Liet. matem. rink., T. 41, spec. nr., 596–602

Skaitmeninių vaizdų panašumo problema ir jos sprendimas

Nerijus MORKEVIČIUS, Jonas VALANTINAS (KTU)
el. paštas: valanti@if.ktu.lt

Įvadas

Realaus pasaulio vaizdų apdorojimas dažniausiai yra siejamas su jų matematinių modelių – skaitmeninių vaizdų (kompiuterinių analogų) – kūrimu bei informacinio jų turinio kodavimu, užtikrinančiu efektyvų vaizdų saugojimą ir perdavimą. Daugelio žinomų efektyvaus vaizdų kodavimo algoritmų kokybę stipriai įtakoja įvairios tuose algoritmuose taikomos procedūros, būtent: adaptyvaus vaizdo kodavimo principo organizavimas, kvantavimo matricų, kvantavimo lygių parinkimas, po kvantavimo išlikusių nenulinių vaizdo (ar jo diskrečiojo spektro) elementų nuskaitymas bei kodavimas, ir panašiai, [1–4]. Vis didesnį populiarumą įgyjančiuose fraktalinio vaizdų kodavimo algoritmuose viena svarbiausių procedūrų – dviejų vaizdų, arba atskirų to paties vaizdo fragmentų, panašumo nustatymas, [5, 6]. Tai didelių laikinių sąnaudų reikalaujanti procedūra. Nuo jos organizavimo labai priklauso, ar konkretus fraktalinis vaizdo kodavimo algoritmas gali „dirbti" realiame laike, ar algoritmas iš vis turi praktinį pritaikomumą. Beje, ši procedūra taikoma ir reguliariosiomis išraiškomis pagrįstuose skaitmeninių vaizdų kodavimo algoritmuose, [7, 8].

Žemiau trumpai pristatomas kontekstas (pagrindinės sąvokos ir apibrėžimai), kuriame formuluojama ir sprendžiama skaitmeninių vaizdų panašumo nustatymo problema.

Skaitmeninių vaizdų erdvė

Imkime realaus pasaulio vaizdų matematinių modelių – d-mačių skaitmeninių vaizdų (duomenų masyvų) – aibę

$$S^d(n) = \{[X(m)] \mid m = (m_1, \ldots, m_d) \in I^d\};$$

čia: $I = \{0, 1, \ldots, N-1\}$, $N = 2^n$, $n \in \mathbf{N}$, $d \in \{1, 2, 3\}$; $X(m) \in \{0, 1, \ldots, 2^p - 1\}$, $p \geqslant 1$; pastebėsime, jog parametras n nusako realaus pasaulio vaizdo detalizacijos lygį, tuo tarpu kai parametras p – bitų, skirtų vaizdo $[X(m)]$ elementų (pikselių) reikšmių kodavimui, skaičių; atskiru atveju, $p = 1$ atitinka dvejetainį (juodai-baltą, „siluetinį") vaizdą, $p > 1$ – nespalvotą vaizdą, kurio kodavimui panaudojami 2^p kvantavimo lygiai (pilkumo atspalviai). Reikia pabrėžti, jog tinkamai parinkus n ir p reikšmes (pavyzdžiui,

$n \geqslant 9, p \geqslant 8$), realaus pasaulio vaizdas ir jo modelis (skaitmeninis vaizdas) yra vizualiai neatskiriami.

Atstumas (metrika) δ tarp bet kurių dviejų aibės $S^d(n)$ elementų-vaizdų $[X_1(m)]$ ir $[X_2(m)]$ – apibrėžiamas taip:

$$\delta = \delta(X_1, X_2) = \sqrt{\frac{1}{N^d} \sum_{m \in I^d} \left(X_2(m) - X_1(m)\right)^2}.$$

Ši metrika naudojama kaskart, kai reikia palyginti pradinio vaizdo ir po kodavimo atkurto vaizdo (pradinio vaizdo įverčio) kokybę.

Taigi, $(S^d(n), \delta)$ yra baigtinė metrinė d-mačių n-ojo detalizacijos lygio skaitmeninių vaizdų erdvė. Būtent tokio tipo metrinėse erdvėse formuluojama ir sprendžiama vaizdų panašumo nustatymo problema.

Skaitmeninių vaizdų panašumo nustatymo problema

Imkime skaitmeninių vaizdų erdvę $(S^d(n), \delta)$. Taikomojo pobūdžio uždaviniuose (jau minėtieji kodavimo algoritmai, [5–8]) susiduriama su tokiu uždaviniu: kiekvienam vaizdui $[U(m)] \in S_1^d(n) \subset S^d(n)$ rasti panašų į jį vaizdą $[V(m)] \in S_2^d(n) \subset S^d(n)$ (bendru atveju, $S_1^d(n) \cap S_2^d(n) = \varnothing$). Vaizdų $[U(m)]$ ir $[V(m)]$ panašumas vertinamas metrikos δ prasme, t.y. šie vaizdai laikomi panašiais tada ir tik tada, kai $\delta = \delta(U, V) \leqslant \varepsilon_0$; čia ε_0 yra iš anksto pasirinktas teigiamas skaičius.

Kartais, siekiant padidinti panašių vaizdų aptikimo tikimybę, vaizdui $[V(m)]$ iš aibės $S_2^d(n)$ papildomai taikomos įvairios transformacijos. Jų tarpe:

1) *Posūkis*; šiai transformacijai būdinga tai, kad ji atliekama nekeičiant vaizdo koordinatinių ašių tarpusavio padėties d-matėje erdvėje. Rezultate gaunamas naujas (modifikuotas) vaizdas $[\tilde{V}(\tilde{m})]$, $\tilde{m} = (\tilde{m}_{i_1}, \ldots, \tilde{m}_{i_d}) \in I^d$; čia: $\tilde{m}_{i_s} = m_{i_t}$ arba $\tilde{m}_{i_s} = N - m_{i_t} - 1$, $i_s, i_t \in \{1, \ldots, d\}$, $s = 1, \ldots, d$. Detaliau neanalizuodami šios transformacijos, pastebėsime, jog bendras (posūkio pagalba) modifikuotų vaizdų skaičius yra lygus 23, kai $d = 3$, lygus 3, kai $d = 2$, ir lygus 0, kai $d = 1$;

2) *Veidrodinis atspindys*; šios transformacijos metu keičiama vaizdo koordinatinių ašių tarpusavio padėtis d-matėje erdvėje. Gaunamas naujas vaizdas $[\tilde{V}(\tilde{m})]$, $\tilde{m} = (m_{i_1}, \ldots, m_{i_d}) \in I^d$, $i_s \in \{1, \ldots, d\}$, $s = 1, \ldots, d$. Bendras modifikuotų vaizdų skaičius yra lygus 3, kai $d = 3$, ir lygus 1, kai $d \in \{1, 2\}$;

3) *Inversija*; ši transformacija apibrėžiama lygybe

$$\tilde{V}(m) = 2^p - V(m) - 1, \quad m = (m_1, \ldots, m_d) \in I^d.$$

4) *Kontrastiškumo keitimas*; šiuo atveju, $\tilde{V}(m) = \lambda \cdot V(m)$, $m = (m_1, \ldots, m_d) \in I^d$. Pastebėsime, jog (šiame kontekste) koeficientas λ parenkamas taip, kad atstumas

$\delta = \delta(U, \tilde{V})$ įgytų minimalią reikšmę. Centruotų vaizdų atveju (vaizdų $[U(m)]$ ir $[V(m)]$ pastoviosios dedamosios lygios nuliui) optimali λ reikšmė parenkama taip:

$$\lambda = \lambda_{opt.} = \sum_{m \in I^d} U(m) \cdot V(m) \Big/ \sum_{m \in I^d} V^2(m).$$

Pridursime, jog vaizdui $[V(m)] \in S_2^d(n) \subset S^d(n)$ sėkmingai galima taikyti ir bet kurią baigtinę aukščiau išvardytų transformacijų kombinaciją.

Turint omenyje tai, kad praktiškai (pavyzdžiui, fraktalinėse vaizdų kodavimo procedūrose) $10^2 < |S_1^d(n)| < 10^4$, $10^4 < |S_2^d(n)| < 10^6$, galime drąsiai teigti, jog metrikos (kriterijaus) δ panaudojimas vaizdų panašumo problemos sprendime, įskaitant ir galimas transformacijas, reikalauja didžiulių laikinių sąnaudų. Be to, dažnai tenka palyginti vaizdus, priklausančius skirtingo detalizacijos lygio erdvėms. Iškyla lygių suvienodinimo problema.

Kaip sumažinti šias laikines sąnaudas? Vienas iš galimų sprendimo būdų – kriterijaus, leidžiančio kiekvienam vaizdui $[U(m)] \in S_1^d(n)$ operatyviai išskirti aibės $S_2^d(n)$ poaibį („baseiną") $S_{20}^d(n)$ ($|S_{20}^d(n)| << |S_2^d(n)|$), kurį sudarantys vaizdai būtų „potencialiai" panašūs į $[U(m)]$, parinkimas.

Mūsų (šia prasme) siūlomas kriterijus – vaizdų panašumo nustatymo problemos sprendimas – remiasi vaizdo glodumo samprata.

Diskretusis vaizdo spektras, vaizdo glodumo samprata

Imkime skaitmeninį vaizdą $[X(m)] \in S^d(n)$. Įvesime vaizdo glodumo sąvoką. Tam panaudosime d-matį diskretųjį vaizdo $[X(m)]$ spektrą $[Y(k)]$, nusakomą, bendru atveju, lygybe

$$Y(k) = \frac{1}{N^d} \sum_{m \in I^d} X(m) \cdot \Phi(k, m), \tag{1}$$

$k = (k_1, \ldots, k_d) \in I^d$; čia: $\{\Phi(k, m)\}$ yra baigtinė d-mačių diskretizuotų ortogonaliųjų funkcijų (Volšo, kosinusinių, ir panašiai, [9]) sistema; paprastai, šios funkcijos užrašomos, panaudojant vienmates ortogonaliąsias funkcijas, t.y. $\Phi(k, m) = \prod_{i=1}^{d} \Phi(k_i, m_i)$, $k, m \in I^d$; beje, jeigu $\{\Phi(k, m)\}$ yra ortonormuota funkcijų sistema, tai (1) išraiškoje nerašomas daugiklis $1/N^d$.

Pradinis vaizdas $[X(m)]$ vienareikšmiškai atkuriamas, taikant atvirkštinę d-matę diskrečiąją transformaciją

$$X(m) = \sum_{k \in I^d} Y(k) \cdot \overline{\Phi(k, m)}, \tag{2}$$

$m \in I^d$; čia $\overline{\Phi(k, m)}$ yra kompleksinė jungtinė išraiška $\Phi(k, m)$ atžvilgiu.

Iš Parsevalio sąryšių ortogonaliųjų funkcijų (išdėstytų pagal dažnį) eilutėms daugiamačiu atveju išplaukia, jog diskrečiojo spektro $[Y(k)]$ elementų $Y(k)$, $k \in I^d$, absoliutinėms skaitinėms reikšmėms būdinga mažėjimo tendencija, kai jų numeriai (indeksai $k_1, \ldots, k_d$) didėja, t.y. egzistuoja hiperbolinis „paviršius“

$$z = z(x_1, \ldots, x_d) = C/(x_1 \cdot \ldots \cdot x_d)^\alpha \quad (C \geqslant 0,\ \alpha \geqslant 0),$$

aproksimuojantis duomenų masyvą $\{|Y(k)| \,|\, k_1^2 + \ldots + k_d^2 \neq 0\}$ vidutinės kvadratinės paklaidos (VKP)

$$\mathrm{VKP} = \sqrt{\frac{1}{N^d - 1} \sum_{\substack{k \in I^d \\ (k_1^2 + \ldots + k_d^2 \neq 0)}} \left(|Y(k)| - \frac{C}{(\bar{k}_1 \cdot \ldots \cdot \bar{k}_d)^\alpha} \right)^2}$$

prasme; čia: $\bar{k}_i = \max\{k_i, 1\}$, $i = 1, \ldots, d$.

Hiperbolinį „paviršių“ charakterizuojantis dydis α interpretuojamas kaip vaizdo $[X(m)] \in S^d(n)$ glodumo parametras (klasė, lygis). Beje, sudaryta iteracinė vaizdų $[X(m)] \in S^2(n)$ glodumo klasės nustatymo procedūra, [10]. Jos apibendrinimas kitokio matavimo erdvės vaizdams nėra sudėtingas. Atskiru atveju, kai vaizdo $[X(m)] \in S^d(n)$ diskrečiojo spektro $[Y(k)]$ elementai $|Y(k)| > 0$, su visais $k = (k_1, \ldots, k_d) \in I^d$, vaizdo glodumo parametro reikšmė $\alpha = \alpha_x$ apskaičiuojama pagal formulę, gautą taikant mažiausiųjų kvadratų metodą, t.y.

$$\alpha_x = \frac{1}{A_N} \cdot \sum_{k \in K} \log \frac{\prod\limits_{k \in K} (\bar{k}_1 \cdot \ldots \cdot \bar{k}_d)}{(\bar{k}_1 \cdot \ldots \cdot \bar{k}_d)^{N^d - 1}} \cdot \log |Y(k)|; \tag{3}$$

čia: $A_N = (N^d - 1) \sum_{k \in K} \log^2(\bar{k}_1 \cdot \ldots \cdot \bar{k}_d) - \left(\sum_{k \in K} \log(\bar{k}_1 \cdot \ldots \cdot \bar{k}_d) \right)^2$ ir yra neneigiamas pastovus dydis, esant fiksuotoms N ir d reikšmėms; $K = \{k | k \in I^d,\ k_1^2 + \ldots + k_d^2 \neq 0\}$.

Tačiau, jeigu egzistuoja bent vienas spektro elementas $Y(k) = 0$, tai sumavimas (3) išraiškoje atliekamas tik pagal indeksus $k \in K \backslash \{k | Y(k) = 0\}$. Apskaičiuotoji (pagal (3) formulę) reikšmė α_x laikoma pirmąja vaizdo $[X(m)] \in S^d(n)$ glodumo parametro aproksimacija, ir jos patikslinimui taikoma iteracinė optimizavimo pagal koordinates (α ir C) procedūra, analogiška aprašytąjai straipsnyje [10].

Svarbu pabrėžti tai, jog vaizdo $[X(m)]$ glodumo parametras α_x yra invariantiškas vaizdą $[X(m)]$ veikiančių transformacijų (posūkis, veidrodinis atspindys, inversija, ir panašiai) atžvilgiu. Šios glodumo parametro savybės praktikoje svarbiam atvejui ($d = 2$) įrodymą galima rasti straipsnyje [10].

Būtina vaizdų panašumo sąlyga – maži glodumo pokyčiai

Tarkime, kad $[U(m)] \in S_1^d(n) \subset S^d(n)$, $[V(m)] \in S_2^d(n) \subset S^d(n)$, o α_U ir α_V žymi atitinkamai šių vaizdų glodumo parametro reikšmes. Įrodysime, jog būtina vaizdų $[U(m)]$

ir $[V(m)]$ panašumo sąlyga yra mažas šių vaizdų glodumo parametro reikšmių α_U ir α_V skirtumas, t.y.

$$\left(\delta(U,V) \leqslant \varepsilon_0\right) \Rightarrow \left(|\alpha_V - \alpha_U| \leqslant \mu_0\right); \tag{4}$$

čia: ε_0 yra vaizdų $[U(m)]$ ir $[V(m)]$ panašumą sąlygojantis teigiamas skaičius; reikšmė μ_0 priklauso nuo ε_0. Kita vertus, jeigu vaizdų glodumo parametro reikšmių skirtumas yra didelis ($|\alpha_V - \alpha_U| > \mu_0$), tai vaizdai $[U(m)]$ ir $[V(m)]$ negali būti panašūs, nes $\delta(U,V) > \varepsilon_0$.

Tarkime, kad vaizdas $[V(m)] \in S^d(n)$ yra gautas, vaizdui $[U(m)] \in S^d(n)$ suteikus pokytį $[\Delta U(m)]$, t.y. $V(m) = U(m) + \Delta U(m)$, su visais $m \in I^d$. Be to, priimkime, kad

$$\delta = \delta(U,V) = \sqrt{\frac{1}{N^d}\sum_{m\in I^d}\left(V(m)-U(m)\right)^2} = \sqrt{\frac{1}{N^d}\sum_{m\in I^d}\left(\Delta U(m)\right)^2} \leqslant \varepsilon_0.$$

Įvertinsime skirtumą $|\alpha_V - \alpha_U|$. Pirmiausia, pažymėkime diskrečiuosius vaizdų $[U(m)]$ ir $[V(m)]$ spektrus atitinkamai $[Y_U(k)]$ ir $[Y_V(k)]$. Tada, su bet kuriuo k ($k \in I^d$), teisinga nelygybė

$$\begin{aligned}|Y_V(k) - Y_U(k)| &= |\Delta Y_U(k)| = \frac{1}{N^d}\left|\sum_{m\in I^d}\Delta U(m)\cdot\Phi(k,m)\right| \\ &\leqslant \max_{k,m\in I^d}\left\{|\Phi(k,m)|\right\}\cdot\frac{1}{N^d}\sum_{m\in I^d}|\Delta U(m)| \leqslant B_N\cdot\delta(U,V) \leqslant B_N\cdot\varepsilon_0;\end{aligned}$$

čia $B_N = \max_{k,m\in I^d}\left\{|\Phi(k,m)|\right\}$ ir yra pastovus dydis, esant fiksuotai N reikšmei ir konkrečiai naudojamai diskrečiąjai transformacijai. Toliau, pasinaudosime (3) išraiška, pastebėję, jog

$$\log|Y_V(k)| = \log|Y_U(k) + \Delta Y_U(k)| = \log|Y_U(k)| + \log\left(1 + \frac{\Delta Y_U(k)}{Y_U(k)}\right),$$

kai $|\Delta Y_U(k)/Y_U(k)| < 1$, su visais $k \in I^d$. Taigi, pažymėję

$$D_k = \log\frac{\prod\limits_{k\in K}(\bar{k}_1\cdot\ldots\cdot\bar{k}_d)}{(\bar{k}_1\cdot\ldots\cdot\bar{k}_d)^{N^d-1}},$$

turime

$$\begin{aligned}|\alpha_V - \alpha_U| &= |\Delta\alpha_U| = \frac{1}{A_N}\cdot\left|\sum_{k\in K}D_k\cdot\left(\log|Y_V(k)| - \log|Y_U(k)|\right)\right| \\ &= \frac{1}{A_N}\cdot\left|\sum_{k\in K}D_k\cdot\log\left|1+\frac{\Delta Y_U(k)}{Y_U(k)}\right|\right| \leqslant \frac{1}{A_N}\cdot\sum_{k\in K}|D_k|\cdot\frac{\left|\frac{\Delta Y_U(k)}{Y_U(k)}\right|}{1-\left|\frac{\Delta Y_U(k)}{Y_U(k)}\right|}\end{aligned}$$

$$\leqslant \frac{1}{A_N \cdot M} \cdot \sum_{k \in K} |D_k| \cdot \left| \frac{\Delta Y_U(k)}{Y_U(k)} \right| \leqslant \frac{B_N}{A_N \cdot M} \cdot \sum_{k \in K} |D_k| \cdot \frac{\varepsilon_0}{|Y_U(k)|} = \mu_0;$$

čia $M = 1 - \max\limits_{k \in K} \left\{ \left| \frac{\Delta Y_U(k)}{Y_U(k)} \right| \right\}.$

Iš pastarosios nelygybės, įvertinant anksčiau priimtas prielaidas, išplaukia, jog maži pokyčiai vaizde sąlygoja mažus vaizdo glodumo pokyčius, t.y.

$$\left(\delta(U, V) = \delta(U, U + \Delta U) \leqslant \varepsilon_0\right) \Rightarrow \left(|\alpha_V - \alpha_U| = |\Delta\alpha_U| \leqslant \mu_0\right). \tag{5}$$

Šis sąryšis išlieka teisingas ir tuo atveju, kai dalis spektro $[Y(k)]$ elementų yra lygūs nuliui, tuo pačiu, vaizdo $[V(m)]$ glodumo parametro reikšmei α_V įvertinti papildomai taikoma iteracinė (su baigtiniu žingsnių skaičiumi) optimizavimo procedūra.

Priklausomybės $\mu_0 = \mu_0(\varepsilon_0, N)$ nustatymo svarbiausiems praktikoje sutinkamiems atvejams būdas – eksperimentinis tyrimas. Teoriniais paskaičiavimais, laikinės sąnaudos turėtų sumažėti apie $10^2 - 10^3$ kartų, palyginus su visų galimų vaizdų porų perrinkimo ir analizės (vaizdų panašumo problemos sprendimo procese) laikinėmis sąnaudomis. Preliminarūs eksperimentai, atlikti bandant modeliuoti fraktalines vaizdų kodavimo procedūras, tai patvirtina.

Išvados

Šiame straipsnyje aprašytą kriterijų (būtiną skaitmeninių vaizdų panašumo sąlygą) tikslinga taikyti tada, kai konkrečiam vaizdui $[U(m)] \in S_1^d(n)$ reikia rasti„potencialiai“ panašių į jį vaizdų $[V(m)]$ poaibį („baseiną“) erdvėje $S_2^d(n)$. Gali pasirodyti, jog praktinė siūlomo kriterijaus realizacija perdėm sudėtinga jau vien dėl tos priežasties, kad vaizdo glodumo įverčiui gauti, bendru atveju, taikoma iteracinė procedūra. Tai tiesa, tačiau, vaizdo glodumo įverčiams apskaičiuoti visada galima panaudoti supaprastintas procedūras – imti nedidelę (ketvirtąją, šešioliktąją, ir panašiai) diskrečiojo vaizdo spektro dalį, riboti iteracijų skaičių, ir panašiai, [10]. Siūlomo kriterijaus teigiamo panaudojimo efektas slypi tame, kad palyginamiems vaizdams glodumas nustatinėjamas vienąkart.

Literatūra

[1] P. Zinterhof, P. Zinterhof jun., Hyperbolic filtering of Walsh series, *RIST++*, University of Salzburg (1993).

[2] J. Valantinas, A new approach to hyperbolic filtering of gray-level images, *Information Technology and Control*, Kaunas, „Technologija“, 1(7), 35–42 (1998).

[3] G.K. Wallace, The JPEG still picture compression standard, *Comm. of the ACM*, 34(4), 30–44 (1991).

[4] P. Fränti, O. Nevalainen, T. Kaukoranta, Compression of digital images by block truncation coding: a survey, *The Computer Journal*, 37(4), 308–333 (1994).

[5] A. Jacquin, Fractal image coding: a review, *Proceedings of the IEEE*, 81(10), 1451–1465 (1993).

[6] M.F. Barnsley, L.P. Hurd, Fractal Image Compression, *AK Peters*, Ltd. Wellesley, Massachusetts (1993).

[7] K. Culik II, J. Kari, Image compression using weighted finite automata, *Computer and Graphics*, 17, 305–313 (1993).

[8] J. Valantinas, Apie baigtinių automatų teorijos taikymą apdorojant dvimačius vaizdus, *LMD XXVIII konferencijos darbai* (spec. Liet. matem. rink. priedas), Vilnius, „Technika", 336–340 (1997).

[9] N. Ahmed, K.R. Rao, *Orthogonal Transforms for Digital Signal Processing*, Springer Verlag, Berlin, Heidelberg-New York (1975).

[10] J. Valantinas, N. Morkevičius, Smoothness analysis of two-dimensional gray-level images, *Information Technology and Control*, Kaunas, „Technologija", 1(14), 15–24 (2000).

Stating and solving digital image similarity problem

N. Morkevičius, J. Valantinas

Digital image similarity problem is posed. Necessary conditions (criterion) for the existence of similar digital d-dimensional images are formulated and proved. Areas of application of the criterion are mentioned.