvou nerovnic monotonii h . \square

Co nám tato věta říká: Energie segmentace se zhoršuje s rostoucím požadavkem na krátkost hran a jakákoliv segmentace s hranami je považována za lepší než „segmentace“ zcela bez hran. Délky hran naopak klesají čím více jsou penalizovány, limitně dosahují nulové hodnoty, neboli Γ_{α} se zmenšuje až na \mathcal{H}^1 -nulovou množinu. Z toho lze usoudit, že jelikož

$$\operatorname{argmin}_u E_{\alpha}(u, \emptyset) = \operatorname{argmin}_u E^R(u),$$

bude se řešení u_{α} čím dál více blížit asymptotickému řešení u .

Po částech konstantní aproximace

Nyní si představíme funkcionál, který je ze segmentačního hlediska velice důležitý a často používaný (případně v mírně upravených verzích), také v této práci se mu budeme věnovat ještě v příští kapitole. Podívejme se však na něj ze stejného hlediska jako v předchozím případě na „rekonstrukci“ obrazu. Budeme se zabývat MS funkcionálem bez gradientního členu, tedy

$$E^C(u, \Gamma) = \int_{\Omega} (u - u_0)^2 dx + \alpha \mathcal{H}^1(\Gamma).$$

Stejně jako v předchozím případě je potřeba omezit definiční obor tohoto funkcionálu, aby jeho minimalizace nebyla triviální. Tentokrát není nijak penalizována derivace řešení u , které se tak může libovolně přiblížit k u_0 — opět je tedy potřeba tuto hodnotu předepsat a přirozeně požadovat, aby

$$\nabla u = 0.$$

Protože po funkci u požadujeme regularitu jen mimo hrany na $\Omega \setminus \Gamma$ a také MS funkcionál penalizuje gradient funkce jen na této množině, budeme uvažovat funkce u po částech konstantní v Ω s přípustnými skoky přes Γ . Formálně stanovme definiční obor \mathcal{D}^C funkcionálu E^C následujícím způsobem: Nechť Ω_i, Γ je dělení oblasti Ω podle 2.1, pak

$$\mathcal{D}^C = \{(u, \Gamma), u|_{\Omega_i} \equiv u_i = \text{konst. } \forall i, \Gamma \text{ je uzavřená v } \Omega, \mathcal{H}^1(\Gamma) < \infty\}.$$

Všechny takové funkce jsou přípustné i pro plný MS funkcionál, jedná se tak o omezení jeho definičního oboru. Z hlediska segmentace dává omezení na po částech konstantní funkce perfektní smysl — jestliže nás nezajímá zjednodušení (nebo chceme-li rekonstrukce) obrazu v rámci jednotlivých objektů, jako je tomu u plného MS modelu, ale zajímáme se pouze o rozdělení oblasti Ω na jednotlivé objekty Ω_i a hrany Γ , tak funkcionál E^C nám dává přesně tuto informaci, Ω_i pak odpovídají množinám, na kterých je hodnota u konstantní, a Γ bodům, na kterých je u nespojitě. Z tohoto důvodu je této variantě MS funkcionálu věnováno mnoho pozornosti.

Podívejme se na funkcionál E^C blíže. Nejprve učiňme pozorování, že funkce u je pro pevné Γ jednoznačně určena (ve smyslu L^p) svými hodnotami u_i na jednotlivých Ω_i . Protože druhý člen E^C na u nezávisí, jsou hodnoty u_i dány jen prvním členem, který je možné přepsat

$$\int_{\Omega} (u - u_0)^2 dx = \sum_i \int_{\Omega_i} (u_i - u_0)^2 dx.$$

Pro pevné Γ je minimalizace tohoto členu triviální (nejmenší čtverce) — na každé Ω_i musí mít u_i hodnotu, která se rovná průměru u_0 přes Ω_i , tedy

$$u_i = \langle u_0 \rangle_{\Omega_i} = \frac{1}{|\Omega_i|} \int_{\Omega_i} u_0 dx.$$

Funkce u se pak pro dané Γ rovná

$$u = \langle u_0 \rangle_{\Gamma} = \sum_i \langle u_0 \rangle_{\Omega_i} \mathbf{1}_{\Omega_i}. \quad (2.17)$$

Proměnné u a Γ tak nejsou nezávislé, funkce u je jednoznačně určena, jmile je známé Γ . Funkcionál E^C tak můžeme uvažovat jen jako funkci Γ , které určuje dělení (Ω_i, Γ) a tím i hodnoty u_i

$$E^C(\Gamma) = \int_{\Omega} (\langle u_0 \rangle_{\Gamma} - u_0)^2 dx + \alpha \mathcal{H}^1(\Gamma). \quad (2.18)$$

Podívejme se nyní na některé vlastnosti tohoto funkcionálu v podobném duchu, jako jsme to udělali s předchozím zjednodušeným modelem. Pokusíme se opět ukázat, že funkcionál E^C je limitním případem plného MS funkcionálu, tentokrát pro $\beta \rightarrow \infty$. Označme plný MS funkcionál (se zdůrazněnou závislostí na parametru β)

$$E_\beta(u, \Gamma) = \int_{\Omega} (u - u_0)^2 dx + \beta \int_{\Omega \setminus \Gamma} |\nabla u|^2 dx + \mathcal{H}^1(\Gamma),$$

dále necht' (u_β, Γ_β) resp. Γ a jím určené $u = \langle u_0 \rangle_\Gamma$ jsou řešení minimalizace E_β resp. E^C , s hodnoty minim

$$E_\beta = E_\beta(u_\beta, \Gamma_\beta), \quad E^C = E^C(\Gamma)$$

a nakonec označme hodnotu gradientního členu

$$s(\beta) = \int_{\Omega \setminus \Gamma_\beta} |\nabla u_\beta|^2 dx.$$

Věta 2.20 *Pak je splněno následující:*

1. $E_\beta \leq E^C, \quad \forall \beta > 0,$
2. E_β je neklesající a $s(\beta)$ je nerostoucí v β ,
3. $s(\beta) \leq \frac{E^C}{\beta}$, mj. tak $s(\beta) \rightarrow 0$ pro $\beta \rightarrow \infty$.

Důkaz Je zcela analogický k důkazu věty 2.19. Platí

$$E_\beta = E_\beta(u_\beta, \Gamma_\beta) \leq E_\beta(u, \Gamma) = \int_{\Omega} (u - u_0)^2 dx + \beta \int_{\Omega \setminus \Gamma} \underbrace{|\nabla u|^2}_{\equiv 0} dx + \mathcal{H}^1(\Gamma) = E^C(\Gamma) = E^C$$

z čehož plyne 1. Dále je $\beta s(\beta) \leq E_\beta \leq E^C$, tedy

$$s(\beta) \leq \frac{E^C}{\beta} \rightarrow 0 \text{ pro } \beta \rightarrow \infty,$$

což dokazuje 3. Zbývá monotonie E_β a $s(\beta)$. Mějme $\beta_1 < \beta_2$, pak platí

$$\begin{aligned} E_{\beta_1} &= E_{\beta_1}(u_{\beta_1}, \Gamma_{\beta_1}) \leq E_{\beta_1}(u_{\beta_2}, \Gamma_{\beta_2}) \\ &= \int_{\Omega} (u_{\beta_2} - u_0)^2 dx + \beta_1 \int_{\Omega \setminus \Gamma_{\beta_2}} |\nabla u_{\beta_2}|^2 dx + \mathcal{H}^1(\Gamma_{\beta_2}) = E_{\beta_2} - \underbrace{(\beta_2 - \beta_1)s(\beta_2)}_{\geq 0}, \end{aligned}$$

čímž je dokázána monotonie E_β . Analogickým postupem lze naopak napsat

$$E_{\beta_2} \leq E_{\beta_1} + (\beta_2 - \beta_1)s(\beta_1).$$

Z posledních dvou nerovnic dostáváme monotonii $s(\beta)$. □

Nad tvrzením lze provést stejnou diskuzi jako v případě věty 2.19. S rostoucím parametrem β se celková variace funkce u na jednotlivých oblastech Ω_i monotónně zmenšuje až do nuly, řešení se tak čím dál více podobá po částech konstantnímu řešení. Energie segmentace naopak monotónně roste, ale je shora omezena energií po částech konstantní segmentace.

Předpokládáme-li i pro rostoucí β jednotlivé oblasti Ω_i dostatečně regulární, lze asymptotický vztah řešení u_β a u ukázat ještě jiným způsobem. Necht' pro Ω_i platí Poincarého nerovnost (např. [14]) pro odchylku funkce od jejího průměru

$$\int_{\Omega_i} (v - \langle v \rangle_{\Omega_i})^2 dx \leq C_i \int_{\Omega_i} |\nabla v|^2 dx. \quad (2.19)$$

Označme $C^\beta = \max_i C_i$ pro dané β . Jsou-li oblasti Ω_i natolik regulární, že pro β v okolí nekonečna roste C^β pomaleji než lineárně, neboli

$$\lim_{\beta \rightarrow \infty} \frac{C^\beta}{\beta} = 0, \quad (2.20)$$

platí pro řešení u_β a jeho po částech konstantní aproximaci $\langle u_\beta \rangle_{\Gamma_\beta}$

$$\int_{\Omega} (u_\beta - \langle u_\beta \rangle_{\Gamma_\beta})^2 = \sum_i \int_{\Omega_i} (u_\beta - \langle u_\beta \rangle_{\Omega_i})^2 \leq \sum_i C_i \int_{\Omega_i} |\nabla u_\beta|^2 \leq C^\beta s(\beta) \rightarrow 0. \quad (2.21)$$

Řešení u_β tedy přechází na svůj po částech konstantní průměr $\langle u_\beta \rangle_{\Gamma_\beta}$. Z hlediska segmentce je však zajímavé a podstatné, zda se asymptoticky shoduje segmentace (tedy dělení oblasti $\Omega = \bigcup_i \Omega_i \cup \Gamma_{(\beta)}$) daná funkcí $\langle u_\beta \rangle_{\Gamma_\beta}$ a funkcí $u = \langle u_0 \rangle_{\Omega}$. Můžeme se tak ptát, jak blízko je $\langle u_\beta \rangle_{\Gamma_\beta}$ k minimalizaci E^C . Odpověď dává následující věta.

Věta 2.21 *Nechť $C^\beta = \max_i C_i$, kde C_i jsou Poincarého konstanty z (2.19), roste s β pomaleji než lineárně, (2.20). Označme (u_β, Γ_β) resp. (u, Γ) řešení minimalizace E_β resp. E^C jako v předchozím případě, $\langle u_\beta \rangle_{\Gamma_\beta}$ po částech konstantní zprůměrování u_β vzhledem k dělení indukovaném Γ_β .*

Pak je

$$\lim_{\beta \rightarrow \infty} E^C(\langle u_\beta \rangle_{\Gamma_\beta}, \Gamma_\beta) = E^C(u, \Gamma) \quad (2.22)$$

Neboli — funkce $\langle u_\beta \rangle_{\Gamma_\beta}$ (která je po částech konstantní aproximací a navíc limitou u_β , jak je vidět z (2.21)) asymptoticky minimalizuje E^C , segmentace po částech konstantního modelu E^C je tak limitou plné segmentace pomocí E_β .

Důkaz Je zřejmé

$$E^C(u, \Gamma) \leq E^C(\langle u_\beta \rangle_{\Gamma_\beta}, \Gamma_\beta),$$

ukážme opačnou nerovnost (v limitě). Tvrdíme nejprve, že

$$\left| \|\langle u_\beta \rangle_{\Gamma_\beta} - u_0\|_{L^2}^2 - \|u_\beta - u_0\|_{L^2}^2 \right| \leq K \sqrt{\frac{C^\beta}{\beta}}, \quad (2.23)$$

kde K je nezávislé na β . Potom je

$$\begin{aligned} E^C(\langle u_\beta \rangle_{\Gamma_\beta}, \Gamma_\beta) &= E_\beta(\langle u_\beta \rangle_{\Gamma_\beta}, \Gamma_\beta) \quad (= \|\langle u_\beta \rangle_{\Gamma_\beta} - u_0\|_{L^2}^2 + \mathcal{H}^1(\Gamma_\beta)) \\ &\leq E_\beta(u_\beta, \Gamma_\beta) + K \sqrt{\frac{C^\beta}{\beta}} \quad (= \|u_\beta - u_0\|_{L^2}^2 + \beta s(\beta) + \mathcal{H}^1(\Gamma_\beta) + K \dots) \\ &\leq E_\beta(u, \Gamma) + K \sqrt{\frac{C^\beta}{\beta}} \\ &= E^C(u, \Gamma) + K \sqrt{\frac{C^\beta}{\beta}}. \end{aligned}$$

Je tedy

$$E^C(u, \Gamma) \leq E^C(\langle u_\beta \rangle_{\Gamma_\beta}, \Gamma_\beta) \leq E^C(u, \Gamma) + K \sqrt{\frac{C^\beta}{\beta}},$$

z čehož za předpokladu $C^\beta/\beta \rightarrow 0$ plyne tvrzení věty. Zbývá ukázat (2.23). Je $u_0 \in L^\infty(\Omega)$ a tedy i $u_\beta \in L^\infty(\Omega)$, přičemž $\|u_\beta\|_{L^\infty} \leq \|u_0\|_{L^\infty}$ (viz 2.17 tvrzení 2). Z

trojúhelníkové nerovnosti a standardního odhadu integrálu dostáváme

$$\begin{aligned}
\|\langle u_\beta \rangle_{\Gamma_\beta} - u_0\|_{L^2}^2 - \|u_\beta - u_0\|_{L^2}^2 &\leq \left(\|\langle u_\beta \rangle_{\Gamma_\beta} - u_0\|_{L^2} + \|u_\beta - u_0\|_{L^2} \right) \|u_\beta - \langle u_\beta \rangle_{\Gamma_\beta}\|_{L^2} \\
&\leq 4\sqrt{|\Omega_i|} \|u_0\|_{L^\infty} \|u_\beta - \langle u_\beta \rangle_{\Gamma_\beta}\|_{L^2} \\
&\leq 4\sqrt{|\Omega_i|} \|u_0\|_{L^\infty} \sqrt{C^\beta} \sqrt{s(\beta)} \\
&\leq 4\sqrt{|\Omega_i|} \|u_0\|_{L^\infty} \sqrt{E^c} \sqrt{\frac{C^\beta}{\beta}},
\end{aligned}$$

kde předposlední resp. poslední nerovnost plyne z Poincarého nerovnosti v (2.21) resp. z věty 2.20 vlastnosti 3. \square

Existenci minimalizace E^C lze ukázat snáze než existenci minimalizace E , pro u_0 spojitou byla ukázána již v původním článku [25], posléze jiným způsobem pro $u_0 \in L^\infty(\Omega)$ v [23]. Lze také samozřejmě uplatnit argumenty, kterými byla dokázána existence minima funkcionalu E a existenci E^C získat stejným způsobem. Shrňme výsledek přehledně do věty.

Věta 2.22 ([23]) *Nechť $u_0 \in L^\infty(\Omega)$, pak existuje $(u, \Gamma) \in \mathcal{D}^C$ minimalizující E^C na této množině. Navíc, Γ se skládá z konečně mnoha $C^{1,1}$ křivek γ , které se protínají jen ve svých koncových bodech. Jediný možný koncový bod křivky γ je buď bod, kde se potkávají právě 3 křivky pod úhly $2\pi/3$, nebo bod, kde křivka γ vede až na $\partial\Omega$ pod úhlem $\pi/2$.*

Řešení minimalizace E^C tak splňuje MS domněnku vyslovenou v [25] s následujícím rozdílem. Je známo, že řešení (u, Γ) plně MS segmentace může obsahovat tzv. slepé hrany (v originálním článku nazvané *crack-tips*), což jsou křivky $\gamma \subset \Gamma$, které jsou z obou stran obklopeny stejnou souvislou Ω_i — hrana tak netvoří hranici mezi objekty, jak je očekáváno, ale „díru“ v jednom objektu. Jsou případy, kdy je to jediné rozumné řešení, například obraz $u_0(r, \theta) = \theta$ v polárních souřadnicích se skládá z jednoho hladkého celku s nepojitostí na úsečce $\theta = 0$. V praktických aplikacích se však takové případy objektů většinou nevyskytují a, co je podstatné, takovéto hrany mohou způsobovat implementační obtíže (například při implementaci MS modelu pomocí tzv. level setů, o kterých pojednává příští kapitola). V po částech konstantním MS modelu, kde jsou aproximace jednotlivých objektů Ω_i beze zbytku konstantní, již slepé hrany nedávají žádný smysl vůbec. Ukažme si tedy, že řešení E^C takové hrany ani nemůže obsahovat, což je dobrá zpráva z hlediska implementace. Nejprve formální definice.

Definice 2.23 *Nechť Γ je uzavřená v Ω a $\mathcal{H}^1(\Gamma) < \infty$. Řekneme, že $\gamma \subset \Gamma$ je slepá hrana, jestliže $\mathcal{H}^1(\gamma) > 0$ a existuje otevřená množina $Q \subset \Omega \setminus (\overline{\Gamma \setminus \gamma})$, $\mathcal{H}^1(\gamma \setminus Q) = 0$ taková, že $Q \setminus \gamma$ je souvislá.*

Definice tedy říká, že slepá hrana γ jako součást množiny hran Γ je zbytečná v tom smyslu, že nerozděluje žádné dvě podmnožiny dělení Ω_i a Ω_j , které by jinak tvořily jednu souvislou množinu. Přístupme k tvrzení.

Věta 2.24 *Nechť Γ je řešení minimalizace E^C . Pak Γ neobsahuje žádnou slepou hranu.*

Důkaz Předpokládejme pro spor, že $\gamma \subset \Gamma$ je slepá. Označme $\Gamma' = \overline{\Gamma \setminus \gamma}$, ukážeme, že Γ' tvoří lepší řešení minimalizace E^C . Nejprve zřejmě

$$\mathcal{H}^1(\Gamma') < \mathcal{H}^1(\Gamma).$$

Tvrdíme dále, že pro po částech konstantní řešení (dané (2.17)) indukovaná dělení Γ resp. Γ' platí

$$\langle u_0 \rangle_{\Gamma'} = \langle u_0 \rangle_{\Gamma}. \quad (2.24)$$

Pak je tedy $E^C(\Gamma') < E^C(\Gamma)$, což je spor. Zbývá ukázat (2.24).

Nechť Q je množina z definice 2.23, pak $Q \setminus \gamma$ je nutně obsažena právě v jedné Ω_i , označme ji Ω_1 . Kdyby ne — necht' má pro spor Q neprázdný průnik s Ω_1 a Ω_2 (s ostatními Ω_i ne), protože neobsahuje žádnou část $\Gamma \setminus \gamma$ tak musí být $Q \setminus \gamma \subset (\Omega_1 \cup \Omega_2)$, tzn. ale že $\Omega_1 \cup \Omega_2$ je souvislá (existuje cesta přes $Q \setminus \gamma$) přičemž Ω_1 a Ω_2 jsou disjunktí otevřené, což je spor se souvislostí. Pokračujme, označme $\Omega'_1 = \Omega_1 \cup Q$. Ta je nutně otevřená a souvislá (Ω_1 i Q jsou otevřené, souvislé a mají neprázdný průnik). Protože Γ a Γ' se liší jen o γ , lze si snadno rozmyslet, že otevřené oblasti dělení Ω indukovaných Γ resp. Γ' se liší právě jen množinou Ω_1 resp. Ω'_1 , ostatní $\Omega_i = \Omega'_i$ musí být shodné až na přecíslování. Funkce $\langle u_0 \rangle_\Gamma$ resp. $\langle u_0 \rangle_{\Gamma'}$ se tak liší jen ve členu (viz definice (2.17)) $\langle u_0 \rangle_{\Omega_1}$ resp. $\langle u_0 \rangle_{\Omega'_1}$ a protože γ má nulovou 2D Lebesgueovu míru, musí být tyto příspěvky shodné. \square

Active Contours

V poslední části diskuze o limitních verzích MS funkcionálu se podíváme na situaci, kdy ve funkcionálu

$$E(u, \Gamma) = \lambda \int_{\Omega} (u - u_0)^2 dx + \int_{\Omega \setminus \Gamma} |\nabla u|^2 dx + \alpha \mathcal{H}^1(\Gamma) \quad (2.25)$$

chceme provést limitu $\lambda \rightarrow \infty$ (neboli uvažovat funkcionál E bez prvního členu). Aby výsledek lépe zapadl do kontextu, je vhodné nejprve představit jinou třídu variačních segmentačních modelů, tzv. active contours metody. To je žádoucí i ze dvou dalších důvodů, jednak jako doplnění ke studiu MS modelu, aby čtenář získal komplexnější obraz o variačních segmentačních metodách, a také jako vhodný úvod pro následující kapitolu.

MS model, kterým jsme se doteď zabývali, se snaží rozdělit vstupní obraz u_0 na oblasti Ω_i s přibližně homogenní intenzitou. Za duální přístup je možné považovat nalezení křivek, které odpovídají hranám v obraze u_0 . Hrany se ve zpracování obrazu tradičně rozpoznávají z gradientu intenzity, tedy $|\nabla u_0|$, zjednodušeně lze říci, že čím vyšší gradient, tím větší je v daném místě hrana⁶. Za hrubou variační formulaci detekce hran lze tak považovat toto: Nalézt křivku Γ , podél které má obraz u_0 veliký gradient a která je rozumně dlouhá. Jednou z možností jak toho dosáhnou jsou zmíněné active contours metody, které lze popsat následujícím způsobem. Pro křivku Γ definujeme její energii

$$J(\Gamma) = \int_{\Gamma} f(\Gamma, \Gamma', \Gamma'', \dots, u_0, \nabla u_0, \dots).$$

Dále zavedme umělý čas t a označme $\Gamma(t)$ křivku v čase t , $\Gamma(0) = \Gamma_0$, a řešme diferenciální rovnici

$$\frac{\partial \Gamma}{\partial t} = -J'(\Gamma), \quad (2.26)$$

neboli metodou největšího spádu minimalizujeme energii křivky $J(\Gamma)$. Nejedná se tak o „pravou“ minimalizaci ve smyslu

$$J(\Gamma) = \inf_{\Gamma'} J(\Gamma'),$$

protože ve většině případu by taková úloha měla triviální řešení, proces se však zastaví pro $\partial \Gamma / \partial t = 0$ v lokálním minimu. Je-li například $J(\Gamma) = \text{délka}(\Gamma)$, pak evoluční rovnice má tvar

$$\frac{\partial \Gamma}{\partial t} = \text{curv } \Gamma \mathbf{n}, \quad (2.27)$$

⁶Protože detekce hran pomocí gradientu je silně nestabilní s přítomností šumu, namísto $|\nabla u_0|$ se většinou používá $|\nabla(u_0 * \eta_\sigma)|$, kde η_σ je vhodné konvoluční jádro.

což je tzv. *mean curvature motion*, pohyb nejvíce zmenšující délku křivky (v tomto případě se proces zastaví až pro $\Gamma = \emptyset$).

Podívejme se nyní na několik konkrétních active contours modelů. Uvažujme vždy křivku s parametrizací $c(p) = (x(p), y(p))$, $p \in [a, b]$ po částech dostatečně hladkou. Jeden z prvních modelů tohoto typu pochází z [21], kde je položeno

$$J_s(c) = \int_a^b (|c'(p)|^2 + |c''(p)|^2) dp - \int_a^b |\nabla u_0(c(p))|^2 dp.$$

První integrál představuje vnitřní energii křivky, první člen měří délku, druhý křivost — evoluce křivky tak zmenšuje a narovnáva křivku. Druhý integrál je vnější energie (daná obrazem), kde je křivka zvýhodněna, pokud vede v místě hrany obrazu. Princip (společný pro většinu active contours metod) je tedy takový, že počáteční uzavřená křivka c_0 obklopuje objekt, který chceme segmentovat. Evolucí podle (2.26) (v tomto případě rovnice 4. řádu, lze snadno odvodit, viz [21]) se křivka zmenšuje, dokud nekopíruje hrany objektu (druhý integrál tak sníží energii). Tam se její evoluce zastaví, protože pokud by objekt „přeskočila“ tak naroste hodnota druhého integrálu a celková energie tím vzroste. Vzhledem k evoluci křivky nazvali autoři tuto metodu *snake model*.

Model s energií J_s má několik nevýhod. Člen s $|c''(p)|$, který snižuje zakřivení, přidává na složitosti evoluční rovnice a bývá často vynecháván. Člen s $|c'(p)|$ totiž snižuje celkovou délku křivky a jak je vidět z (2.27), tento člen způsobí právě narovnávání křivky a tedy také snižuje její křivost. Lze tak uvažovat jen

$$J_{s'}(c) = \int_a^b |c'(p)|^2 dp - \int_a^b |\nabla u_0(c(p))|^2 dp. \quad (2.28)$$

Druhou a podstatnější nevýhodou je fakt, že hodnota energie křivky závisí na její parametrizaci, takže například dvě stejné segmentace mohou mít jinou hodnotu energie, nebo ze dvou „stejných“ počáteční křivky můžeme dostat rozdílná řešení, jestliže budou křivky jinak parametrizovány. Další nevýhodou je praktická realizace numerického výpočtu (navržena v [21]), kde je potřeba řešit mnoho dílčích problémů pro správné sledování křivky v průběhu jejího vývoje (lze si představit, že je velice náročná například změna topologie křivky — rozdělení na dvě křivky pro detekci více objektů nebo objektů, které nejsou jednoduše souvislé).

Tyto problémy z velké části řeší tzv. *geodesic active contours* model pocházející z [7]. Položme

$$J_{\text{gac}}(c) = \int_a^b g(|\nabla u_0(c(p))|) |c'(p)| dp, \quad (2.29)$$

kde $g : \mathbb{R} \cup \{\infty\} \rightarrow \mathbb{R}_0^+$ je tzv. *edge-detector function*, klesající funkce, která má vysokou hodnotu okolo nuly a blíží se k nule v nekonečnu (tzn. invertuje velikost gradientu). Nejčastějším případem je $s \rightarrow 1/(1 + s^2)$.

Kdyby $g \equiv 1$, tak $J_{\text{gac}}(c)$ měří jen délku křivky. Je-li však třeba $g \in [0, 1]$ a g klesající s rostoucím $|\nabla u_0|$, délka křivky je vážena tím, zda křivka vede po hraně či nikoliv. Vede-li křivka homogenní plochou, je J_{gac} rovno její délce, pokud vede křivka po ideální hraně, je $J_{\text{gac}} = 0$. Hodnota energie nyní také nezáleží na parametrizaci křivky, protože $ds = |c'(p)| dp$ je délkový element, můžeme napsat

$$J_{\text{gac}}(c) = \int_0^{\text{délka } c} g(|\nabla u_0(c(s))|) ds$$

zcela bez závislosti na konkrétní parametrizaci. Další výhodou, která však nyní ještě není zřejmá, je dobrá možnost numerické implementace pomocí tzv. level set metody,

s tou se setkáme v následující kapitole. Pro více informací o vztahu mezi J_{gac} a $J_{s'}$ odkažme na [5, 7].

Než opustíme active contours a vrátíme se k MS funkcionálu, pokusme se stručně shrnout základní vlastnosti těchto metod. Jak jsme již řekli, jedná se nalezení lokálního minima z předem dané počáteční polohy křivky, výsledek segmentace je tak jistě závislý na počáteční křivce. Určujícím kritériem pro zastavení křivky na hraně objektu je lokální minimum energie díky $g(|\nabla u_0|)$ nebo jinému ekvivalentnímu členu, toto minimum však může být velice malé (zejména pokud je hrana objektu neostrá nebo příliš potlačena počátečním vyhlazením $u_0 * \eta_\sigma$). V numerické implementaci se tak snadno může stát, že křivka (často jen její část) hranu objektu „přeskočí“ a s ní i celé lokální minimum. Jámile se tak stane, tak již není pomoci, protože křivka se začne opět zmenšovat a není žádná síla, který by ji vrátila zpět na hranu objektu. Zmíněné active contours metody jsou v tomto smyslu poměrně nestabilní, o úspěšné či neúspěšné segmentaci může rozhodnout malá změna některého ze vstupních parametrů. Naopak mohou tyto metody dávat dobré výsledky při segmentaci objektů s ostrými vnějšími hranami, libovolně složitou vnitřní texturou na víceméně homogenním pozadí. Křivka se pak zastaví na hraně objektu a nevadí, že vnitřek objektu tvoří komplikovaná nehomogenní textura s mnoha vlastními „hranami“, protože křivka se tam vůbec nedostane. Přímá aplikace MS modelu na tento případ (stačí například homogenní objekt, který je však nehomogenně osvětlený) nemusí dát dobré výsledky, protože metoda začne rozdělovat vnitřek objektu.

Zcela specifický je zatím nezmíněný model z [8] nazvaný *active contours without edges*, který má navzdory svému názvu mnohem blíže k MS segmentaci (jde víceméně o E^C) než k active contours, z nich však přebírá myšlenku evoluce křivky. Tímto modelem se budeme zabývat v další kapitole.

Vraťme se ke zmíněnému limitnímu případu MS funkcionálu. Jak již bylo řečeno, chceme v (2.25) provést limitu $\lambda \rightarrow \infty$. Ukazuje se, že je vhodné zároveň položit $\alpha = 2\alpha_0 \frac{1}{\sqrt{\lambda}}$, přeznačme tedy funkcionál E na

$$E(u, \Gamma) = \lambda \int_{\Omega} (u - u_0)^2 dx + \int_{\Omega \setminus \Gamma} |\nabla u|^2 dx + \frac{2\alpha_0}{\sqrt{\lambda}} \mathcal{H}^1(\Gamma).$$

Dále zavedme nový funkcionál (jako asymptotickou verzi MS funkcionálu)

$$E^{\text{AC}}(\Gamma) = \int_{\Gamma} \left(\alpha_0 - \left(\frac{\partial u_0}{\partial \mathbf{n}} \right)^2 \right). \quad (2.30)$$

Lze pak ukázat, že E^{AC} je výše popsanou limitou MS funkcionálu E ve smyslu následující věty.

Věta 2.25 ([25]) *Předpokládejme, že $u_0 \in C^{1,1}(\overline{\Omega})$ a Γ je složena z konečně mnoha $C^{1,1}$ křivek, které netvoří „hroty“⁷. Označme pro větší přehlednost $\varepsilon = 1/\lambda$. Pak je pro $\varepsilon \rightarrow 0+$*

$$E(u_{\Gamma}, \Gamma) = 2\sqrt{\varepsilon} E^{\text{AC}}(\Gamma) + \int_{\Omega} |\nabla u_0|^2 - \sqrt{\varepsilon} \int_{\partial\Omega} \left(\frac{\partial u_0}{\partial \mathbf{n}} \right)^2 - K(\varepsilon \log \varepsilon),$$

kde u_{Γ} je jednoznačné řešení minimalizace $E(\cdot, \Gamma)$ pro pevné Γ .

Omezení $u_0 \in C^{1,1}$ je sice restriktivní z hlediska praktické segmentace, ale zde se jím nemusíme příliš zabývat, protože nám jde o asymptotické chování E , navíc v active

⁷Tzn. že z jednoho bodu vychází dvě křivky se stejným tečným vektorem a směrem, jako v obrázku srdíčka.

contours modelech se téměř vždy uvažuje u_0 vyhlazené konvolucí. Požadavek $\Gamma \in C^{1,1}$ je také restriktivní, ale charakterizace regularity řešení MS funkcionálu ukazují, že téměř s touto regularitou lze počítat. Pak druhý a třetí člen na pravé straně jsou konstanty vzhledem ke Γ , poslední člen jde do nuly rychleji než všechny předchozí, stačí nám tak uvažovat z pravé strany jen první člen, $2\sqrt{\varepsilon} E^{\text{AC}}(\Gamma)$. Věta nám tak říká, že pro vysoké λ (tj. nízké ε) je energie segmentace téměř výhradně koncentrována podél množiny hran Γ . Podívejme se však na tvar E^{AC} — energie roste s délkou křivky (první člen), ale klesá, pokud má obraz u_0 kolmo na křivku vysokou derivaci (druhý člen), jsou tak preferovány krátké křivky, které vedou kolmo na směr největšího růstu intenzity u_0 . Porovnáme-li nyní $J_{s'}(c)$ z (2.28) a E^{AC} z (2.30), vidíme, že se jedná o prakticky stejné funkcionály — oba penalizují délku křivky a favorizují průběh křivky po hraně obrazu. Pro rostoucí λ tak MS model asymptoticky přechází na active contours model.

Důkaz věty 2.25 je poměrně dlouhý a technický, výchozím krokem je charakterizace energie vázané na Γ ve tvaru ([25] část 6, lemma)

$$E(u_\Gamma, \Gamma) = E(\tilde{u}, \emptyset) - \int_\Gamma \left(\frac{\partial \tilde{u}}{\partial \mathbf{n}} (h_\Gamma^+ - h_\Gamma^-) - \alpha \right),$$

kde \tilde{u} je řešení bez hran (tedy $\tilde{u} - \Delta \tilde{u} = u_0$ v Ω) a $h_\Gamma^+, -$ jsou stopy funkce $h_\Gamma = u_\Gamma - \tilde{u}$ na levé resp. pravé straně Γ . Toto lze ukázat jednoduše rozepsáním $E(u_\Gamma, \Gamma) - E(\tilde{u}, \emptyset)$ a aplikací Greenovy věty. Dále je potřeba vhodně odhadnout h_Γ , aby bylo možné provést limitu $\lambda \rightarrow \infty$. Plné znění věty a důkaz je možné nalézt v [25], Theorem 6.1, jiný a jednodušší důkaz omezený na 1D problém je uveden v [12].

Kapitola 3

Active Contours Without Edges

V této kapitole se podíváme na implementaci konkrétní zjednodušené verze MS segmentačního modelu. Jedná se o metodu nazvanou active contours without edges, neboli active contours, které nejsou založeny na gradientních hranách. Prakticky se jedná o restrikcí MS modelu na po částech konstantní funkce, která byla uvedena v předchozí kapitole jako limitní případ MS modelu. Svým způsobem má však metoda blízko i k active contours metodám (z toho pochází název), neboť se zpravidla implementuje jako evoluce křivky ve fiktivním čase.

Začátek kapitoly věnujeme představení tzv. level set metody pocházející z [26], 1988. Jde právě způsob, jak snadno popsat křivku a její vývoj v čase. To má v kontextu variační segmentace obrazu dobrý smysl, neboť většina active contours modelů a zjednodušené verze MS modelu bývají implementovány právě pomocí této metody, což bude i případ implementace active contours without edges v této práci.

Level set metoda

Ústřední myšlenka spočívá v tom, že namísto parametrizace křivky v rovině je možné tuto křivku popsat jako nulovou hladinu jiné funkce. Mějme například parametrizaci kružnice $c = \{r(\cos p, \sin p), p \in [0, 2\pi]\}$, pak můžeme položit $\phi(x, y) = x^2 + y^2 - r^2$ a tentýž kruh popsat jako $c = \{(x, y), \phi(x, y) = 0\}$. Funkce ϕ se pak nazývá level setová funkce. Výhod tohoto přístupu je hned několik. Máme-li variační model založený na energii křivky (například všechny active contours modely z předchozí kapitoly), pak Euler–Langrangeova rovnice obsahuje derivaci křivky (nikoliv parametrizace), což je pracné přenést do numerické implementace, kde pracujeme s konkrétní parametrizací. V případě level setové funkce tento problém odpadá, neboť funkci ϕ můžeme snadno diskretizovat standardním způsobem na pravidlné pevné mřížce a přitom bude stále reprezentovat správnou křivku. Další často zmiňovanou (a podstatnou) výhodou je snadná změna topologie křivky (tzn. například rozdělení jedné uzavřené křivky na dvě uzavřené křivky, nebo změna uzavřené křivky na neuzavřenou, která vede z jednoho bodu $\partial\Omega$ do jiného). Pro parametrizovanou křivku je taková změna opět implementačně náročná, zatímco u funkce ϕ se o ni vůbec nemusíme starat.

Chceme-li této metody využít, je potřeba Euler–Langrangeovu rovnici energetického funkcionálu vyjádřit pomocí funkce ϕ . Zde máme dvě možnosti, buď vyjdeme z rovnice pro křivku c a převedeme ji na rovnici pro level setovou funkci ϕ , nebo vyjádříme energii funkcionálu přímo pomocí funkce ϕ a teprve potom ho zderivujeme. Ukážeme si nyní první možnost, v další části uvidíme druhou možnost.

Máme počáteční křivku $c_0 = c_0(0, p)$ a metodou největšího spádu chceme nalézt minimalizační křivku $c(t, p)$. Nechť je pak evoluční rovnice křivky v umělém čase t

popsána jako

$$\frac{\partial c}{\partial t} = F \mathbf{n}, \quad (3.1)$$

kde F je obecná funkce „síly“ a \mathbf{n} je normálový vektor křivky. To je velice časté, například geodesic active contours z minulé kapitoly jsou popsány rovnicí

$$\frac{\partial c}{\partial t} = (K g - \nabla g \cdot \mathbf{n}) \mathbf{n},$$

připomínáme že g je edge-detector function, K je křivost křivky. Nebo zmíněný mean-curvature motion, pohyb zmenšující délku křivky, je popsán rovnicí

$$\frac{\partial c}{\partial t} = K \mathbf{n},$$

viz (2.12). Položme tedy nyní křivku c jako množinu $\phi(x, y) = 0$ a přepišme rovnici (3.1) pro $\phi(t)$. Pro všechny časy t můžeme psát

$$\phi(t, c(t, p)) = 0.$$

Derivací podle času dostaneme

$$\frac{\partial \phi}{\partial t} + \left\langle \nabla \phi, \frac{\partial c}{\partial t} \right\rangle = 0.$$

Dosazením z (3.1) dostáváme

$$\frac{\partial \phi}{\partial t} + \langle \nabla \phi, F \mathbf{n} \rangle = 0. \quad (3.2)$$

Nyní si lze uvědomit, že $\mathbf{n} = \lambda \nabla \phi$ pro nějaké λ , neboli $\nabla \phi$ je (ne nutně jednotkový) normálový vektor ke křivce. Proč tomu tak je: Křivka $c = (x(s), y(s))$, kde s je délkový parametr, je určena rovnicí $\phi(x(s), y(s)) = 0$, pak zderivováním podle s máme

$$\phi_x x'(s) + \phi_y y'(s) = 0, \quad (3.3)$$

neboli vektor $(x'(s), y'(s))$, což je tečný vektor křivky, je kolmý na $\nabla \phi$ (ϕ_x apod. značí parciální prostorové derivace). Vraťme se k (3.2) a dosaďme za $\mathbf{n} = -\nabla \phi / |\nabla \phi|^1$, dostáváme finální level setovou rovnici

$$\frac{\partial \phi}{\partial t}(t, x) = F |\nabla \phi(t, x)|. \quad (3.4)$$

Rovnice je odvozena pro (x, y) na křivce c , princip metody však spočívá v tom, že ji zobecníme na celou oblast Ω a máme tak evoluční rovnici pro ϕ , která v každém časovém okamžiku t nese plnou informaci o křivce c , aniž bychom se museli starat o jakoukoliv parametrizaci. Podmínkou je, že silové pole F potřebujeme znát i mimo křivku c (v praxi alespoň v nějakém pásu na okolí $\phi(x, y) = 0$).

Jiný tvar, se kterým je možné se setkat, je

$$\frac{\partial \phi}{\partial t}(t, x) = F \delta(\phi(t, x)), \quad (3.5)$$

kde δ je Diracova funkce (distribuce). Evoluce je tak efektivně omezena skutečně jen na křivku c . Tento tvar přímo vychází z integrální formulace energie pomocí level setové funkce a následným zderivováním. V numerické implementaci je tak někdy zaměňováno

¹Znaménko záleží na orientaci normály \mathbf{n} a volbě ϕ vs. $-\phi$, tedy například zda je ϕ kladná či záporná uvnitř uzavřené křivky. Zde používáme nejčastěji uváděný tvar.

$|\nabla\phi|$ a $\delta(\phi)$ kvůli rozdílnému chování, podle toho, co je pro danou aplikaci vhodnější. Namísto distributivní δ funkce se v praxi pochopitelně používá vhodná aproximace δ_ε , uvidíme dále.

Ukázali jsme, že jednotkový normálový vektor křivky lze pomocí level setové funkce vyjádřit jako $\mathbf{n} = \pm \frac{\nabla\phi}{|\nabla\phi|}$. Je ještě další důležitá veličina, křivost křivky, kterou lze snadno vyjádřit pomocí level setové funkce. Konkrétně

$$K = \operatorname{div} \left(\frac{\nabla\phi}{|\nabla\phi|} \right),$$

neboli divergence normálového vektoru (znaménko opět závisí na orientaci). Odvození je podobné jako v případě normálového vektoru, konkrétně rovnici (3.3) zderivujeme ještě jednou podle s

$$x''\phi_x + \phi_{xx}(x')^2 + 2\phi_{xy}x'y' + y''\phi_y + \phi_{yy}(y')^2 = 0.$$

Víme, že $\mathbf{n} = (-y', x') = \lambda\nabla\phi = \lambda(\phi_x, \phi_y)$, dosadíme to do rovnice

$$\lambda^2 \left((\phi_y)^2 \phi_{xx} + (\phi_x)^2 \phi_{yy} - 2\phi_x \phi_y \phi_{xy} \right) + \frac{1}{\lambda} (y''x' - x''y') = 0.$$

Člen v poslední závorce je již (až na případné znaménko) křivost K (viz (2.10)), protože $(x', y')^2 = 1$ při parametrizaci délkou. Z toho důvodu je i $\lambda^2 = \frac{1}{|\nabla\phi|}$, dosazením tak máme (až na případné znaménko)

$$K = \frac{(\phi_y)^2 \phi_{xx} + (\phi_x)^2 \phi_{yy} - 2\phi_x \phi_y \phi_{xy}}{|\nabla\phi|^3} = \operatorname{div} \left(\frac{\nabla\phi}{|\nabla\phi|} \right),$$

což lze již vidět snadno rozderivováním pravé strany.

Viděli jsme tedy, že za jednoduchou myšlenkou se skrývá velice efektivní způsob, jak popsat a hlavně numericky implementovat evoluci křivky, přičemž neztratíme možnost poměrně snadno vyjádřit základní geometrické vlastnosti křivky. Namísto rovnice (3.1) použijeme (3.4) nebo (3.5) a funkci ϕ můžeme diskretizovat mnohem snáze na pevné mřížce. Změna však není úplně zadarmo. Stabilní diskretizace rovnice (3.4) může být často poměrně složitá a dále je třeba se starat o to, aby funkce ϕ měla stále rozumný tvar. U difusních rovnic se totiž stává, že funkce ϕ se v čase postupně zplošťuje ($|\nabla\phi| \rightarrow 0$) a určení křivky c jako $\phi(x, y) = 0$ pak může být zatíženo velikou chybou, nebo to bude zcela nemožné. Naopak (zejména při použití (3.5)) může být ϕ čím dál strmější ($|\nabla\phi| \rightarrow \infty$), což v (3.4) způsobí nestabilitu a v (3.5) zastaví vývoj křivky, protože žádné body (x, y) nebudou v rozmezí šířky aproximace δ funkce. Čas od času (každých několik iterací) je tak potřeba provést tzv. reinicializaci, kdy se aktuální funkce ϕ nahradí jinou funkcí, která má stejnou nulovou hladinu, ale stabilnější tvar. Vhodným kandidátem je funkce měřící (kladnou resp. zápornou) vzdálenost od křivky c (*signed distance function*), taková funkce má skoro všude $|\nabla\phi| = 1$, což je výhodné. „Standardní“ postup reinicializace tedy vypadá tak, že namísto řešení rovnice související s minimalizačním problémem se jednou za několik iterací vyřeší rovnice

$$\begin{aligned} \frac{\partial\psi}{\partial t} + \operatorname{sign}(\psi)(|\nabla\psi| - 1) &= 0, \\ \psi(0) &= \phi, \end{aligned} \tag{3.6}$$

kde $\operatorname{sign}(\cdot)$ je znaménková funkce, nulová v nule. Stacionární řešení ψ této rovnice má (v ideálním případě) stejnou nulovou hladinu jako ϕ , přičemž $|\nabla\psi| = 1$ s.v., lze pak položit $\phi := \psi$ a pokračovat v řešení minimalizačního problému. Tato rovnice se objevuje v různých variantách, které zahrnují zejména přesný tvar znaménkové funkce

(používá se různě aproximovaná) nebo dokonce záměnu $\text{sign}(\psi) \leftrightarrow \text{sign}(\phi)$, píšeme proto „standardní“ postup do uvozovek, neboť nejde o jedinou ustálenou možnost. Kromě řešení reinicializační rovnice existují i zcela jiné možnosti, jednu z nich uvidíme v jedné z dalších částí při implementaci active contours without edges. Jde zhruba o to, že k energii pocházející z variačního problému přidáme člen, který nutí křivku, aby měla gradient přibližně rovný jedné.

Pro účely této práce nám toto stručné shrnutí level set metody stačí, pro hlubší seznámení (mj. ve více dimenzích) lze doporučit například [29, 27, 28].

Formulace problému

V minulé kapitole byla diskuze o restrikci MS modelu na aproximaci po částech konstantními funkcemi, neboli minimalizaci

$$E^C(\Gamma) = \sum_i \int_{\Omega_i} (u_0 - c_i)^2 dx + \alpha \mathcal{H}^1(\Gamma), \quad (2.18)$$

kde Ω_i jsou souvislé celky Ω určené množinou hran Γ a konstanty c_i jsou střední hodnoty u_0 na jednotlivých Ω_i . Několikrát zmiňovaný segmentační model active contours without edges, kterým se v této kapitole zabýváme, lze formulovat velice jednoduše: Jedná se o minimalizaci E^C pro dělení Ω na právě dvě Ω_1, Ω_2 , tedy $\Omega = \Omega_1 \cup \Omega_2 \cup \Gamma$. Ačkoliv byl model objeven jiným způsobem, jedná se o takto jednoduchou restrikci MS funkcionálu (rozšíření na vyšší ale pevný počet Ω_i je pak již poměrně triviální).

Formulace pochází z článku [8], 1999, kde byl model formulován jako varianta active contours modelů. Princip active contours je tradičně takový, že počáteční křivka obklopující segmentovaný objekt se postupně smršťuje, dokud se nedostane na hranici objektu, kde se evoluce zastaví (lze postupovat i obráceně rozpínáním křivky). Protože hranu objektu se nepovede vždy určit přesně, snadno se může stát, že křivka hranu přeskočí a smršťuje se dále. V článku [8] autoři T. F. Chan a L. A. Vese formulovali podobný model, který však nespolehá na zastavení evoluce hranou objektu, ale pracuje s celkovou datovou informací uvnitř a vně křivky (tj. v Ω_1 a Ω_2), což již není lokální informace jakou jsou hrany, a křivka tak není náchylná na „přeskočení“, odtud pochází název metody. Efektivně se však jedná o výše zmíněnou restrikci MS segmentačního modelu. Další rozšíření modelu lze najít v článkách [9] (větší pevný počet celků Ω_i), [10] (barevné obrázky) a [11] (použití jiných dekriptorů pro segmentaci, než je přímo intenzita, tzn. např. segmentaci na základě textury). Tato rozšíření jsou z praktického hlediska užitečná, ale nepřinášejí žádnou podstatnou teoretickou či implementační náročnost, budeme se tak primárně zabývat jen prvním zmíněným modelem.

S formulací (2.18) jsme již dobře seznámeni, formulujme tedy nyní energetický funkcionál přímo v řeči level setové funkce. Vstupní obraz u_0 na Ω chceme rozdělit na (nejvýše) dvě části oddělené křivkou Γ . Tato křivka nemůže obsahovat žádná dělení (křížovatky) na tři či více křivek, protože souvislé celky $\Omega \setminus \Gamma$ jsou jen dva a po částech konstantní model neobsahuje žádné „zbytečné“ hrany ve smyslu definice 2.23 a věty 2.24. Γ tak může být složena z několika navzájem izolovaných komponent, kde každou komponentu tvoří jednoduchá uzavřená křivka, nebo jednoduchá křivka vedoucí o jednoho bodu $\partial\Omega$ ke druhému. Z výsledků o regularitě řešení po částech konstantního MS modelu lze tyto křivky předpokládat C^1 hladké. Zavedme tedy level setovou funkci $\phi : \Omega \rightarrow \mathbb{R}$ tak, že $\Gamma = \{(x, y), \phi(x, y) = 0\}$, jak jsme uvažovali v předchozí části. Γ pak rozděluje množinu Ω na dvě části, kde na jedné bude $\phi(x, y) > 0$ a na druhé $\phi(x, y) < 0$. Označme je Ω_I jako vnitřek Γ a Ω_O jako vnějšek Γ . V případě že Γ není uzavřená tak tyto množiny ozačují prostě jednu resp. druhou stranu Γ , přičemž tyto strany jsou pevně dané znaménkem ϕ . Pro pořádek tak shrňme, že v tomto případě

je $\mathbf{n} = -\frac{\nabla\phi}{|\nabla\phi|}$ jednotkový vnější normálový vektor křivky a $K = \operatorname{div}\left(\frac{\nabla\phi}{|\nabla\phi|}\right)$ je křivost, která je záporná v místě, kde je vnitřek Ω_I křivky lokálně konvexní a kladná v místě, kde je vnějšek Ω_O křivky lokálně konvexní. Konkrétně — je-li Γ kružnice, pak \mathbf{n} míří pryč od středu kružnice a K je na celé křivce záporná.

Chceme pomocí ϕ přepsat energetický funkcionál našeho modelu, označme jej F . Máme zatím

$$F(\Gamma) = \lambda_1 \int_{\Omega_I} (u_0 - c_I)^2 dx + \lambda_2 \int_{\Omega_O} (u_0 - c_O)^2 dx + \alpha \int_{\Gamma} ds, \quad (3.7)$$

kde jsou přidány váhové koeficienty λ_1, λ_2 u aproximačních členů. Uvažujme Heavisideovu funkci H a Diracovu funkci (míru) δ definované standardně

$$H(x) = \begin{cases} 0 & x < 0, \\ 1 & x > 0, \end{cases} \\ \delta = H'.$$

Za pomoci těchto funkcí můžeme funkcionál F napsat jako

$$F(\phi) = \lambda_1 \int_{\Omega} H(\phi)(u_0 - c_I)^2 dx + \lambda_2 \int_{\Omega} (1 - H(\phi))(u_0 - c_O)^2 dx + \alpha \int_{\Omega} |\nabla H(\phi)|,$$

kde poslední integrál značí totální variaci $H(\phi)$, neboli totální variaci charakteristické funkce množiny $\{(x, y), \phi(x, y) > 0\}$, což je rovno délce Γ , neboli $\mathcal{H}^1(\{(x, y), \phi(x, y) = 0\})$ (viz (2.4) a (2.4)), kde pak S_u odpovídá právě nulové hladině funkce ϕ . Přepíšeme-li poslední rovnici pomocí δ funkce, máme konečný tvar funkcionálu jako

$$F(\phi) = \lambda_1 \int_{\Omega} H(\phi)(u_0 - c_I)^2 dx + \lambda_2 \int_{\Omega} (1 - H(\phi))(u_0 - c_O)^2 dx + \alpha \int_{\Omega} \delta(\phi)|\nabla\phi|. \quad (3.8)$$

Jak již bylo několikrát zmíněno, pro pevné Γ (resp. nyní pro pevné ϕ) jsou konstanty c_I resp. c_O jednoznačně určeny jako průměry u_0 na odpovídajících množinách. Pomocí level setové funkce a Heavisideovy funkce je můžeme zapsat jako

$$c_I = \frac{\int_{\Omega} H(\phi)u_0 dx}{\int_{\Omega} H(\phi)dx}, \quad c_O = \frac{\int_{\Omega} (1 - H(\phi))u_0 dx}{\int_{\Omega} (1 - H(\phi))dx}, \quad (3.9)$$

tam kde jsou jmenovatele zlomků nenulové (opačný případ by znamenal $\Gamma = \emptyset$ a tudíž $\Omega_I = \emptyset$ nebo $\Omega_O = \emptyset$ a v tom případě budeme odpovídající člen s c_I resp. c_O z funkcionálu F uvažovat nulový).

Pro numerickou implementaci i derivaci funkcionálu je obvyklé namísto distributivních funkcí H a δ uvažovat jejich hladké aproximace H_ε resp. δ_ε , pro které stále platí $H'_\varepsilon = \delta_\varepsilon$. Obě uvažujeme hladké, omezené, konkrétní tvary uvedeme v další části u popisu numerické implementace. Budeme značit F_ε funkcionál F , kde nahradíme H_ε za H a δ_ε za δ .

Spočítejme nyní Gâteauxovu derivaci funkcionálu F_ε . Pro obecný funkcionál

$$F(u) = \int_{\Omega} f\left(x, u, \frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}\right) dx, \quad f(x, u, \xi_1, \xi_2) : \Omega \times \mathbb{R}^3 \rightarrow \mathbb{R},$$

platí, že Gâteauxova derivace $F'(u)$ funkcionálu je dána vztahem (viz např. [14], 8.1.2)

$$\begin{aligned} F'(u) &= \frac{\partial f}{\partial u}\left(x, u, \frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}\right) - \sum_{i=1}^2 \frac{\partial}{\partial x_i} \left(\frac{\partial f}{\partial \xi_i}\left(x, u, \frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}\right) \right) \\ &= \frac{\partial f}{\partial u}\left(x, u, \frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}\right) - \operatorname{div} \nabla_{\xi} f\left(x, u, \frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}\right). \end{aligned}$$

Aplikujeme tuto formuli na náš případ. Derivace prvních dvou členů je zřejmá, uved'eme však derivaci posledního členu, kde máme pod integrálem $\delta_\varepsilon(\phi)|\nabla\phi|$, což odpovídá

$$f(\phi, \xi_1, \xi_2) = \delta_\varepsilon(\phi)\sqrt{\xi_1^2 + \xi_2^2}.$$

Derivací po vzoru předchozí rovnice dostaneme v tomto členu

$$\frac{\partial f}{\partial \phi} = \delta'_\varepsilon(\phi)|\nabla\phi|, \quad \nabla_\xi f = \delta_\varepsilon(\phi)\frac{\nabla\phi}{|\nabla\phi|}.$$

S využitím snadno ověřitelné formule $\operatorname{div}(u\mathbf{v}) = \nabla u \cdot \mathbf{v} + u \operatorname{div} \mathbf{v}$ pro $u = \delta_\varepsilon(\phi)$ a $\mathbf{v} = \nabla\phi/|\nabla\phi|$ tak můžeme napsat

$$\frac{\partial f}{\partial \phi} - \operatorname{div} \nabla_\xi f = \delta'_\varepsilon(\phi)|\nabla\phi| - \delta'_\varepsilon(\phi)\frac{|\nabla\phi|^2}{|\nabla\phi|} - \delta_\varepsilon(\phi) \operatorname{div} \left(\frac{\nabla\phi}{|\nabla\phi|} \right) = -\delta_\varepsilon(\phi) \operatorname{div} \left(\frac{\nabla\phi}{|\nabla\phi|} \right).$$

Pro celkovou derivaci pak máme

$$F'_\varepsilon(\phi) = \delta_\varepsilon(\phi) \left(-\alpha \operatorname{div} \left(\frac{\nabla\phi}{|\nabla\phi|} \right) + \lambda_1(u_0 - c_I)^2 - \lambda_2(u_0 - c_O)^2 \right). \quad (3.10)$$

Numericky tedy budeme řešit evoluční problém (pohyb proti směru gradientu)

$$\begin{aligned} \frac{\partial \phi}{\partial t} &= \delta_\varepsilon(\phi) \left(\alpha \operatorname{div} \left(\frac{\nabla\phi}{|\nabla\phi|} \right) - \lambda_1(u_0 - c_I)^2 + \lambda_2(u_0 - c_O)^2 \right), \\ \frac{\partial \phi}{\partial \mathbf{n}} &= 0 \text{ na } \partial\Omega, \\ \phi(0) &= \phi_0, \end{aligned} \quad (3.11)$$

kde počáteční funkci ϕ_0 zvolíme nějakým vhodným způsobem, buď bez zřetele na segmentační problém, nebo (pro urychlení konvergence do minima) tak, aby se její nulová hladina přibližně shodovala s hranicí segmentovaného objektu (máme-li o něm apriorní informaci).

Řešením segmentace tak pro nás bude stacionární stav této rovnice, neboli ϕ splňující (3.10), segmentovaný objekt Ω_I , pozadí Ω_O a hranice mezi nimi Γ pak budou určeny po řadě jako množiny bodů (x, y) , kde $\phi(x, y) > 0$ resp. $\phi(x, y) < 0$ resp. $\phi(x, y) = 0$.

Existence řešení

V této části se budeme věnovat otázce existence řešení segmentačního modelu active contours without edges. Odpověď nás nepřekvapí, protože model je speciálním případem po částech konstantní segmentace, jejíž existence byla dokázána například v [25], část 5, nebo v samostatném článku [23]. Navíc, i tento model je opět jen speciálním případem MS segmentačního modelu, u něhož jsme existenci řešení podrobně zmapovali v předchozí kapitole. Přesto není na škodu se před numerickou implementací na existenci znovu podívat, mimo jiné proto, že zde budou dobře zúročeny poznatky z předchozí kapitoly.

Nejprve si uvědomme, že funkcionál F z (3.8) lze snadno přepsat pomocí charakteristických funkcí. Je-li χ charakteristická funkce Ω_I , pak

$$F(\phi) = G(\chi) = \lambda_1 \int_{\Omega} \chi(x)(u_0 - c_I)^2 dx + \lambda_2 \int_{\Omega} (1 - \chi(x))(u_0 - c_O)^2 dx + \alpha \int_{\Omega} |\nabla\chi|,$$

kde poslední integrál opět označuje totální variaci funkce χ , která je shodná (pokud Ω_I má \mathcal{H}^1 -měřitelnou hranici) s obvodem Ω_I neboli délkou její hranice. Tento fakt jsme již využili při formulaci F pomocí Heavisideovy funkce, což je totéž. S tímto pozorováním můžeme formulovat následující existenční větu.

Věta 3.1 *Nechť je $\Omega \subset \mathbb{R}^2$ otevřená omezená s lipschitzovskou hranicí, $u_0 \in L^\infty(\Omega)$. Pak existuje aspoň jedno řešení $u \in BV(\Omega)$, $u(x) \in \{0, 1\}$ minimalizující funkcionál*

$$G(v) = \lambda_1 \int_{\Omega} v(x)(u_0 - c_I)^2 dx + \lambda_2 \int_{\Omega} (1 - v(x))(u_0 - c_O)^2 dx + \alpha \int_{\Omega} |\nabla v|,$$

$$v \in BV(\Omega), v(x) \in \{0, 1\} \forall x \in \Omega.$$

Je jasné, že takové řešení u představuje i řešení minimalizace F .

Důkaz Nechť $\{u_n\} \subset BV(\Omega)$ je posloupnost charakteristických funkcí minimalizujících G , tedy $\lim_n G(u_n) = \inf_u G(u)$. Pak $\{u_n\}$ je omezená v $BV(\Omega)$, neboť $\|u_n\|_{L^1} < |\Omega|$ a $\int_{\Omega} |\nabla u_n| \leq G(\mathbf{0}) < \infty$ pro s.v. n , kde $\mathbf{0}$ označuje identicky nulovou funkci. Pak víme (věta 2.6, část 5), že existuje $u \in L^1(\Omega)$ taková, že (až na vybranou posloupnost) $u_n \xrightarrow{L^1} u$. Dále víme (věta 2.6, část 2.3), že

$$\int_{\Omega} |\nabla u| \leq \liminf_n \int_{\Omega} |\nabla u_n|,$$

z čehož mimo jiné plyne, že $u \in BV(\Omega)$. Poslední člen funkcionálu G je tedy zdola polospojité v topologii L^1 . Konstanty c_I^n resp. c_O^n se zdůrazněnou závislostí na n jsou dané vztahy (3.9), resp. v řeči charakteristických funkcí jako

$$c_I^n = \frac{\int_{\Omega} u_n(x)u_0(x) dx}{\int_{\Omega} u_n(x) dx}, \quad c_O^n = \frac{\int_{\Omega} (1 - u_n(x))u_0(x) dx}{\int_{\Omega} (1 - u_n(x)) dx},$$

ty jsou zjevně spojitě v L^1 topologii, tedy $c_I^n \rightarrow c_I$, analogicky c_O^n . Nakonec člen $\int_{\Omega} (c_I^n - u_0)^2 dx$ je také spojitý, neboť

$$\left| \int_{\Omega} ((c_n - u_0)^2 - (c - u_0)^2) dx \right| \leq |\Omega| |c_n^2 - c^2| + 2\|u_0\|_{L^\infty} |\Omega| |c_n - c| \rightarrow 0.$$

Ostatní členy G jsou tedy dokonce spojitě a z přímé metody variačního počtu, která byla vysvětlena v předchozí kapitole v části o existenci řešení, plyne, že u je řešením minimalizace G na množině charakteristických funkcí z $BV(\Omega)$. \square

Existenci analytického řešení tak máme potvrzenou, jednoznačnost samozřejmě stále neplatí.

Numerická implementace

Dostáváme se konečně k numerické implementaci segmentačního modelu daného funkcionálem F , původně formulovaným jako (3.7) a pomocí level setové funkce přeformulovaným na (3.8). Minimum budeme hledat metodou největšího spádu, řešíme tak evoluční rovnici

$$\frac{\partial \phi}{\partial t} = \delta_\varepsilon(\phi) \left(\alpha \operatorname{div} \left(\frac{\nabla \phi}{|\nabla \phi|} \right) - \lambda_1 (u_0 - c_I)^2 + \lambda_2 (u_0 - c_O)^2 \right). \quad (3.11)$$

Nejprve upřesníme, jakou δ_ε aproximaci δ funkce budeme používat. V článku [8] jsou zmíněné dvě konkrétní aproximace Heavisideovy funkce, z nich pak plynou dvě možnosti aproximace δ funkce.

$$H_{1,\varepsilon}(s) = \begin{cases} 1 & s > \varepsilon, \\ 0 & s < -\varepsilon, \\ \frac{1}{2} \left(1 + \frac{s}{\varepsilon} + \frac{1}{\pi} \sin \left(\frac{\pi s}{\varepsilon} \right) \right) & |s| \leq \varepsilon \end{cases}$$

a

$$H_{2,\varepsilon}(s) = \frac{1}{2} \left(1 + \frac{2}{\pi} \arctan \left(\frac{s}{\varepsilon} \right) \right).$$

Hlavní rozdíl je v tom, že zatímco $H_{1,\varepsilon}$ je nenulová jen v ε okolí nuly, $H_{2,\varepsilon}$ má nosič na celém \mathbb{R} , totéž pak platí o jejich derivacích, tedy $\delta_{1,\varepsilon}$ a $\delta_{2,\varepsilon}$. Autoři [8] doporučují použití $\delta_{2,\varepsilon}$ s vysvětlením, že v prvním případě je vývoj level setové funkce omezen jen na úzké okolí křivky (nulové hladiny), takže i když křivka „cítí“ data z u_0 v celé ploše obrazu (prostřednictvím c_I a c_O), reagovat může jen lokálně ve svém okolí. Je-li pak na obraze objekt, který není jednoduše souvislý (např. mezikružší), pak obepíná-li křivka tento objekt, nemůže správně vysegmentovat „díru“ uprostřed, neboť tam již nedosáhne vliv δ funkce, která omezuje vývoj křivky (bylo by to možné, pokud by level setová funkce ϕ byla hodně plochá, tomu však za prvé zabráňuje reinicializace, o které jsme se zmínili, a za druhé není obecně dobré pracovat s plochou level setovou funkcí, protože tím mj. vzniká chyba při určení nulové hladiny a dále uvidíme, že metoda má pak i horší stabilitu).

Experimentoval jsem s oběma zmíněnými možnostmi δ funkcí a lze jen souhlasit s tímto doporučením. Pokud však v rovnici položíme $\lambda_1 = \lambda_2 = 0$, pak nám zbyde jen $\phi_t = \delta(\phi) \operatorname{div}(\nabla\phi/|\nabla\phi|)$, což je několikrát zmíněný mean curvature motion, pohyb nejvíce zkracující délku křivky. V tomto konkrétním případě se mi funkce $\delta_{1,\varepsilon}$ osvědčila lépe, neboť lépe lokalizuje změny jen na úzké okolí křivky, kde je má funkce ϕ vysoký gradient a diskretizace rovnice je v tom případě stabilnější (uvidíme). V konečné numerické implementaci je tedy použito

$$\frac{\partial\phi}{\partial t} = \delta_{1,\varepsilon}(\phi)\alpha \operatorname{div} \left(\frac{\nabla\phi}{|\nabla\phi|} \right) + \delta_{2,\varepsilon}(\phi) \left(-\lambda_1(u_0 - c_I)^2 + \lambda_2(u_0 - c_O)^2 \right). \quad (3.12)$$

Pro diskretizaci této rovnice je použita pravidelná mříž, kde každý uzel odpovídá jednomu pixelu vstupního obrazu, rozměr úlohy je tak rovný rozměru vstupního obrazu. Funkci $\phi(t, x, y)$ aproximujeme diskrétní verzí $\phi_{i,j}^n \approx \phi(n\Delta t, i, j)$, $i \in \{1, 2, \dots, N\}$, $j \in \{1, 2, \dots, M\}$, $n \in \{1, 2, \dots\}$. Je tedy uvažována vzdálenost uzlů $\Delta x = \Delta y = h = 1$. Označme centrální diference

$$\Delta^x \phi_{i,j}^n = \frac{\phi_{i+1,j}^n - \phi_{i-1,j}^n}{2} \quad \text{a} \quad \Delta^y \phi_{i,j}^n = \frac{\phi_{i,j+1}^n - \phi_{i,j-1}^n}{2}.$$

Rovnice (3.12) je pak diskretizována explicitním schématem

$$\begin{aligned} & \frac{\phi_{i,j}^{n+1} - \phi_{i,j}^n}{\Delta t} = \\ & = \delta_{1,\varepsilon}(\phi_{i,j}^n)\alpha \left(\Delta^x \left(\frac{\Delta^x \phi_{i,j}^n}{\sqrt{(\Delta^x \phi_{i,j}^n)^2 + (\Delta^y \phi_{i,j}^n)^2}} \right) + \Delta^y \left(\frac{\Delta^y \phi_{i,j}^n}{\sqrt{(\Delta^x \phi_{i,j}^n)^2 + (\Delta^y \phi_{i,j}^n)^2}} \right) \right) \\ & + \delta_{2,\varepsilon}(\phi_{i,j}^n) \left(-\lambda_1(u_{0,i,j} - c_I)^2 + \lambda_2(u_{0,i,j} - c_O)^2 \right), \end{aligned} \quad (3.13)$$

kde $u_{0,i,j} = u_0(i, j)$.

Počáteční funkce $\phi_{i,j}^0 = \phi_0(i, j)$ je volena po částech konstantní s hodnotami

$$\phi_0(x, y) = \begin{cases} C & \text{uvnitř křivky,} \\ -C & \text{vně křivky,} \end{cases} \quad (3.14)$$

kde konstantu $C > 0$ je potřeba zvolit s ohledem na volbu ε v aproximaci δ funkce funkcemi $\delta_{1,\varepsilon}$ a $\delta_{2,\varepsilon}$. Gradient funkce ϕ již po první iteraci nebude nekonečný, ale zůstane

vysoký, což je dokonce i výhodné (uvidíme dále). Je-li pak C vysoké, bude jen velice málo bodů (v diskrétním případě i žádné), pro které je $|\phi(x, y)| < \varepsilon$ a tedy časový vývoj funkce ϕ bude velice pomalý nebo žádný. Při volbě nízkého C je situace zřejmě opačná, vývoj není koncentrován na okolí křivky, jak předepisuje integrální formulace (3.11), mimo jiné se pak pracuje s místy, kde je $|\nabla\phi|$ nízký, což vede k nestabilitě. V numerických experimentech je voleno $C = 2$, což se osvědčilo. Lze doporučit zhruba $C \in [2, 10]$ jako vhodný kompromis mezi rychlostí a stabilitou.

Zde je na místě zmínit, že výsledek segmentace v principu nezávisí na kontrastu vstupního obrazu u_0 v tom smyslu, že když provedeme transformaci $u_0 \rightarrow \theta u_0$, je to totéž, jako bychom provedli $\lambda_{1,2} \rightarrow \theta^2 \lambda_{1,2}$, to je snadno vidět dosazením. Není tedy možné takto uměle zlepšovat kvalitu segmentace (ačkoliv opticky působí pro $\theta > 1$ obraz kontrastněji). Je ale vhodné uvažovat vstupní obraz nějak normovaný, aby nebylo nutné právě změnami $\lambda_{1,2}$ kompenzovat špatný kontrast, je tak vždy předpokládáno $u_0(x, y) \in [0, 1]$ s tím, že funkce je na.

Funkci ϕ_0 lze vybrat s ohledem na segmentovaný objekt, tzn. například aby vnitřek objektu zhruba odpovídal bodům, kde $\phi(x, y) > 0$, což podstatně urychlí konvergenci do minima (tj. sníží počet nutných iterací). V numerických experimentech je však většinou volena bez tohoto ohledu, aby se ukázalo, jak si metoda poradí se segmentací bez apriorní informace o segmentovaném objektu.

Konstanty c_I, c_O z rovnice (3.11) jsou v každém kroku spočteny pomocí vztahů (3.9), v numerické implementaci jsou spočteny s „pravou“ skokovou funkcí H , což je rychlejší a na řešení rovnice to nemá žádný vliv. Rozdíl je srovnatelný se zcela nepatrnou změnou konstant λ_1, λ_2 .

Algoritmus segmentace pak probíhá v následujících krocích

1. Načíst vstupní obrázek, provést případné roztážení histogramu, aby hodnoty intenzit ležely právě v $[0, 1]$, sestavit diskrétní mříž stejné velikosti a inicializovat ji počáteční funkcí $\phi_{i,j}^0$ podle (3.14).
2. Provést výpočet konstant c_I, c_O podle (3.9).
3. Provést výpočet jednotlivých členů na pravé straně (3.13) (k nim přibude ještě regularizační člen, o kterém bude řeč na konci této části).
4. Provést iteraci podle (3.13) s pevným časovým krokem Δt .
5. Dopočítat hodnoty ϕ v krajních bodech Ω podle Neumannovy okrajové podmínky.
6. Zobrazit aktuální výsledek.
7. Kroky 2 až 6 opakovat po stanovený počet iterací.

Podívejme se ještě na stabilitu metody. S datovými členy, tzn. typu $\phi_t = \delta_\varepsilon(\phi)\lambda(u_0 - c)^2$, není problém. Omezíme-li δ_ε funkci konstantou, tak se jedná v podstatě o časovou integraci typu

$$\phi^n = K n + \phi^0,$$

kde do konstanty K je zahrnuto omezení δ funkce, velikost časového kroku, konstanty λ a $\|u_0\|_{L^\infty}$. Věnujme se tedy disipativnímu členu, který zkracuje délku křivky, tzn. $\phi_t = \delta_\varepsilon(\phi)\alpha \operatorname{div} \left(\frac{\nabla\phi}{|\nabla\phi|} \right)$. Po vzoru [5], str. 64, je možné difuzní rovnici

$$\frac{\partial\phi}{\partial t} = \alpha \operatorname{div} \left(\frac{\varphi(|\nabla\phi|)}{|\nabla\phi|} \nabla\phi \right)$$

rozložit na tečnou část a normálovou část vzhledem ke směru, na kterém je ϕ konstantní (tedy kolmo na gradient a ve směru gradientu), konkrétně je pak

$$\frac{\partial \phi}{\partial t} = \underbrace{\left(\alpha \frac{\varphi(|\nabla \phi|)}{|\nabla \phi|} \nabla \phi \right)}_{=c_t} \frac{\partial^2 \phi}{\partial t^2} + \underbrace{\left(\alpha \varphi'(|\nabla \phi|) \right)}_{=c_n} \frac{\partial^2 \phi}{\partial \mathbf{n}^2},$$

kde \mathbf{t} resp. \mathbf{n} značí tečný resp. normálový vektor vzhledem k isočárám ϕ . Pro stabilitu explicitní diskretizace této rovnice pak máme obecnou podmínku

$$\Delta t \leq \frac{h^2}{4 \max\{|c_t|, |c_n|\}}.$$

Aplikujeme-li tento postup na náš případ, máme $\varphi \equiv 1$, $h = 1$, a $\delta_\varepsilon(\phi)$ aproximujeme omezující konstantou $|\delta_\varepsilon| \leq C_\varepsilon$, dostáváme na časový krok omezení

$$\Delta t \leq \frac{|\nabla \phi|}{4\alpha C_\varepsilon}, \quad \text{tedy } \Delta t \leq \frac{\inf |\nabla \phi|}{4\alpha C_\varepsilon}.$$

Časový krok byl ve všech případech stanoven pevně $\Delta t = 1$, podmínku stability pak dostáváme pro α

$$\alpha \leq \frac{\inf |\nabla \phi|}{4C_\varepsilon}.$$

Toto omezení je třeba mít při volbě konstanty α na paměti. Ze zkušenosti s numerickým experimentováním lze říct, že mírné porušení této podmínky ještě nevede k oscilacím v řešení. Vývoj ϕ je úzce koncentrován na okolí nulové hladiny, jednotlivé iterace tak působí v trochu jiných bodech $x \in \Omega$, takže případné chyby se nebudou příliš zvětšovat. Navíc, vzápětí uvidíme, že k rovnici je z jiného důvodu přidán ještě externí regularizační člen, který má také za následek eliminaci případných menších chyb vzniklých porušením této podmínky. Hrubé porušení však skutečně vede k prudkému nárůstu oscilačního šumu ve funkci ϕ a řešení přestane mít smysl. Při použití $\delta_{1,\varepsilon}$, kde je pro $\varepsilon = 1$ konstanta $C_\varepsilon = 1$, lze jako bezpečnou doporučit volbu $\alpha \in (0, 5]$. Tato podmínka mimo jiné ospravedlňuje některé poznatky z předchozí diskuze: $|\nabla \phi|$ je vhodné udržovat relativně vysoký v blízkosti křivky na nulové hladině ϕ (toto je na úkor rychlosti konvergence do minima, samozřejmě), volba $\delta_{1,\varepsilon}$ je z toho důvodu pro divergenční člen vhodnější než $\delta_{2,\varepsilon}$ neboť koncentruje vývoj ϕ jen na zmíněné okolí.

Reinicializace

Ke konci části o level setové metodě bylo zmíněno, že level setová funkce ϕ bývá vlivem evolučního procesu nežádoucím způsobem deformována tak, že to buď zhoršuje přesnost stanovení její nulové hladiny (sledované křivky Γ), nebo to negativně ovlivňuje další řešení a jeho stabilitu, často obojí. Jedním ze způsobů jak tento problém řešit je jednou za několik iterací vyřešit reinicializační rovnici (3.6), položit $\phi(t) := \psi$ a pokračovat v řešení rovnice zkoumaného problému. V případě segmentace pomocí active contours without edges není tento přístup ideální. Výhoda level setové segmentace spočívá mj. v tom, že segmentace není striktně lokalizována na křivku nulové hladiny, tzn. level setová funkce se může „prohnout“, změnit znaménko a způsobit tak vznik nové hrany a tedy změnu celé segmentace i relativně daleko od původní hranové křivky (tím je umožněna správná segmentace objektů, které nejsou jednoduše souvislé). Reinicializace ale každých několik iterací nutí level setovou funkci mít $|\nabla \phi| = 1$, což nutně omezuje tuto možnost detekce vnitřních hran objektů. Existují však i jiné možnosti, jak v průběhu řešení rovnice zachovávat vhodný tvar level setové funkce. V této práci je

při implementaci použita metoda z [22] nazvaná *distance regularized level set evolution*. Její princip spočívá v tom, že k segmentační energii F přidáme regularizační člen R a minimalizujeme modifikovaný funkcionál

$$F_R(\phi) = \mu R(\phi) + F(\phi),$$

kde $\mu > 0$ je váhový parametr a regularizační člen R má tvar

$$R(\phi) = \int_{\Omega} p(|\nabla\phi|) dx.$$

Funkce $p : [0, \infty) \rightarrow \mathbb{R}$ je vhodný potenciál, který upravuje $|\nabla\phi|$. Jak bylo řečeno, za ideální lze považovat, pokud $|\nabla\phi| \approx 1$, což je primární cíl reinitializace. Potenciál $p(s)$ je tedy zvolen tak, aby měl minimum právě pro $s = 1$, jednou z možností je pak například $p_1(s) = \frac{1}{2}(s - 1)^2$.

Spočteme-li Gâteauxovu derivaci nového funkcionálu F_R , máme

$$F'_R(\phi) = -\mu \operatorname{div} \left(\frac{p'(|\nabla\phi|)}{|\nabla\phi|} \nabla\phi \right) + F'(\phi).$$

Vidíme, že derivace nového členu R' představuje (po přidání do časové rovnice) difuzi s difuzním koeficientem $d(s) = \frac{p'(s)}{s}$ závislým na $s = |\nabla\phi|$. Znaménko tohoto členu určuje, zda se jedná o dopřednou ($d(s) > 0$) či zpětnou ($d(s) < 0$) difuzi. Pro potenciál p_1 je vidět, že $d(s) > 0$ pro $s > 1$, $d(s) < 0$ pro $s < 1$, tedy potenciál dopřednou či zpětnou difuzí udržuje level setovou funkci ϕ tak, aby $|\nabla\phi| = 1$. Nevýhodou však je, že $d(s) \rightarrow -\infty$ pro $s \rightarrow 0+$, neboli pokud je funkce ϕ konstantní, regularizační člen představuje zpětnou difuzi s neomezeným difuzním koeficientem. To je nepraktické jednak kvůli stabilitě, ale hlavně proto, že v numerické implementaci popsané výše je $|\nabla\phi| = 0$ hlavně v místech daleko od křivky na nulové hladině, kde tato nekontrolovaná difuze způsobuje v podstatě šum v jinak konstantní funkci a důsledkem bývá narušení nulové hladiny level setové funkce, což poškodí sledovanou křivku. V témže článku proto autoři navrhují použít jiný potenciál p_2 takový, který bude mít minimum nejen pro $s = 1$ ale i pro $s = 0$, takže bude-li se level setová funkce ϕ blížit konstantě tak na ni regularizace nebude mít vliv. Druhým přirozeným požadavkem je, aby odpovídající $d_2(s) = \frac{p'_2(s)}{s}$ byla omezená pro $s \in [0, \infty]$. Konkrétní navrženou volbou je

$$p_2(s) = \begin{cases} \frac{1}{(2\pi)^2} (1 - \cos(2\pi s)) & s < 1, \\ \frac{1}{2} (s - 1)^2 & s \geq 1. \end{cases}$$

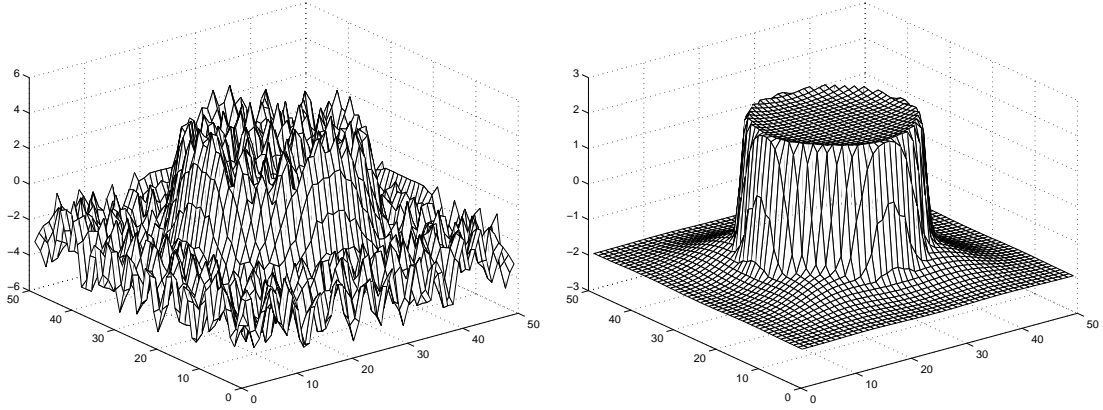
Pak je

$$d_2(s) = \begin{cases} \frac{1}{2\pi s} \sin(2\pi s) & s < 1, \\ 1 - \frac{1}{s} & s \geq 1 \end{cases}$$

a platí

- $d_2(s) \geq 0$ na $s \in [0, \frac{1}{2}] \cup [1, \infty]$ a $d_2(s) \leq 0$ na $s \in [\frac{1}{2}, 1]$,
- v limitách je $d_2(0) = 1$ a $d_2(\infty) = 1$,
- $|d_2(s)| \leq 1 \forall s$.

Tento potenciál tak produkuje difuzi s omezeným difuzním koeficientem, která je dopředná pro $|\nabla\phi| < \frac{1}{2}$, kde dále snižuje $|\nabla\phi|$ k nule, taktéž dopředná pro $|\nabla\phi| > 1$, kde snižuje $|\nabla\phi|$ k jedné, a zpětná v opačném případě, kdy zvedá $|\nabla\phi|$ zpět k jedné.



Obrázek 3.1: Level setová funkce po 250 iteracích stejné rovnice, vlevo při regularizaci s potenciálem p_1 , vpravo s potenciálem p_2 .

Rozdílný vliv těchto potenciálů na level setovou funkci je vidět na obrázku 3.1. Potenciál p_2 tak splňuje všechny naše předchozí požadavky a byl použit pro regularizaci level setové funkce v numerické implementaci active contours without edges v této práci. Diskretizace tohoto difuzního členu byla převzata z článku [22], je provedena analogickým explicitním schématem, jaké je použito pro divergenční člen v (3.13).

V citovaném článku [22] je provedena diskuze stability metody naprosto stejným způsobem jako v našem případě pro mean curvature motion (divergenční člen z (3.13)), díky omezenosti difuzního koeficientu lze podmínku stability určit jako

$$\mu\Delta t \leq \frac{1}{4}.$$

Kapitola 4

Experimentální výsledky

V této poslední části práce budou představeny výsledky praktických testů segmentace pomocí active countours without edges. Součástí práce je kód numerické implementace napsaný v MATLABu¹. Hlavním prvkem je funkce `acwe` v souboru `acwe.m`, která provádí iterace podle rovnice (3.13), nezobrazuje žádnou grafiku a veškeré chování je řízeno vstupními parametry, funkce je tak vhodná pro začlenění do externího projektu. Příložený jsou soubory `test_*.m`, kde každý odpovídá jednomu experimentálnímu testu, které budou vzápětí představeny. Jednotlivé ukázky je tak možné velice snadno rekonstruovat, stačí spustit odpovídající skript `test_*.m`, není nutná žádná kompilace či jiná příprava. Všechny použité obrázky jsou přiloženy v adresáři `/img`. Průběh každého testu je uložen jako videosoubor, tyto soubory jsou uloženy v adresáři `/videos`, takže i čtenář, který nemá k dispozici MATLAB si může ukázky segmentace bez problémů prohlédnout. Mechanismus ukládání segmentace jako video je součástí testovacích souborů, takže není problém jej převzít do jiného projektu. Nakonec, vybrané snímky z průběhu jednotlivých segmentací (mimo jiné celkový výsledek segmentace) jsou uloženy v adresáři `/frames` a označeny jménem testu a pořadovým číslem snímku. Zdrojový kód funkce `acwe` je uveden v příloze A.

Zbytek kapitoly bude tvořit přehled a diskuze jednotlivých testovacích ukázek, na konci kapitoly provedeme diskuzi celkového chování segmentace tímto modelem. U každého testu budou uvedeny použité hodnoty parametrů. Zde nebyla snaha pečlivou volbou dosáhnout ideálního chování, ale spíše hodnotami parametrů z běžných rozmezí (která tak lze z jednotlivých ukázek usoudit) vyzkoušet, zda a jak segmentace na daném snímku funguje. Pro všechny testy je volen časový krok $\Delta t = 1$, váha regularizace $\mu = 0.2$, šířka δ_ϵ funkcí $\epsilon = 1$ a váhy datových členů shodné $\lambda_1 = \lambda_2 = \lambda$, v praxi se tak měnily hodnoty váhy délkového členu α , datových členů λ a počet iterací.

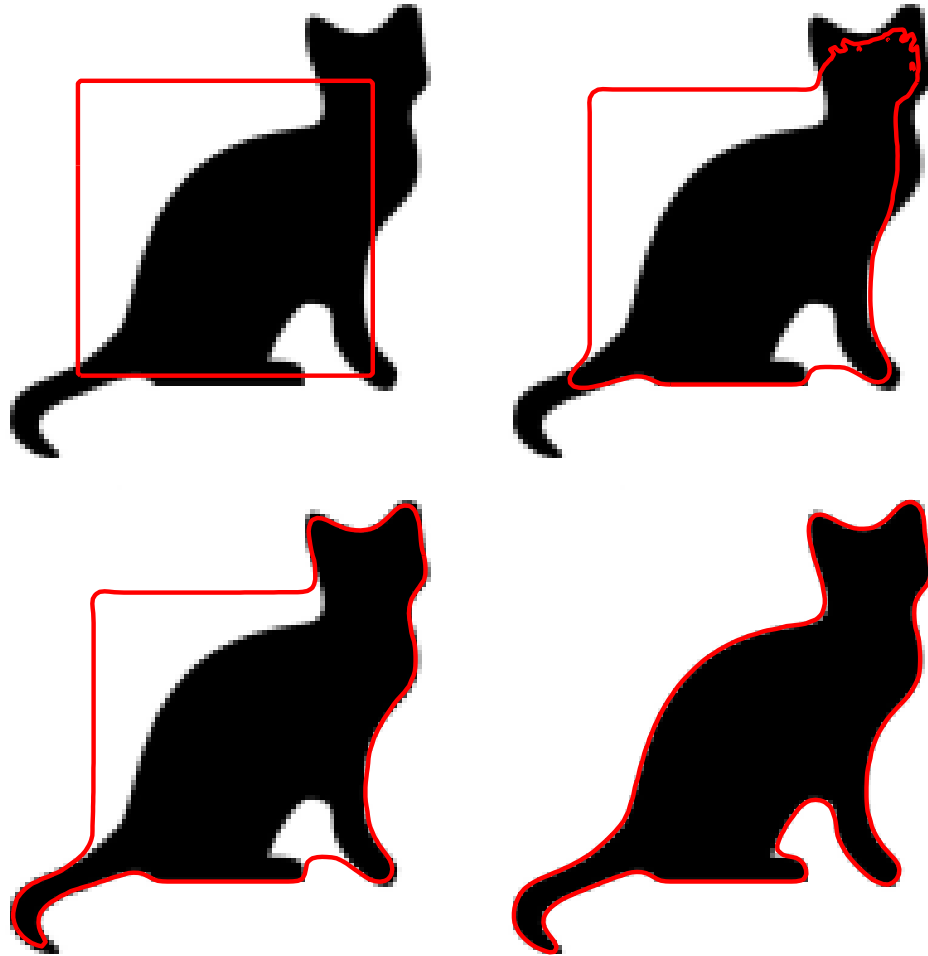
Každý představený test má své `jmeno`, spouští se skriptem `test_jmeno.m`, pracuje s obrázkem `/img/jmeno.bmp` a videosekvence průběhu segmentace je uložena v `/videos/jmeno.avi`. K popisu testu je připojen obrázek, na kterém je červenou křivkou, znázorňující nulovou hladinu level setové funkce a tedy křivku hran Γ , vyznačena je vždy počáteční poloha, konečná segmentace a mezitím několik snímků z průběhu segmentace.

Dílčí testy

Test simple, snadno segmentovatelný objekt

Obrázek 4.1. Jak lze očekávat, v takto jednoduchém případě funguje algoritmus bez problému, obrázek má jasně bimodální charakter (tím označujeme tmavý objekt a světlé

¹Zdrojový kód byl vyvinut a testován ve verzi R2010b, 7.11



Obrázek 4.1: Test `simple`, $\alpha = 0.1$, $\lambda = 1$, 80 iterací.

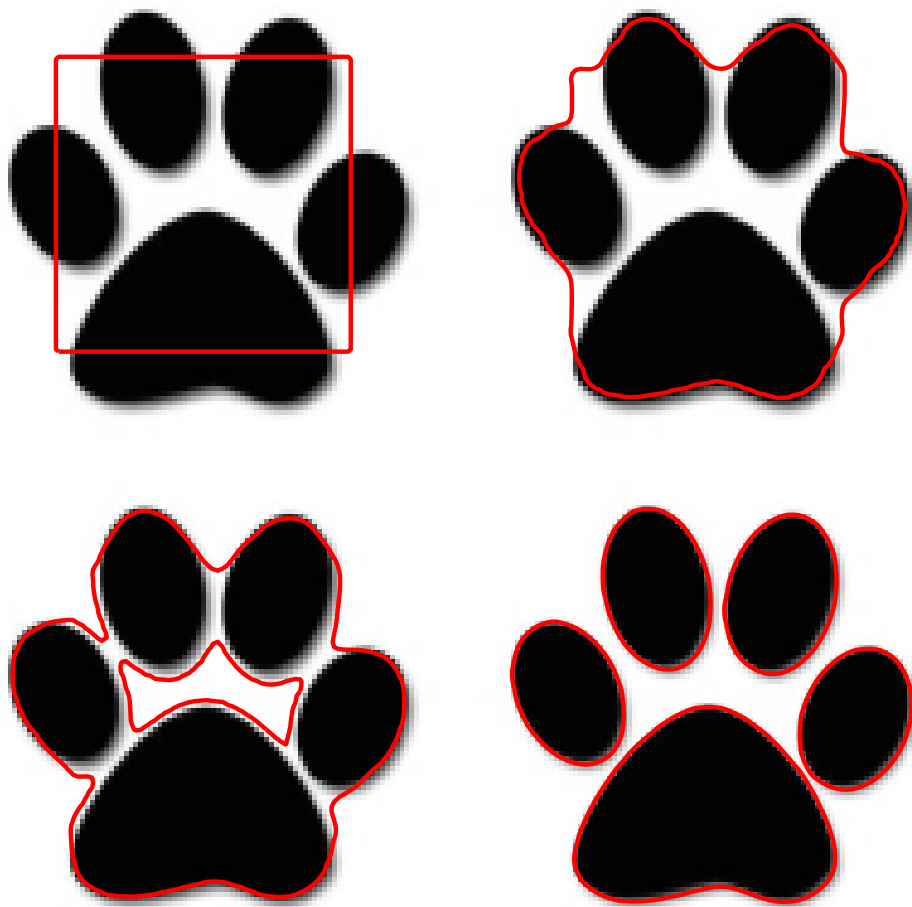
pozadí či obráceně) a to je pro segmentaci touto metodou ideální. Nastavením většího λ nebo Δt jde docílit ještě rychlejší konvergence.

Test topology, změna topologie křivky

Obrázek 4.2. Obrázek v tomto testu je opět ze segmentačního hlediska jednoduchý, slouží k demonstraci jedné ze zmiňovaných výhod level set metody. Počáteční křivka ϕ_0 je jednoduchá uzavřená křivka, zatímco konečná segmentace je složena z několika takových křivek.

Test edge1, umělý obrázek bez gradientních hran

Obrázek 4.3. Cílem tohoto testu je segmentace obrázku, na kterém objekt není ohraničen souvislými hranami rozpoznatelnými podle gradientu (poslední snímek ukazuje výstup Cannyho detektoru na vyhlazený obrázek). Z této představy model active contours without edges vychází, takže není překvapivé, že segmentace (na tomto umělém obrázku) proběhne dobře. Nejedná se však o segmentaci textury, jak by se mohlo zdát, protože „obličej“ na obrázku má jinou průměrnou hodnotu šedi než pozadí a to je jediná informace, která je pro segmentaci použita. Hranové metody jsou pro tento typ obrázku nevhodné, neboť i při značném rozmazání jsou hranice objektů nesouvislé, což znemožňuje správnou segmentaci. Mimochodem je v tomto testu vidět schopnost me-

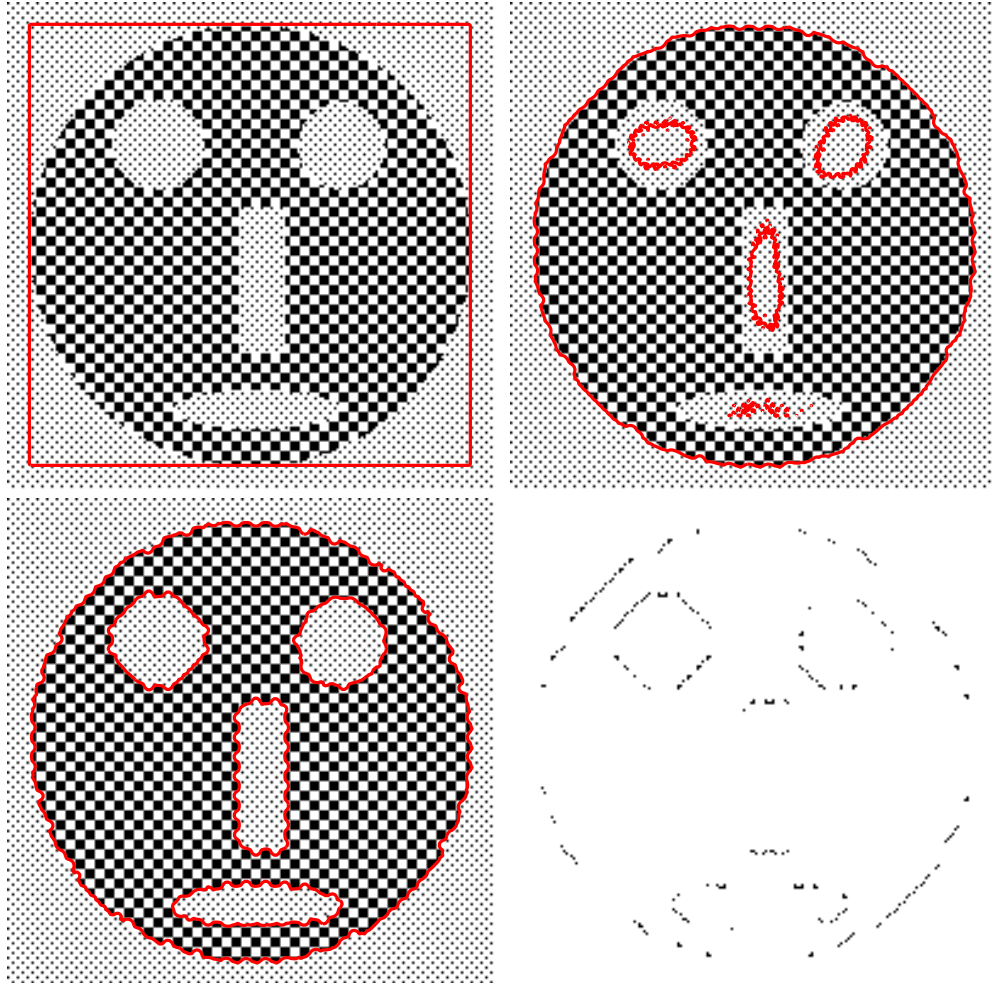


Obrázek 4.2: Test topology, $\alpha = 0.2$, $\lambda = 2$, 70 iterací.

tody detekovat vnitřní hrany objektu („oči, nos, pusou“), ačkoliv je od nich křivka zprvu daleko. Na tom se podílí mimo jiné použití funkce $\delta_{2,\varepsilon}$ u datového členu, při použití $\delta_{1,\varepsilon}$ je segmentace vnitřních okrajů umožněna jen pro vybrané hodnoty váhových parametrů a zejména vhodnou volbu počáteční křivky, čímž ale metoda značně ztrácí na samostatnosti.

Test edge2, reálný obrázek bez gradientních hran

Obrázek 4.4. Tento test má podobný charakter jako předchozí, použitý obrázek je však reálná fotografie. Galaxie na obrázku nemá zřetelné gradientní hrany, obrázek je však poměrně dobře bimodální, takže by metoda mohla fungovat. Vidíme však, že segmentace není zcela správná. Galaxie je v zhruba vysegmentována, ale chvosty vpravo dole a vlevo nahoře jsou uříznuté, neboť jsou příliš tmavé a zapadají do průměrné šedi pozadí. Metoda nebere ohled na souvislost vysegmentovaných objektů, takže nezohlední, že tyto části náležejí ke galaxii, zatímco podobně šedé pozadí na levé straně obrazu již ne. Dále je vidět, že spolu s galaxií jsou vysegmentované ještě tři hvězdy, což je logické, protože tvoří jedny z nejsvětlejších bodů obrázku. Jejich vysegmentování lze zabránit nastavením vyšší hodnoty škálového parametru α , například $\alpha = 3$, v tom případě je však galaxie ještě více uříznutá.



Obrázek 4.3: Test edge1, $\alpha = 0.2$, $\lambda = 2$, 350 iterací. Poslední obrázek ozačuje výstup z Cannyho hranového detektor (předtím tímto byl obrázek vyhlazen gaussovským jádrem velikosti 5, $\sigma = 5$).

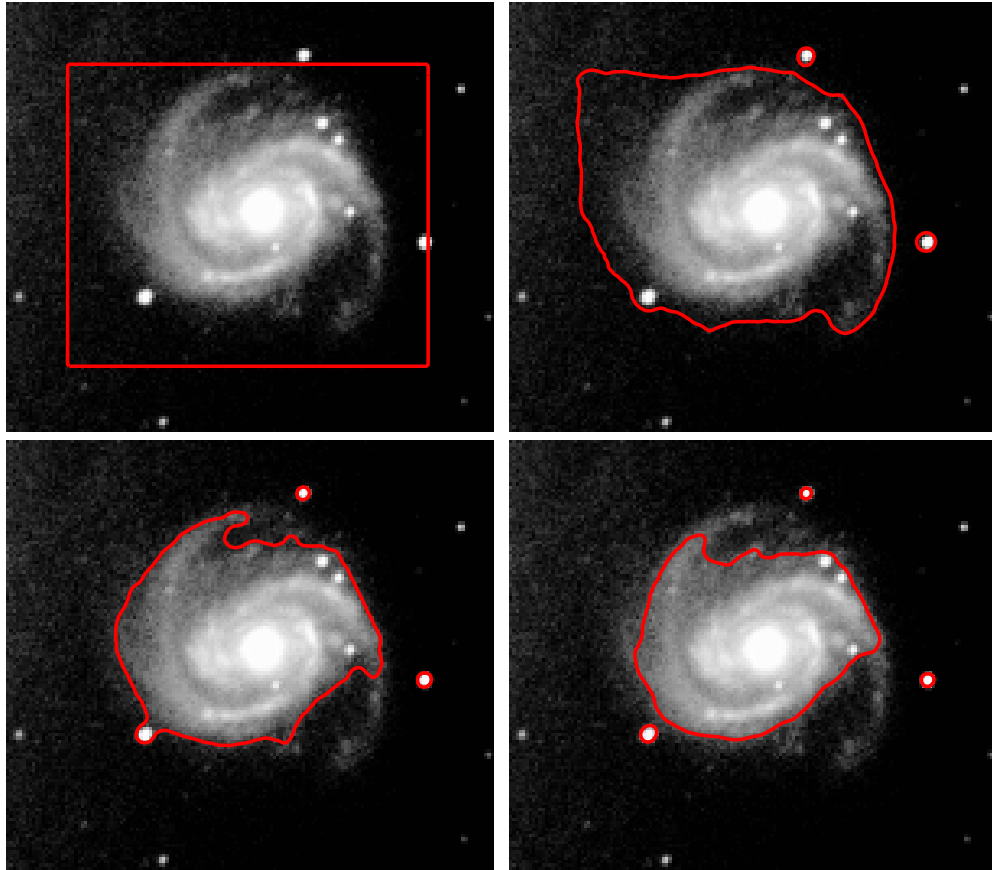
Test nonconvex, segmentace silně nekonvexního objektu

Obrázek 4.5. Obrázek v tomto testu je opět umělý. Cílem je ověřit schopnost segmentovat objekty, které jsou silně nekonvexní, neboli mají velký obvod vzhledem k zaujímané ploše. Tento typ objektů je nesmírně nepříjemný pro active contours modely založené na mean curvature motion a detekci hran. Vezmeme-li jako příklad geodesic active contours s funkcí (2.29), pak evoluční rovnice křivky má tvar

$$\frac{\partial c}{\partial t} = (gk - \nabla g \cdot \mathbf{n})\mathbf{n}.$$

Připomínáme, že funkce g detekuje hrany a k značíme křivost. Lze ukázat, že konvekční člen $\nabla g \cdot \mathbf{n}$ přitahuje křivku k hraně, má však velice malý dosah a není příliš účinný, dokud je křivka daleko od hrany. Pak je podstatný jen člen s k zkracující křivku. Dokud je vnitřek křivky u hrany lokálně konvexní, má k stále stejné znaménko a křivka se zkracuje, jakmile se však křivka narovná, křivost změní znaménko (resp. bude nulová) a vývoj křivky se zastaví — křivka se tak nikdy neprohne aby přestala být konvexní, nebude-li v těsné blízkosti hrany. Z toho důvodu je do rovnice potřeba přidat navíc „umělý“ člen νg ,

$$\frac{\partial c}{\partial t} = (gk - \nabla g \cdot \mathbf{n} + \nu g)\mathbf{n},$$



Obrázek 4.4: Test `edge2`, $\alpha = 0.1$, $\lambda = 2$, 1000 iterací.

který lze interpretovat jako minimalizaci plochy, kterou křivka uzavírá (z rovnice je jasné, že se jedná o pohyb v normálovém směru). Tento člen pak musí vykompenzovat opačné znaménko k , aby celkově znaménko $k + \nu$ bylo stále stejné. Problém však je, že je-li koeficient ν natolik vysoký, aby umožnil segmentaci silně nekonvexních objektů, bude pak tento člen velice silný a na neostře hraně, kde není $g = 0$, snadno způsobí, že křivka hranu přeskočí. Na takovém typu objektů jsou tedy hranové metody poměrně nestabilní a navíc velice pomalé, protože křivka při svém vývoji musí postupně okopírovat celou hranici objektu, nemůže se najednou „objevit uprostřed“, jako je tomu v případě `active contours without edges`.

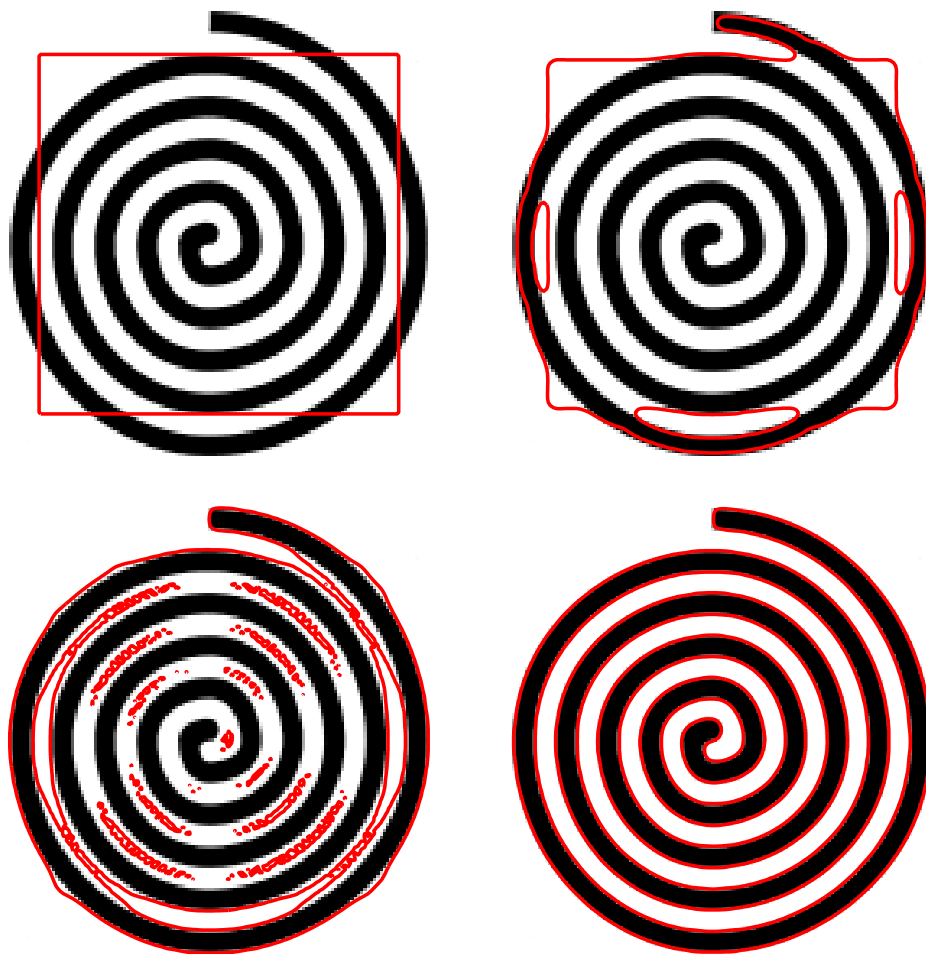
Náš model na tomto obrázku funguje velice dobře, konvergence je rychlá, neboť segmentace probíhá v celé ploše obrázku, což je velká výhoda tohoto způsobu segmentace.

Test `real1`, reálný obrázek s bimodálním charakterem

Obrázek 4.6. V tomto testu je jako obrázek opět použita reálná fotografie. Pozadí (tráva) a částečně i objekt (ruka) mají svou vlastní strukturu (texturu), přesto segmentace funguje velice dobře a rychle, protože objekt je výrazně světlejší než pozadí.

Test `noise`, reálný obrázek se šumem

Obrázek 4.7. Tento test je stejný jako předchozí, do obrázku byl však přidán aditivní šum. Segmentace funguje správně se stejnými parametry jako v předchozím případě, tedy bez šumu. To je pochopitelné, model pracuje primárně s globální (integrální) informací a ta není šumem příliš narušena, hranové segmentace pracující s lokální informací se s šumem vypořádávají daleko hůře. Šum způsobí snížení rozdílu mezi c_I a c_O , což

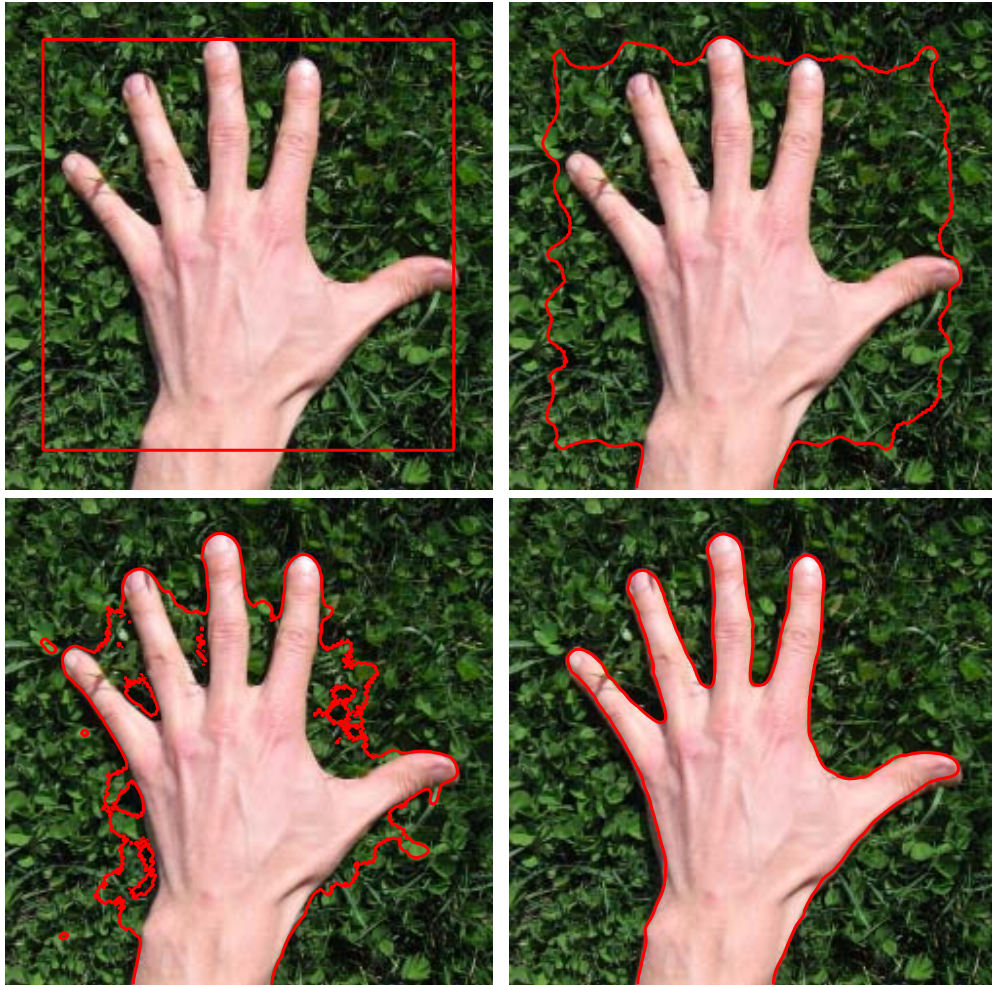


Obrázek 4.5: Test nonconvex, $\alpha = 0.1$, $\lambda = 1$, 200 iterací.

znamená, že segmentace (konvergence do minima) bude pomalejší. Dokud však zůstane zachován bimodální charakter snímku, je metoda proti šumu dobře odolná.

Test real12, reálný obrázek s rušivým prvkem

Obrázek 4.8. Na obrázku k tomuto testu je objekt zřetelně tmavší než pozadí, pozadí však obsahuje části (zábradlí, hrana chodníku), které jsou také tmavé a mají tak svou intenzitou blíže k intenzitě segmentovaného objektu. Segmentace s uvedenými parametry a počáteční křivkou je uspokojivá, i když ne zcela správná — okolo zadních nohou kočky je k objektu připojen i kus zábradlí a chodníku. Nepovedlo se mi žádným nastavením parametrů dosáhnout toho, aby byla kočka vysegmentována zcela správně. Situace však není tak dobrá, jak se může zdát. Pokud posuneme počáteční křivky ϕ_0 tak, aby byla dál od segmentovaného objektu (stačí například ve skriptu `test_real12.m` nastavit parametr `offset=20`, což přiblíží počáteční křivku k okrajům obrázku), potom se již křivka „chytne“ tmavých částí pozadí a segmentace není nikdy provedena správně. Tento test tak ukazuje, že u složitějších obrázků je nejen rychlost, ale i úspěšnost segmentace závislá na volbě počáteční křivky, apriorní informace o přibližné poloze objektu tak může být pro výsledek segmentace rozhodující.

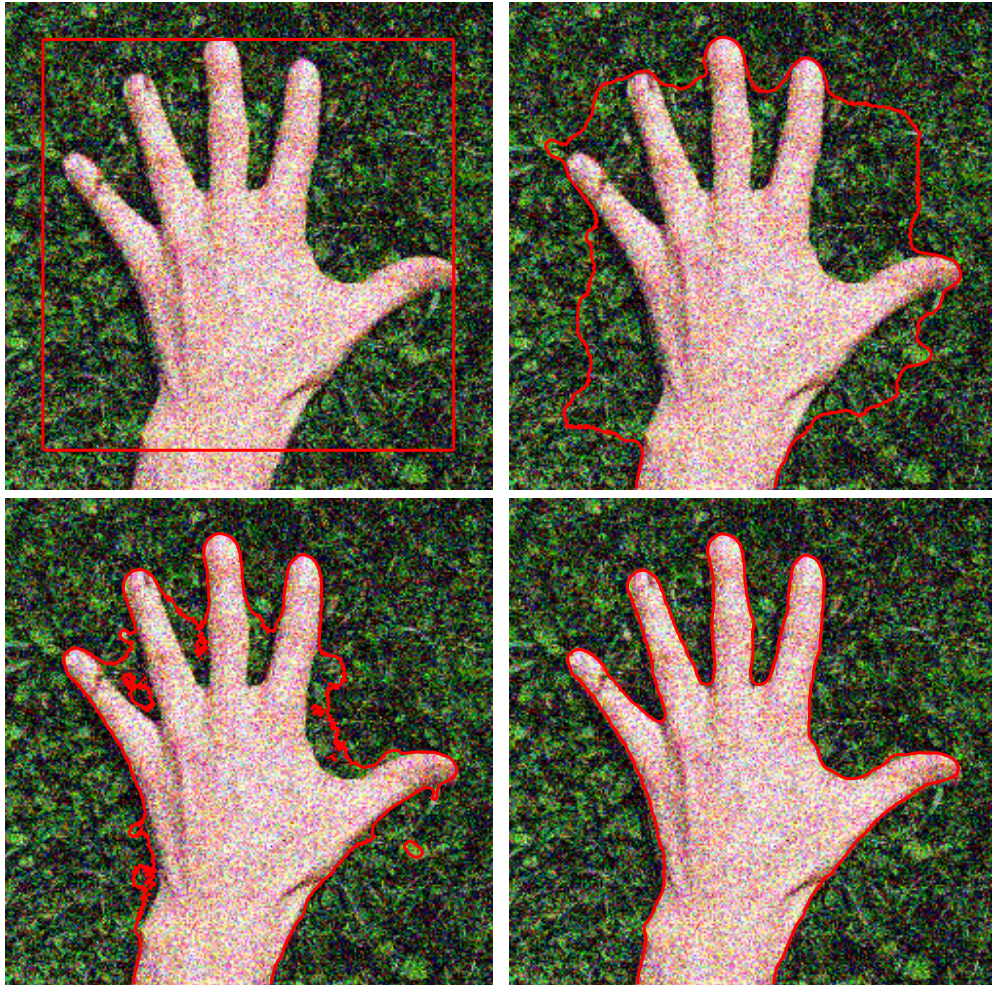


Obrázek 4.6: Test `real11`, $\alpha = 0.3$, $\lambda = 1$, 350 iterací.

Test `real3`, univerzální volba ϕ_0

Obrázek 4.9. V tomto testu se snažíme adresovat problém, na který jsme narazili v minulém testu, tedy volbu počáteční funkce ϕ_0 . Obrázek k tomuto testu je relativně dobře vysegmentovatelný. Pokud zvolíme ϕ_0 stejně „ledybyle“ jako ve většině předchozích testů, kde křivka vede 20px od okrajů obrázku, můžeme se přesvědčit, že správná segmentace nastane po zhruba 700 iteracích. V ukázce je však ϕ_0 zvolena jinak, a sice s rovnoměrně rozloženými „ostrůvky“ kladných a záporných hodnot, takže vzniklou počáteční křivku tvoří několik uzavřených křivek rovnoměrně rozložených v ploše obrázku. Vidíme, že v tomto případě nastane správná segmentace již po zhruba 150 iteracích, což je podstatné zrychlení, a přitom k volbě ϕ_0 nebyla potřeba žádná informace o poloze segmentovaného objektu. Proč je v tomto případě segmentace rychlejší? Na počátku jsou konstanty c_I, c_O přibližně stejné jako v případě jednoduché volby ϕ_0 , neboť obě křivky byly voleny náhodně, takže datový člen má zhruba stejnou váhu. Podstatné však je, že ve druhém případě má křivka mnohem větší délku, takže má možnost „reagovat“ na podstatně větší počet pixelů vstupního obrazu a měnit podle toho svůj tvar, důsledkem toho je, že se konstanty c_I, c_O rychle aktualizují (vzdalují), což vede k rychlé konvergenci. Stojí za to upozornit, že u hranových active contours metod možnost takovéto inicializace není, neboť křivka musí vždy obklopotovat segmentovaný objekt.

Pokud jde o vlastní výsledek segmentace, vidíme že tělo helikoptéry je vysegment-

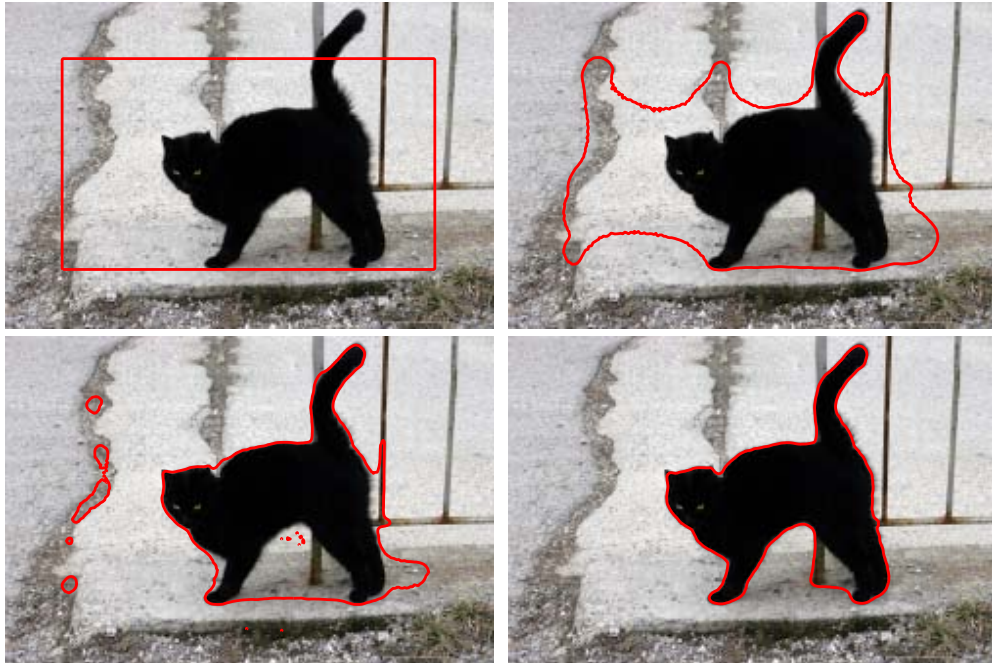


Obrázek 4.7: Test noise, $\alpha = 0.3$, $\lambda = 1$, 400 iterací.

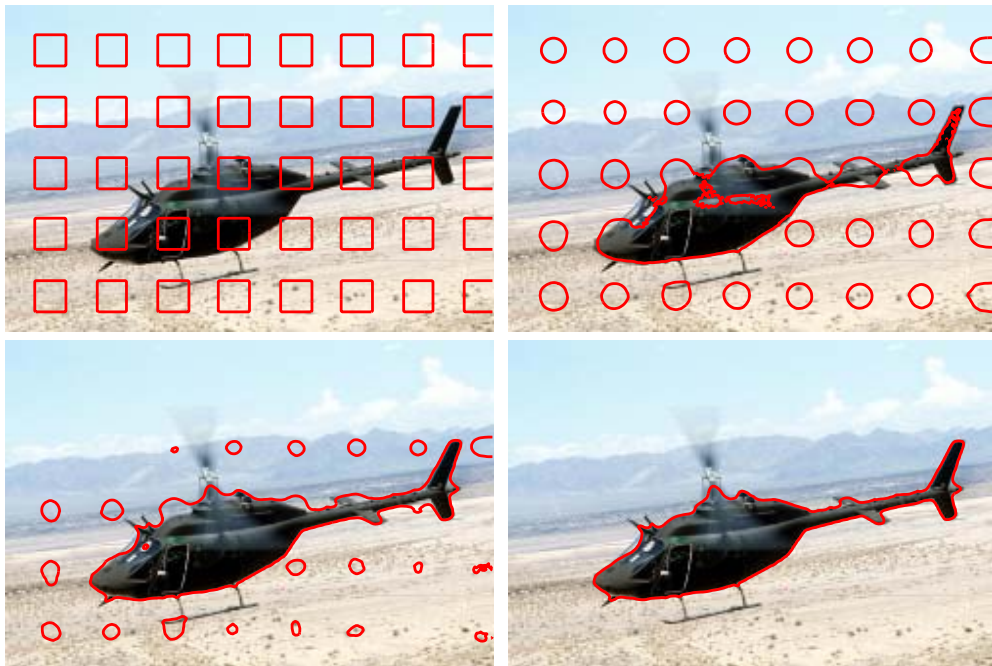
tované dobře, ale chybí lyžiny a rotor. Zda je taková segmentace postačující závisí na dalším zpracování, pokud segmentaci následuje rozpoznání objektu, mohou být chybějící periferie příčinou selhání. Zlepšení segmentace lyžin je možné dosáhnout při extrémních hodnotách parametrů, například $\alpha = 0$, to však již znamená víceméně oprahování intenzity obrazu, takže výsledek segmentace pak v podstatě ani neodpovídá teoretickému očekávání. Použití metody s takovou volbou parametrů nelze doporučit. Správná segmentace rotoru je zcela mimo schopnosti této metody.

Test texture, segmentace textury

Obrázek 4.10. Obrázek k tomuto testu je opět umělý, segmentovaný objekt je srdíčko, které se však od pozadí liší pouze jinou texturou. Přímá aplikace metody nemůže objekt správně vysegmentovat, neboť průměrná intenzita objektu je pochopitelně rovna průměrné intenzitě pozadí. Pro segmentaci jde však použít i jiný deskriptor než přímo intenzitu. V tomto případě je tak nejprve v každém bodě obrazu spočtena velikost derivace ve směru (1,1), teprve podle tohoto deskriptoru je prováděna segmentace. Výsledkem je úspěšná segmentace objektu.



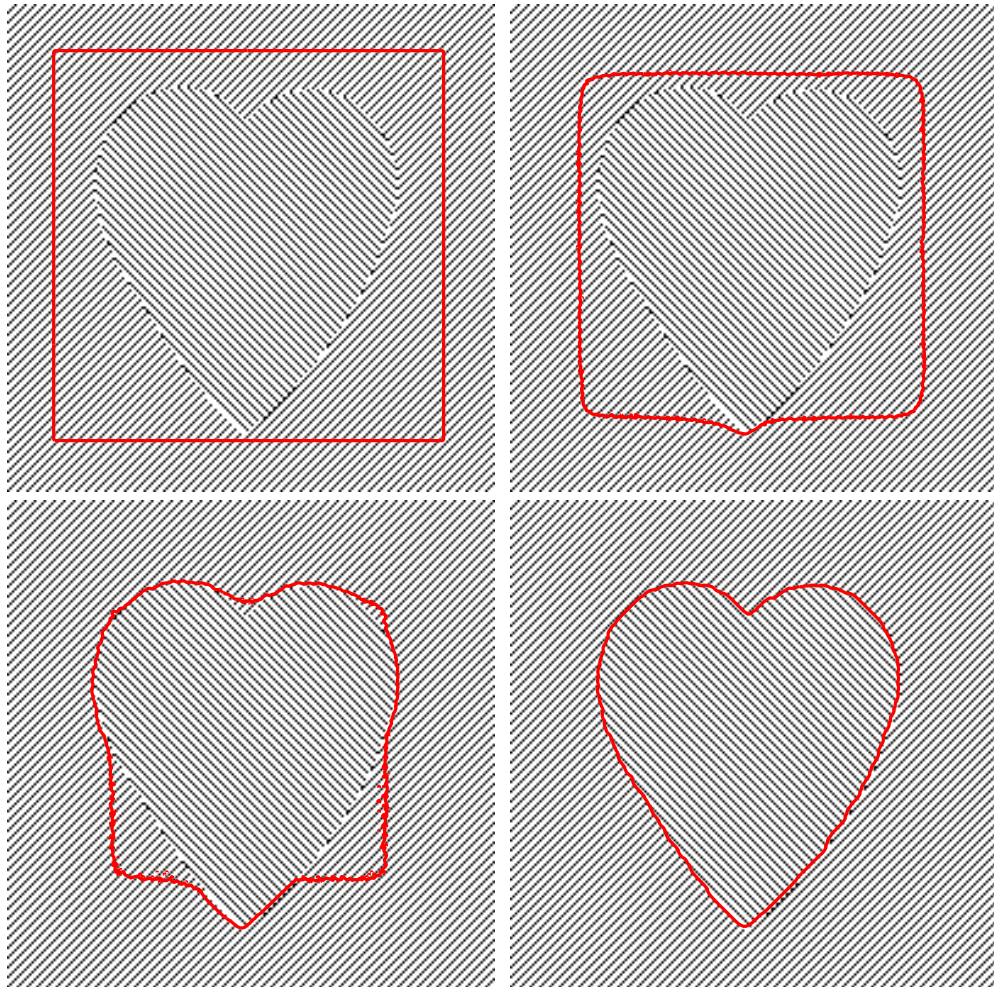
Obrázek 4.8: Test real2, $\alpha = 1$, $\lambda = 4$, 600 iterací.



Obrázek 4.9: Test real3, $\alpha = 0.5$, $\lambda = 4$, 200 iterací.

Shrnutí praktických poznatků

Na provedených testech je vidět, že metoda dobře funguje v případě, že je vstupní obraz přibližně bimodální, čímž myslíme, že průměrná intenzita objektu je nezanedbatelně vyšší či nižší, než průměrná intenzita pozadí. To není překvapivé, protože z této představy vychází idea této segmentační metody. Pokud není tento předpoklad splněn a objekt má průměrnou intenzitu stejnou jako pozadí, není přímá segmentace touto metodou ani teoreticky možná. Lze však v některých případech použít jiný deskriptor než jen intenzitu, jak bylo ukázáno v testu `texture`, většinou jde o funkci koeficientů



Obrázek 4.10: Test `texture`, $\alpha = 0.5$, $\lambda = 2$, 400 iterací.

rozkladu obrázku do nějaké báze.

Výsledek i rychlost segmentace je závislá na počáteční poloze křivky. Máme-li přibližnou informaci o poloze segmentovaného objektu, jde vhodným umístěním počáteční křivky dosáhnout výrazně lepšího výsledku. Pokud taková informace není, lze v některých případech segmentaci zlepšit, pokud se počáteční křivky rozmístí rovnoměrně v ploše obrázku, jako tomu je v testu `real3`, počet komponent křivky v takovém případě nesmí být příliš vysoký (zejména při vysokém koeficientu α), neboť se může stát, že po několika iteracích bude $\phi(x, y) > 0$ nebo $\phi(x, y) < 0$ všude v Ω a segmentace se tak zcela zastaví.

Metoda není příliš vhodná pro segmentaci objektů s vysokým poměrem délky hranice ku ploše objektu, tzn. například segmentaci mnoha malých objektů na jednom obrázku. Datová informace pocházející z plochy objektů je pak velice slabá a člen zkracující délku křivky se stane dominantní. Ten však nedělá žádnou skutečnou segmentaci, slouží jen jako regularizace a škálový prvek.

U snadno segmentovatelných obrázků (tj. silně bimodálních) je metoda stabilní vzhledem k nevelkým změnám vstupních parametrů, takže správného výsledku jde dosáhnout s nějakým univerzálním nastavením. U obrázků s komplikovanější strukturou je však nastavení parametrů (konkrétně tedy vždy parametru α vzhledem k nastavené hodnotě λ) důležité a při špatném nastavení není výsledek segmentace správný. Pokud tedy není předem znám charakter obrázku a z něho plynoucí nastavení parametrů, není metoda vhodná pro automatickou neřízenou segmentaci.

Metoda je velice dobře odolná vůči šumu do té míry, dokud přítomnost šumu nepoškodí bimodalitu obrazu. Metoda si dobře poradí se změnou topologie křivky, počáteční křivka nemusí obklopovat segmentovaný objekt a počet komponent počáteční křivky nemusí odpovídat počtu segmentovaných objektů.

Je-li segmentovaný objekt od pozadí dobře ohraničen gradientní hranou, ale přitom má vlastní vnitřní texturu podobnou textuře pozadí, tak metoda pravděpodobně selže, neboť hranovou informaci nebere v potaz, pro tento druh obrázků je třeba použít hranové segmentační metody. Naopak, metoda nevyžaduje, aby byl segmentovaný objekt od pozadí oddělen gradientní hranou.

Metoda je pomalá, velikost všech testovaných obrázků nepřesáhla 250px v jednom rozměru a segmentace trvá jednotky, vyjímečně desítky vteřin, včetně vykreslování grafiky (časy je třeba brát za orientační, neboť při experimentech nebyl cíl dosáhnout vhodné volbou parametrů nejrychlejší segmentace). Čas jedné iterace však roste lineárně s počtem pixelů vstupního obrázku, neboť v každé iteraci se provádí konečně mnoho sčítání matic a násobení matic element po elementu (nikoliv standardní násobení matic), což jsou všechno $O(N)$ operace, kde N je počet elementů v matici. Celkový čas úlohy tedy roste minimálně kvadraticky s délkovým rozměrem (tzn. šířkou nebo výškou) obrázku. Ve skutečnosti však roste také potřebný počet iterací. Odhadneme-li hrubě pro představu, že každá iterace provede (vzhledem k lokalizaci na okolí křivky) klasifikaci pouze $k < N$ vstupních pixelů, pak počet potřebných iterací roste lineárně s N , pak by celkový čas segmentace rostl jako $O(N^2)$, kde N je počet pixelů obrázku. Složitost jedné iterace jde podstatně zlepšit jinou implementací, existuje tzv. *narrow band* implementace, která nejen že soustředí vývoj křivky jen na její okolí, ale dokonce ani s jinými pixely vůbec nepočítá, iterace je pak podstatně rychlejší. Nevýhodou ale je, že v tom případě nemohou být správně vysegmentovány objekty, které nejsou jednoduše souvislé, viz například test `edge1`, a výsledek segmentace může ještě více záviset na poloze počáteční křivky.

Závěr

V práci byla pečlivě představena teorie Mumford–Shahova segmentačního modelu digitálního obrazu. Byla vysvětlena úvodní myšlenka a z ní vycházející formulace modelu, dále byla adresována a zodpovězena otázka existence a jednoznačnosti řešení, probrány vlastnosti segmentace a zdůvodněno ústřední postavení tohoto modelu mezi ostatními variačními segmentačními modely, čímž bylo také ukázáno, jak lze modifikacemi získat jiné segmentační modely. Dále byla vypracována konkrétní numerická implementace jedné z těchto variant, modelu *active contours without edges*. Tato implementace byla podrobně vysvětlena a posléze testována na praktických příkladech. Bylo poukázáno na výhody i nevýhody modelu či implementace a tyto vlastnosti byly také teoreticky zdůvodněny. Práce tak představuje dobrý zdroj informací pro kohokoliv, kdo se zajímá o variační segmentaci digitálního obrazu jak po teoretické, tak po praktické stránce.

Literatura

- [1] G.Alberti and C.Mantegazza. A Note on the Theory of SBV Functions. *Bollettino Unione Matematica Italiana Sezione B*, 11(2):375–382, 1997.
- [2] L.Ambrosio. A compactness theorem for a new class of functions of bounded variation. *Bollettino della Unione Matematica Italiana*, VII(4):857–881, 1989.
- [3] L.Ambrosio. Variational problems in SBV and image segmentation. *Acta Applicandae Mathematicae*, 17:1–40, 1989.
- [4] L.Ambrosio. Existence theory for a new class of variational problems. *Archive for Rational Mechanics and Analysis*, 111:291–322, 1990.
- [5] G. Aubert and P. Kornprobst. *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*. New York: Springer-Verlag, 2002.
- [6] A.Bonnet. On the regularity of the edge set of Mumford–Shah minimizers. *Progress in Nonlinear Differential Equations*, 25:93–103, 1996.
- [7] V.Caselles, R.Kimmel, and G.Sapiro. geodesic active contours. In *Proceedings of the 5th International Conference on Computer Vision*, pages 694–699, Boston, MA, June 1995. IEEE Computer Society Press.
- [8] T.Chan and L.Vese. Active contours without edges. Technical Report 98–53, UCLA CAM Report, 1999.
- [9] T.Chan and L.Vese. Image segmentation using level sets and the piecewise–constant Mumford–Shah model. Technical Report 00–14, UCLA CAM Report, 2000.
- [10] T.Chan, B.Sandberg, and L.Vese. Active contours without edges for vector–valued images. *Journal of Visual Communication and Image Representation*, 11:130–141, 2000.
- [11] T.Chan, B.Sandberg, and L.Vese. Active contours without edges for textured images. *UCLA Department of Mathematics CAM Report*, 02–28, 2002.
- [12] T.Chan and J.Chen. *Image processing and analysis: variational, PDE, wavelet and stochastic methods*. Siam, 2005.
- [13] L.C.Evans and R.F.Gariepy. *Measure Theory and Fine Properties of Functions*. CRC Press, 1992.
- [14] L.C.Evans. *Partial Differential Equations*, volume 19 of *Graduate Studies in Mathematics*. Proceedings of the American Mathematical Society, 1998.

- [15] M.Feistauer. *Mathematical Methods in Fluid Dynamics*. Pitman Monographs and Surveys in Pure and Applied Mathematics Series 67, Longman Scientific & Technical, Harlow, 1993.
- [16] E.De Giorgi and L.Ambrosio. Un nuovo tipo di funzionale del calcolo delle variazioni. *Att. Accad. Naz. Lincei, Rend Cl Sci. Fis. Mat. Nat.*, 82:199–210, 1988.
- [17] E.De Giorgi, M.Carriero, and A.Leaci. Existence theorem for a minimum problem with a free discontinuity set. *Archive for Rational Mechanics and Analysis*, 108:195–218, 1989.
- [18] P. Grisvard. *Elliptic Problems in Nonsmooth Domains*, volume XIV of *Monographs and Studies in Mathematics*. Pitman Advanced Publishing Program. Boston–London–Melbourne: Pitman Publishing Inc., 1985.
- [19] R.C.Gonzales and R.E.Woods. *Digital Image Processing*. 2nd ed., Prentice Hall, Upper Saddle River, NJ, 2002.
- [20] A.D.Ioffe. On lower semicontinuity of integral functionals. *SIAM J. Control Optim.* 15, 521–538, 1977.
- [21] M.Kass, A.Witkin and D.Terzopoulos. Snakes: Active contour models. In *First International Conference on Computer Vision*, pages 259–268, London, June 1987.
- [22] C.Li, C.Xu, C.Gui, and M.D.Fox. Distance Regularized Level Set Evolution and its Application to Image Segmentation, *IEEE Trans. Image Processing*, vol. 19(12), 2010.
- [23] J.M.Morel and S.Solimini. Segmentation of images by variational methods: A constructive approach. *Rv. Math. Univ. Complut. Madrid*, 1:169–182, 1988.
- [24] J.M.Morel. The Mumford–Shah conjecture in image processing, *Séminaire N. Bourbaki*, exp.no. 813:221–242, 1996.
- [25] D.Mumford and J.Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42:577–684, 1989.
- [26] S.Osher and J.Sethian. Fronts propagating with curvature dependent speed:algorithms based on Hamilton–Jacobi formulation. *Journal of Computational Physics*, 79:12–49, 1988.
- [27] S.Osher and R.Fedkiw. *Level Set Methods and Dynamic Implicit Surfaces*. Applied Mathematical Sciences Volume 153, Springer–Verlag, 2003.
- [28] G.Sapiro. *Geometric Partial Differential Equations and Image Analysis*. Cambridge University Press, 2001.
- [29] J.A.Sethian. *Level Set Methods and Fast Marching Methods: Evolving Interfaces in Computational Geometry, Fluid Mechanics, Computer Vision, and Material Sciences*. Cambridge Monograph on Applied and Computational Mathematics. Cambridge University Press, 1999.

Příloha A

Zdrojový kód funkce acwe

```
function phi = acwe(phi, u, iterations, timestep,...
    lambda1, lambda2, alpha, mu, epsilon)
% provede dany pocet iteraci metody active contours without edges

for i=1:iterations % provest dany pocet iteraci
    % upravit krajni hodnoty, aby byly splneny okrajove podminky
    phi = Neumann(phi);
    % vypect gradientu phi (centralni diference)
    [phi_x,phi_y] = gradient(phi);
    % velikost gradientu
    s=sqrt(phi_x.^2 + phi_y.^2);
    % vypocet krivosti, male cislo zabranuje deleni nulou
    curvature = div(phi_x./(s + 1e-6), phi_y./(s + 1e-6));

    % vypocet jednotlivych clenu evolucni rovnice
    % regularizacni clen
    distRegTerm = distReg(phi, phi_x, phi_y, s);
    % delkovy clen
    lengthTerm = diracCompact(phi, epsilon).*curvature;
    % vypocet konstant c_I, c_E a datoveho clenu
    % prvky, ktere jsou uvnitr objektu
    in = (phi>0);
    % suma prvku uvnitr objektu
    dsum = sum(in(:));
    % rozhodnout, jestli je objekt prazdny nebo ne
    if dsum > 0
        % vypocet konstanty c_I a odpovidajiciho datoveho clenu
        data_in = (u-sum(u(in)))/dsum).^2;
    else
        % pokud je objekt prazdny tak je tento datovy clen nulovy
        data_in = 0;
    end
    % suma prvku vne objektu
    dsum = sum(~in(:));
    % rozhodnout, zda je pozadi prazdne ci ne
    if dsum > 0
        % vypocet konstanty c_E a odpovidajiciho datoveho clenu
        data_out = (u-sum(u(~in)))/dsum).^2;
```

```

else
    % pokud je pozadi prazdne tak je tento datovy clen nulovy
    data_out = 0;
end
% vypocet datoveho clenu
dataTerm = (-lambda1*data_in+lambda2*data_out).*diracSpread(phi, epsilon);

% explicitni iterace rovnice
phi = phi + timestep*(dataTerm + alpha*lengthTerm + mu*distRegTerm);
end
end

% vypocet divergence (centralni diference)
function K = div(nx,ny)
[nxx,~]=gradient(nx);
[~,nyy]=gradient(ny);
K=nxx+nyy;
end

% funkce delta_1 s kompaktnim nosicem, pouzita pro delkovy clen
function y = diracCompact(x,epsilon)
s = max(min(x,epsilon),-epsilon);
y = (1+cos(pi*s/epsilon))/(2*epsilon);
end

% funkce delta_2 s nosicem na R, pouzita pro datovy clen
function y = diracSpread(x, epsilon)
y = 1/pi*epsilon./(x.^2+epsilon^2);
end

% vypocet hodnoty clenu pro distance regularization
function f = distReg(phi, phi_x, phi_y, s)
a=(s>=0) & (s<=1);
b=(s>1);
ps=a.*sin(2*pi*s)/(2*pi)+b.*(s-1);
dps=((ps~=0).*ps+(ps==0))./((s~=0).*s+(s==0));
f = div(dps.*phi_x - phi_x, dps.*phi_y - phi_y) + 4*del2(phi);
end

% okopirovani krajnich hodnot pro splneni Neumannovy okrajove podminky
function g = Neumann(f)
g = f;
g([1 end], [1 end]) = g([3 end-2], [3 end-2]);
g([1 end],2:end-1) = g([3 end-2], 2:end-1);
g(2:end-1,[1 end]) = g(2:end-1,[3 end-2]);
end

```