

КОГЕРЕНТНІ ОЦІНКИ ІМОВІРНІСНИХ ХАРАКТЕРИСТИК РИТМІЧНИХ ПРОЦЕСІВ ІЗ НЕВІДОМИМ ПЕРІОДОМ

Віктор Михайлишин, Ігор Яворський

Фізико-механічний інститут ім. Г.В.Карпенка НАН України
79601, Україна, Львів, МСП, вул. Наукова, 5
Телефон: 8 0322 63-33-55
Факс: 8 0322 64-94-27

Електронна пошта: mykhai@ipm.lviv.ua, iavor@ipm.lviv.ua

Розглянуто властивості когерентного методу оцінювання перших двох моментних функцій періодично нестационарного випадкового процесу з невідомим періодом стохастичної повторюваності. Отримано вирази для зміщення та дисперсії когерентних оцінок імовірнісних характеристик ритмічних сигналів, коли оцінки періоду знаходились на основі компонентних статистик.

ВСТУП

Метод когерентного оцінювання імовірнісних характеристик періодично корельованих випадкових процесів (ПКВП) дає змогу здійснювати оброблення сигналів безпосередньо у часовій області. Він ґрунтується на усередненні відліків реалізації через період корельованості. Попереднє оцінювання періоду впливає на властивості когерентних оцінок імовірнісних характеристик ПКВП.

КОГЕРЕНТНА ОЦІНКА МАТЕМАТИЧНОГО СПОДІВАННЯ

Когерентна оцінка математичного сподівання періодично корельованого випадкового процесу (ПКВП) при апіорі невідомому періоді корельованості визначається статистикою

$$m(t) = \frac{1}{N} \sum_{n=0}^{N-1} \xi(t + n\tau_m), \quad (1)$$

де N - кількість відліків, що усереднюються, τ_m - оцінка періоду корельованості.

У роботі [1] показано, що для визначення періоду корельованості можуть використовуватися функціонали, вид яких, аналогічний компонентним оцінкам відповідних імовірнісних характеристик ПКВП:

$$\tilde{m}(t, \tau) = \sum_{l=-N_1}^{N_1} \tilde{m}_l(\tau) e^{i\frac{2\pi}{\tau} t}, \quad (2)$$

де

$$\tilde{m}_l(\tau) = (2\theta)^{-1} \int_{-\theta}^{\theta} \xi(t) e^{i\frac{2\pi}{\tau} t} dt$$

Математичне сподівання функціоналу (2) визначає його регулярну складову

$$\tilde{S}(\tau) = E\tilde{m}(t, \tau) = \frac{1}{2\theta} \sum_{l=-N_1}^{N_1} \int_{-\infty}^{\infty} m(s) e^{i\frac{2\pi}{\tau}(t-s)} ds.$$

Відповідна випадкова складова має вигляд

$$\tilde{N}(\tau) = \tilde{m}(t, \tau) - \tilde{S}(\tau) = \frac{1}{2\theta} \sum_{l=-N_1}^{N_1} \int_{-\infty}^{\infty} \xi(s) e^{i\frac{2\pi}{\tau}(t-s)} ds.$$

Оцінка періоду корельованості ПКВП τ_m є розв'язком нелінійного рівняння

$$\left[\frac{dS(\tau)}{d\tau} + \varepsilon \frac{dN(\tau)}{d\tau} \right]_{\tau=\tau_m} = 0, \text{ де } S(\tau) = \frac{\tilde{S}(\tau)}{\tilde{S}(T)},$$

$$N(\tau) = \frac{N(\tau)}{\sqrt{EN^2(\tau)}}, \quad \varepsilon = \frac{\sqrt{EN^2(\tau)}}{\tilde{S}(T)}.$$

При виконанні умови $\lim_{|u| \rightarrow \infty} b(t, u) = 0$, $\varepsilon \rightarrow 0$ при

$\theta \rightarrow \infty$. Для достатньо великих θ параметр $\varepsilon \ll 1$ і розв'язок нелінійного рівняння знаходимо у вигляді ряду

$$\tau_m = \tau_0 + \varepsilon \tau_1 + \varepsilon^2 \tau_2 + \dots$$

Невідомі величини першого наближення $\tau_0 = T + \Delta T$ і τ_1 знаходимо із системи лінійних

рівнянь: $\Delta T = -\frac{s_1}{s_2}$, $\tau_1 = -\frac{n_1}{s_2}$, де

$$s_i = \left[\frac{d^i S(\tau)}{d\tau^i} \right]_{\tau=T}, \quad n_i = \left[\frac{d^i N(\tau)}{d\tau^i} \right]_{\tau=T}$$

Представимо функціонал (1) у вигляді

$$m(t) = \frac{1}{N} \sum_{n=0}^{N-1} \int_{-\infty}^{\infty} \xi(t + nx) \delta(x - \tau_m) dx.$$

Скористаємося формальним розкладом δ -функції у ряд Тейлора

$$\delta(x - \tau_m) = \delta(x - T) - \delta'(x - T)(\tau_m - T) + \frac{1}{2} \delta''(x - T)(\tau_m - T)^2 - \dots$$

Оскільки оцінка періоду корельованості подана у вигляді ряду по степенях малого параметра ε , обмежуючись першим наближенням, запишемо

$$\delta(x - \tau_m) = \delta(x - T) - \varepsilon \tau_1 \delta'(x - T).$$

Тоді

$$\begin{aligned} m(t) &= \frac{1}{N} \sum_{n=0}^{N-1} \int \xi(t + nx) [\delta(x - T) - \varepsilon \tau_1 \delta'(x - T)] dx = \\ &= \frac{1}{N} \sum_{n=0}^{N-1} \xi(t + nT) - \varepsilon \tau_1 \left[\frac{1}{N} \sum_{n=0}^{N-1} \int \xi(t + nx) \delta'(x - T) dx \right]. \end{aligned}$$

З властивостей похідної δ -функції маємо

$$\int_{-\infty}^{\infty} \xi(t + nx) \delta'(x - T) dx = -\xi'_x(t + nT).$$

Перетворимо статистику (1) до вигляду

$$m(t) = \frac{1}{N} \sum_{n=0}^{N-1} \xi(t + nT) + \varepsilon \tau_1 \frac{1}{N} \sum_{n=0}^{N-1} \xi'_x(t + nT). \quad (3)$$

Отримане в [1] співвідношення, що визначає перше наближення

$$\begin{aligned} \tau_1 &= -\frac{n_1}{s_2} = -\frac{1}{\sqrt{EN^2(T) s_2}} \times \\ &\times \left[\frac{i}{2\theta T} \sum_{l=-N_1}^{N_1} \Lambda_l \int_{-\theta}^{\theta} (s-t) \xi(s) e^{i\Lambda_l(t-s)} ds \right] \end{aligned}$$

підставимо в (3), і в результаті маємо статистику

$$\begin{aligned} m(t) &= \frac{1}{N} \sum_{n=0}^{N-1} \xi(t + nT) - \frac{i}{2N\theta T s_2} \sum_{n=0}^{N-1} \sum_{l=-N_1}^{N_1} \Lambda_l \times \\ &\times \int_{-\theta}^{\theta} (s-t) \xi'_x(t + nT) \xi(s) e^{i\Lambda_l(t-s)} ds \end{aligned}$$

Знайдемо її математичне сподівання

$$\begin{aligned} Em(t) &= m(t) - \frac{i}{2N\theta T s_2} \sum_{n=0}^{N-1} \sum_{l=-N_1}^{N_1} \Lambda_l \times \\ &\times \int_{-\theta}^{\theta} (s-t) E \xi(s) \left[\frac{\partial}{\partial x} \xi(t + nx) \right]_{x=T} e^{i\Lambda_l(t-s)} ds \end{aligned}$$

і конкретизуємо вид співмножника

$$\begin{aligned} E \xi(s) \left[\frac{\partial}{\partial x} \xi(t + nx) \right]_{x=T} &= \frac{\partial}{\partial x} b(t + nx, s - t - nx)_{x=T} = \\ &= \left[\frac{\partial b(s, t - s + nx)}{\partial x} \right]_{x=T} \end{aligned}$$

Таким чином, у першому наближенні оцінка математичного сподівання, отримана на основі статистики (1) із врахуванням того, що оцінка періоду корельованості знаходиться на основі функціоналу (2) виявляється зміщеною. Величина зміщення визначається таким співвідношенням:

$$\begin{aligned} \varepsilon_1[m(t)] &= -\frac{i}{2N\theta T s_2} \sum_{l=-N_1}^{N_1} \Lambda_l \times \\ &\times \sum_{n=0}^{N-1} \int_{-\theta}^{\theta} (s-t) \left[\frac{\partial b(s, t - s + nx)}{\partial x} \right]_{x=T} e^{i\Lambda_l(t-s)} ds \end{aligned}$$

Легко переконатися у тому, що при $N \rightarrow \infty$ зміщення $\varepsilon_1[m(t)] \rightarrow 0$. Тому така оцінка асимптотично незміщена.

Тепер дослідимо вплив випадкових флуктуацій на характеристики оцінки (1). Для цього визначимо її другий центральний момент

$$D[m(t)] = Em^2(t) - [Em(t)]^2 \quad (4)$$

Подано статистику (1) у вигляді

$$\begin{aligned} m(t) &= \frac{1}{N} \sum_{n=0}^{N-1} \xi(t + nT) - \frac{i}{2N\theta T s_2} \sum_{l=-N_1}^{N_1} \Lambda_l \times \\ &\times \sum_{n=0}^{N-1} \int_{-\theta}^{\theta} (s-t) \left[\frac{\partial}{\partial x} \xi(t + nx) \right]_{x=T} \xi(s) e^{i\Lambda_l(t-s)} ds + \\ &+ m(t) - \frac{i}{2N\theta T s_2} \sum_{l=-N_1}^{N_1} \Lambda_l \times \\ &\times \sum_{n=0}^{N-1} \int_{-\theta}^{\theta} (s-t) \left[\frac{\partial}{\partial x} m(t + nx) \right]_{x=T} \xi(s) e^{i\Lambda_l(t-s)} ds \end{aligned}$$

У припущенні гаусового ПКВП обчислюємо згідно (4) дисперсію оцінки математичного сподівання

$$\begin{aligned} D[m(t)] &= \frac{1}{N} \sum_{n=-N+1}^{N-1} \left(1 - \frac{|n|}{N}\right) b(t, nT) - \\ &- \frac{1}{4N^2 \theta^2 T^2 s_2^2} \sum_{l, k=-N_1}^{N_1} \Lambda_l \Lambda_k e^{i(l-k)\frac{2\pi}{T}} \times \\ &\times \sum_{n, m=0}^{N-1} \int_{-\theta}^{\theta} \int_{-\theta}^{\theta} (s_1 - t)(s_2 - t) \times \end{aligned} \quad (5)$$

$$\begin{aligned} &\times \left[\frac{\partial^2}{\partial x \partial y} \left[b(t + nx, my - nx) b(s_1, s_2 - s_1) + \right. \right. \\ &\left. \left. b(s_2, t - s_2 + nx) b(s_1, t - s_1 + my) \right]_{x, y=T} \right] \times \\ &\times e^{i\frac{2\pi}{T}(ks_2 - ls_1)} ds_1 ds_2 - \\ &- \frac{i}{N^2 \theta T s_2} \sum_{l=-N_1}^{N_1} \Lambda_l e^{i\Lambda_l t} \times \\ &\times \sum_{m, n=0}^{N-1} \int_{-\theta}^{\theta} (s-t) \left[\frac{\partial m(t + mx)}{\partial x} \right]_{x=T} b(t, s - t - nT) e^{i\Lambda_l(t-s)} ds \end{aligned}$$

Аналіз співвідношення (5) показує, що при $N \rightarrow \infty$ та $\theta \rightarrow \infty$ другий центральний момент оцінки прямує до нуля. Таким чином, оцінка (1) слухна.

КОГЕРЕНТНА ОЦІНКА КОРЕЛЯЦІЙНОЇ ФУНКЦІЇ

Попереднє оцінювання періоду корельованості ПКВП впливає також на якість когерентної оцінки кореляційної функції. Припустимо, що період визначається за максимумом функціоналу математичного сподівання (2). Розглянемо когерентну оцінку кореляційної функції

$$\hat{b}(t, u) = \frac{1}{N} \sum_{n=0}^{N-1} \overset{\circ}{\xi}(t + n\tau_m) \overset{\circ}{\xi}(t + u + n\tau_m). \quad (6)$$

Подамо статистику (6) в інтегральному вигляді, використовуючи фільтруючу властивість δ -функції і, розкладаючи її у ряд Тейлора. У першому наближенні по малому параметру ε отримаємо

$$\begin{aligned} \hat{b}(t, u) &= \frac{1}{N} \sum_{n=0}^{N-1} \int_{-\infty}^{\infty} \overset{\circ}{\xi}(t + nx) \overset{\circ}{\xi}(t + u + nx) \delta(x - \tau_m) dx = \\ &= \frac{1}{N} \sum_{n=0}^{N-1} \overset{\circ}{\xi}(t + nT) \overset{\circ}{\xi}(t + u + nT) - \\ &- \varepsilon \tau_1 \frac{1}{N} \sum_{n=0}^{N-1} \int_{-\infty}^{\infty} \overset{\circ}{\xi}(t + nx) \overset{\circ}{\xi}(t + u + nx) \delta'(x - T) dx = \\ &= \frac{1}{N} \sum_{n=0}^{N-1} \overset{\circ}{\xi}(t + nT) \overset{\circ}{\xi}(t + u + nT) + \\ &+ \varepsilon \tau_1 \frac{1}{N} \sum_{n=0}^{N-1} \left[\frac{\partial}{\partial x} \left[\overset{\circ}{\xi}(t + nx) \overset{\circ}{\xi}(t + u + nx) \right]_{x=T} \right]. \end{aligned}$$

Оскільки $\tau_1 = -\frac{n_1}{s_2}$, то для гаусового ПКВП має місце співвідношення

$$E\hat{b}(t, u) = b(t, u).$$

Отже, в першому наближенні оцінка (6) незміщена.

Визначимо тепер дисперсію такої оцінки.

$$\begin{aligned} E[\hat{b}(t, u)]^2 &= \\ &= \frac{1}{N^2} \sum_{n, m=0}^{N-1} \left[b(t, (m-n)T) b(t + u, (m-n)T) + \right. \\ &\quad \left. + b(t, u + (m-n)T) b(t, u + (n-m)T) \right] + \\ &+ \varepsilon^2 E \left[\tau_1^2 \frac{1}{N^2} \sum_{n, m=0}^{N-1} \left[\frac{\partial}{\partial x} \overset{\circ}{\xi}(t + nx) \overset{\circ}{\xi}(t + u + nx) \right]_{x=T} \times \right. \\ &\quad \left. \times \left[\frac{\partial}{\partial y} \overset{\circ}{\xi}(t + my) \overset{\circ}{\xi}(t + u + my) \right]_{y=T} \right]. \end{aligned}$$

Математичне сподівання другого доданка рівне

$$E \left[\tau_1^2 \frac{1}{N^2} \sum_{n, m=0}^{N-1} \left[\frac{\partial}{\partial x} \overset{\circ}{\xi}(t + nx) \overset{\circ}{\xi}(t + u + nx) \right]_{x=T} \times \left[\frac{\partial}{\partial y} \overset{\circ}{\xi}(t + my) \overset{\circ}{\xi}(t + u + my) \right]_{y=T} \right] =$$

$$\begin{aligned} &= - \frac{1}{4\theta^2 T^2 \sqrt{EN^2(T)S_2^2}} \times \\ &\times \sum_{k, l=-N_1}^{N_1} \Lambda_l \Lambda_k e^{i(l-k)\frac{2\pi}{T}t} \int_{-\theta}^{\theta} \int_{-\theta}^{\theta} (s_1 - t)(s_2 - t) \times \\ &\quad \times E \overset{\circ}{\xi}(s_1) \overset{\circ}{\xi}(s_2) e^{i\frac{2\pi}{T}(ks_2 - ls_1)} ds_1 ds_2 \times \\ &\times \frac{1}{N^2} \sum_{n, m=0}^{N-1} \left[\frac{\partial}{\partial x} \overset{\circ}{\xi}(t + nx) \overset{\circ}{\xi}(t + u + nx) \right]_{x=T} \times \\ &\times \left[\frac{\partial}{\partial y} \overset{\circ}{\xi}(t + my) \overset{\circ}{\xi}(t + u + my) \right]_{y=T}. \end{aligned}$$

Враховуючи співвідношення

$$\begin{aligned} E \left[\overset{\circ}{\xi}(s_1) \overset{\circ}{\xi}(s_2) \overset{\circ}{\xi}(t + nx) \overset{\circ}{\xi}(t + u + nx) \times \right. \\ \left. \times \overset{\circ}{\xi}(t + mx) \overset{\circ}{\xi}(t + u + mx) \right] = \\ = b(s_1, s_2 - s_1) G(t + nx, t + u + nx, t + mx, t + u + mx) + \\ + b(s_1, t - s_1 + nx) G(s_2, t + u + nx, t + mx, t + u + mx) + \\ + b(s_2, t - s_1 + u + nx) G(s_2, t + nx, t + mx, t + u + mx) + \\ + b(s_1, t - s_2 + mx) G(s_2, t + nx, t + u + nx, t + u + mx) + \\ + b(s_1, t - s_2 + u + mx) G(s_2, t + nx, t + u + nx, t + mx), \end{aligned}$$

де

$$\begin{aligned} G(s_1, s_2, s_3, s_4) &= E \left[\overset{\circ}{\xi}(s_1) \overset{\circ}{\xi}(s_2) \overset{\circ}{\xi}(s_3) \overset{\circ}{\xi}(s_4) \right] - \\ &- b(s_1, s_2 - s_1) b(s_3, s_4 - s_3) \end{aligned}$$

запишемо остаточний вираз для дисперсії оцінки кореляційної функції:

$$\begin{aligned} D[\hat{b}(t, u)] &= \\ &= \frac{1}{N^2} \sum_{m, n=0}^{N-1} G(t + nT, t + u + nT, t + mT, t + u + mT) - \\ &- \frac{1}{4N^2 \theta^2 T^2 S_2^2} \sum_{k, l=-N_1}^{N_1} \Lambda_l \Lambda_k e^{i(l-k)\frac{2\pi}{T}t} \times \\ &\times \int_{-\theta}^{\theta} \int_{-\theta}^{\theta} (s_1 - t)(s_2 - t) e^{i\frac{2\pi}{T}(ks_2 - ls_1)} \times \\ &\times \left[\frac{\partial^2}{\partial x \partial y} \left[b(s_1, s_2 - s_1) \times \right. \right. \\ &\quad \times G(t + nx, t + u + nx, t + mx, t + u + mx) + \\ &\quad + b(s_1, t - s_1 + mx) G(s_2, t + u + nx, t + mx, t + u + mx) + \\ &\quad + b(s_1, t - s_1 + u + nx) G(s_2, t + nx, t + mx, t + u + mx) + \\ &\quad + b(s_1, t - s_1 + mx) G(s_2, t + nx, t + u + nx, t + u + mx) + \\ &\quad \left. \left. + b(s_1, t - s_2 + u + mx) \times \right. \right. \\ &\quad \left. \left. \times G(s_2, t + nx, t + u + nx, t + mx) \right] \right]_{x, y=T} ds_1 ds_2. \end{aligned}$$

Метод оцінки періоду за статистикою (2) при $m(t) \approx 0$ непридатний. У цьому випадку слід використати кореляційний функціонал

$$b(t, u, \tau) = \sum_{l=-N_2}^{N_2} B_l(u, \tau) e^{i \frac{2\pi}{\tau} l t} \quad (7)$$

де

$$B_l(u, \tau) = (2\theta)^{-1} \int_{-\theta}^{\theta} \xi(t) \xi(t+u) e^{-i \frac{2\pi}{\tau} l t} dt,$$

$$\xi(t) = \xi(t) - m(t).$$

Якщо період визначати за абсолютним максимумом статистики (7), то перше наближення оцінки періоду має вигляд

$$\tau_1 = -\frac{1}{s_2} \left[\frac{i}{2\theta T} \sum_{l=-N_2}^{N_2} \Lambda_l \int_{-\theta}^{\theta} (s-t) \eta(s) e^{i \frac{2\pi}{T} (t-s)} ds \right], \text{ де}$$

$\eta(s) = \xi(s) \xi(s+u) - b(s, u)$. Тоді із врахуванням співвідношення

$$E \left[\eta(s) \xi(t+nx) \xi(t+u+nx) \right] =$$

$$= E \left[\xi(s) \xi(s+u) - b(s, u) \right] \times$$

$$\times E \eta(s) \left[\frac{\partial}{\partial x} \xi(t+nx) \xi(t+u+nx) \right]_{x=T} ds$$

знайдемо математичне сподівання

$$E \left[\tau_1 \frac{1}{N} \sum_{n=0}^{N-1} \left[\frac{\partial}{\partial x} \left(\xi(t+nx) \xi(t+u+nx) \right) \right]_{x=T} \right] =$$

$$= -\frac{i}{2Ns_2\theta T} \sum_{l=-N_2}^{N_2} \Lambda_l e^{i \frac{2\pi}{T} l t} \sum_{n=0}^{N-1} \int_{-\theta}^{\theta} (s-t) e^{-i \frac{2\pi}{T} s} \times$$

$$\times E \eta(s) \left[\frac{\partial}{\partial x} \xi(t+nx) \xi(t+u+nx) \right]_{x=T} ds.$$

Зміщення когерентної оцінки кореляційної функції (6) у першому наближенні має вигляд

$$\varepsilon_1[b(t, u)] = -\frac{i}{2Ns_2\theta T} \sum_{l=-N_2}^{N_2} \Lambda_l e^{i \frac{2\pi}{T} l t} \sum_{n=0}^{N-1} \int_{-\theta}^{\theta} (s-t) \times$$

$$\times \left[\frac{\partial}{\partial x} G(s, s+u, t+nx, t+u+nx) \right]_{x=T} e^{-i \frac{2\pi}{T} s} ds.$$

Дисперсія такої оцінки визначається восьмим моментом, котрий у випадку гаусового ПКВП виражається через кореляційну функцію. Тому й тут при виконанні умови $\lim_{|u| \rightarrow \infty} b(t, u) = 0$ оцінка (6)

є слушною.

КОГЕРЕНТНІ ОЦІНКИ ПРИ СПРОЩЕНІЙ ПРОЦЕДУРІ ВИЗНАЧЕННЯ ПЕРІОДУ

Нехай оцінка періоду ритміки ПКВП знаходиться шляхом максимізації статистики виду

$$m_l(\tau) = \frac{1}{\theta} \int_{-\theta}^{\theta} \xi(t) \cos l \frac{2\pi}{\tau} t dt \quad (8)$$

Тоді маємо перше наближення оцінки періоду корельованості у вигляді

$$\tau_1 = -\frac{1}{s_2} \left[\frac{\Lambda_l}{\theta T} \int_{-\theta}^{\theta} s \xi(s) \sin \Lambda_l s ds \right].$$

Оскільки

$$E \left[\tau_1 \sum_{n=0}^{N-1} \left[\frac{\partial}{\partial x} \xi(t+nx) \right]_{x=T} \right] =$$

$$= -\frac{\Lambda_l}{s_2^2 \theta T} \sum_{n=0}^{N-1} \int_{-\theta}^{\theta} s \left[\frac{\partial}{\partial x} E \xi(t+nx) \xi(s) \right]_{x=T} \sin \Lambda_l s ds =$$

$$= -\frac{\Lambda_l}{s_2^2 \theta T} \sum_{n=0}^{N-1} \int_{-\theta}^{\theta} s \left[\frac{\partial b(s, t-s+nx)}{\partial x} \right]_{x=T} \sin \Lambda_l s ds,$$

то зміщення оцінки (1) у даному випадку з точністю до величини першого порядку малого параметра ε рівне

$$\varepsilon_1[m(t)] = -\frac{\Lambda_l}{N\theta T s_2} \sum_{n=0}^{N-1} \int_{-\theta}^{\theta} s \left[\frac{\partial b(s, t-s+nx)}{\partial x} \right]_{x=T} \times$$

$$\times \sin \Lambda_l s ds$$

Дисперсія оцінки математичного сподівання (1) у першому наближенні для гаусового ПКВП має вигляд:

$$D[m(t)] = \frac{1}{N} \sum_{n=N+1}^{N-1} \left(1 - \frac{|n|}{N} \right) b(t, nT) + \frac{\Lambda_l^2}{N^2 \theta^2 T^2 s_2^2} \times$$

$$\times \sum_{n, m=0}^{N-1} \int_{-\theta}^{\theta} \int_{-\theta}^{\theta} s_1 s_2 \left[\frac{\partial^2}{\partial x \partial y} G(s_1, s_2, t+nx, t+my) \right]_{x, y=T} \times$$

$$\times \sin \Lambda_l s_1 \sin \Lambda_l s_2 ds_1 ds_2 -$$

$$- \frac{1}{N^2 \theta T s_2} \sum_{m, n=0}^{N-1} \int_{-\theta}^{\theta} s b(s, t-s+nT) \times$$

$$\times \left[\frac{\partial}{\partial x} m(t+mx) \right]_{x=T} \sin \Lambda_l s ds.$$

Аналіз цього співвідношення показує, що оцінка (1) за умови оцінювання періоду корельованості за статистикою (8) також є слушною.

ВИСНОВКИ

Показано слушність когерентних оцінок імовірнісних характеристик ПКВП у випадку компонентного оцінювання періоду корельованості.

Отримані співвідношення для зміщення оцінок дають змогу контролювати величину систематичної похибки. Метод ефективний при вібродіагностиці обертових машин.

ЛІТЕРАТУРА

1. Михайлишин В.Ю., Яворский И.Н. "Компонентное оценивание периода стохастической повторяемости ритмических сигналов" // Автоматика. - 1993. - №4. - С.43-47

SPEEDING-UP THE FRACTAL COMPRESSION WITH CLUSTERING

Nikolay N. Ponomarenko¹, Karen Egiazarian², Vladimir V. Lukin¹, Jaakko T. Astola²

¹Dept 504, State Aerospace University (Kharkov Aviation Institute)
17 Chkalova Street, 61070, Kharkov, UKRAINE,
Tel/fax + 38 0572 441186, E-mail: lukin@mmds.kharkov.ua

²Tampere International Center for Signal Processing, Tampere University of Technology
P.O.Box-553, FIN-33101, Tampere, FINLAND, Tel. +358 3 365 3860,
Fax +358 3 365 3857, E-mail karen@cs.tut.fi, jta@cs.tut.fi

ABSTRACT

A problem of speeding-up the fractal compression of images is discussed. The techniques based on clustering are considered. An approach to clustering based on median split method is proposed. A procedure of reassigning the domain blocks between clusters is shown to improve the PSNR or to increase the compression ratio. The optimal number of clusters that minimizes the computation efforts for this procedure is derived. Some practical limitations of clustering are discussed and the particular ways out are given.

1. INTRODUCTION

Fractal image compression is an intensively developing topic in modern image processing [1]. Along with wavelet compression [2] it has drawn attention of many researchers and exhibited some advantages as a challenge to already used standard compression techniques.

The problems in fractal compression of still images are a very large amount of computations (and time) required and, thus, a necessity of speeding up the compression process without essential loss in quality [3], [4] as well as the recovered image enhancement. This paper mainly addresses the first problem dealing with speeding up the search of optimal domain blocks for range blocks by means of clustering.

There exist a lot of different strategies of domain search [3], but the strategy based on clustering appears to be the most reasonable. First, due to its application it becomes possible to obtain a rather great benefit in compression speeding up with reasonably small degradation of recovered image quality (or, if desirable, with small reduction of compression ratio). Second, this technique is easily adjusted to other methods of compression speeding up or image quality enhancement.

So below we first consider an approach described in [4] more in detail since our proposed methods are in-

tended to improve it. Then an application of median split algorithm to data clustering is considered and its basic steps are given. Its advantages and drawbacks are analyzed. The procedure of reassigning the domain blocks between clusters is put forward and shown to be rather effective for PSNR increasing. All the benefits and drawbacks are illustrated using numerical simulation results for a set of traditional test images.

2. BASIC STEPS IN CLUSTERING

In standard fractal compression a dominant part of time (almost 100%) is spent on finding the correspondence between the compressed range blocks and the domain blocks [1]. Denoting a number of range blocks as K and a number of domain blocks as N the complexity of compression becomes of $O(K*N)$ since it is required to compare each range block with each domain block. The use of clustering permits to avoid a linear search between all variants due to comparison of each range block with only the cluster centers and, then, with domain blocks of the cluster being the closest to the considered range blocks.

To perform these operations Wein and Blake [4] propose the following steps to be done in fractal compression: 1) the calculation of the number of clusters optimal for given number of domain and range blocks, 2) the division of the virtual codebooks (a domain block set) on clusters using the selected clustering method; 3) the search of the nearest cluster (according to its center) for each range block; 4) the search of the optimal domain block for each range block within the determined cluster.

The application of this scheme permits to decrease the complexity of domain search if the dimensions of virtual codebook clusters are equal. If initially it was $O(K*N)$, then, it becomes of order $O(K*M+K*N/M)+Ocl$ where M is the number of clusters and Ocl describes the complexity of clustering algo-

rithm. As shown in [4], the value of the sum $K*M+K*N/M$ is minimized in case of $M=N^{0.5}$. The complexity is expressed then as $O(2*K*N^{0.5})+Ocl$.

Taking this into account one gets the following requirements to clustering algorithm:

1) it should provide the predetermined number of clusters desirably of equal size;

2) the complexity of clustering algorithm has to be considerably less than $O(K*N-K*M-N*M)$ in order to ensure considerable benefit in speeding up the fractal compression in comparison to methods based on linear (full) search;

3) the algorithms are to be able to operate with vectors having large dimensions (like 256 or 1024).

The first and the second requirements can be satisfied for several algorithms. Unfortunately, among these algorithms practically no one is able to satisfy the last condition. This results in noticeable degradation of recovered image quality in comparison to compression based on linear search. In particular, in [4] it is proposed to apply the *pnn*-algorithm with complexity $O(N*logN)$. However, to our opinion, the application of other methods could be more expedient.

3. MEDIAN SPLIT CLUSTERING

The median split clustering algorithm contains the following steps:

1) the calculation of the value variance for each vector coordinate (component);

2) the calculation of the median for the vector coordinate having the largest variance;

3) the division of domain field into two clusters having almost equal numbers of domains (the domains with the value for the considered coordinate less than median are referred to one cluster and otherwise);

4) until the number of clusters reaches the required value the steps 1-3 are repeated.

If the number of clusters is not the power of 2 the obtained clusters do not have equal sizes. However, this drawback is compensated by the higher speed of the considered algorithm. Actually, the median split algorithm complexity is $O(N*logM)$. That is why, if, for instance, the number of domain blocks is 65536 and the number of clusters is 256 the median split algorithm of clustering is approximately two times faster than *pnn*-algorithm.

The abovementioned technique of speeding up the fractal compression permits to get a very good (about $0.5N^{0.5}$) benefit in time needed. One more advantage is that the benefit almost does not depend on the range block number, but only upon the domain block number. Unfortunately, a considerable loss in recovered image quality (characterized by PSNR) is observed.

Let us confirm this by numerical simulation results. Table 1 presents the results of fractal compression for the test image (512x512) Lenna using the linear search

method and two schemes of clustering – without and with reassigning the domain blocks (See next Section). The compression method performance is characterized by the values of PSNR for the same compression ratio (CR). It is seen (columns 3 and 4 of Table 1) that with the cluster number increasing the loss in PSNR becomes larger and for optimal value of M (256) the loss is of about 0.7 dB and it can be even larger. We have checked this for other traditional test images like “Mandrill” and “Peppers” and observed both the same tendency and the same order losses in PSNR.

One of the reasons is the necessity to perform the vector quantization for vectors with large dimensions. For median split algorithm the division of the vector field is done using only few vector coordinates (components) while the other ones remain unanalyzed. This leads to considerable inhomogeneity of the cluster content and the high probability of false decisions in clustering and comparison making. When the number of clusters increases and computations reduce, the PSNR decreases since the range blocks and the domain blocks the most close to them could occasionally drop into different clusters.

4. DOMAIN BLOCK REASSIGNING

Let us consider what can be done to get around this problem or, at least, to minimize its consequences. First, let us take into account the fact that for range blocks we find the closest cluster center and for domain blocks this procedure is not executed. They are considered to belong to that cluster to which they have been referred after initial clustering. But in practice a large percentage of domain blocks occur to be placed more close to the centers of another, neighbor clusters. So if after the initial clustering one makes for each domain the search of the closest cluster center the quality of finding the correspondence between the range blocks and the domain blocks can increase (although the complexity of search then increases by $O(N*M)$).

In this case the scheme of clustering for fractal compression is the following:

1) the calculation of cluster number optimal for given numbers of domain and range blocks;

2) the division of virtual codebooks (the domain block set) into clusters using the selected clustering algorithm;

3) the search of the nearest cluster (according to the distance to its center) for each range block;

4) the search of the nearest cluster (according to the distance to its center) for each domain block and translation of the domain block to this cluster if needed;

5) the search of the optimal domain block for each range block within the determined cluster.

The application of this scheme results in considerable decreasing of loss in the recovered image quality caused by clustering. Table 1 presents the corresponding

results (see columns 5 and 6). The range blocks for all the considered methods have the size 8x8, the domain block number is 62001, the clustering has been performed by median split algorithm.

It is seen from this Table that the increasing of the cluster number for the proposed clustering method with reassigning also results in reduction of PSNR in comparison to linear search. However, if the domain block reassigning is not used the loss in PSNR is about 0.3 dB larger than if the reassigning is applied (compare the results in columns 3 and 5 or analyze data in column 7). Moreover, the number (percentage) of domain blocks that have changed the cluster is large – from 60 to 72%.

Note that because of changing the clustering complexity – for clustering with domain block reassigning it became $O(K*M+K*N/M+K*N/M)+Ocl$ – the optimal number of clusters is not equal to $N^{0.5}$. It depends on both the domain and range block numbers. In situations when $N \gg K$ it is reasonable to select $M=K^{0.5}$ and if N and K are of the same order one should use $M=N^{0.5}$. The latter situation takes place in fractal compression using adaptive partition schemes [3] where the domain block number is restricted to get an opportunity of image partitioning on larger number of range blocks and its optimization. However, for majority of practical tasks one has $N \gg K$ and the search complexity can be approximately expressed as $O(K*K^{0.5}+2*N*K^{0.5})+Ocl$.

Therefore, the procedure of domain block reassigning permits to partly solve the problem of recovered image enhancement. However, the dimensions of clusters become not equal and this can lead to some growth of computations to be performed. So one of problems to be considered in future is the application of methods providing better clustering, for instance, due to more perfect vector quantization approaches.

5. OTHER POSSIBLE SOLUTIONS

One heuristical way out is proposed below. Suppose we do not subtract the mean for every domain block before finding the median value and cluster splitting into two parts. Then the domains with similar (approximately equal) coordinate values will fall down into the same cluster. If one deals with mosaic image compression then it is reasonable to assume the following. If the means for different domain blocks are rather close then these domain blocks are with larger probability placed in image not far away from each other than the domains with considerably different means. In turn, it means that the other vector coordinates of these domains can be highly correlated. If these assumptions are valid then the clustering by one coordinate at every step is “supported” by clustering for some other dependent coordinates of the vector.

Let us analyze the obtained numerical simulation results. Some of them are given in Table 2. The time required for finding the range and domain block corre-

spondence for equal size clusters is considered as T . The ratios of computation time in case of clustering with subtraction of mean and without subtraction are denoted as T_s/T and T_w/T , respectively. The number of domain blocks is given in the first column, the PSNR values are computed for range block size 8x8 and 64 clusters.

The test image is the same – Lenna 512x512 pixels. It is worth noting here that all the numerical simulation results in this paper are obtained for the following technique of color-to-gray scale image transformation $Gr=0.3R+0.59G+0.11B$ where R , G and B are the intensities of red, green and blue components of color image, respectively. We would like to mention also that the results (the PSNR values) can be noticeably (even by 1-2 dB) different for other variants of color-to-gray scale transformation.

As seen from Table 2 there is almost no difference in PSNR values for the considered clustering methods. The increase in computations with respect to ideal case is not too large, however, in the latter case (when the mean values are not subtracted) the increase is smaller due to better clustering with almost equal cluster dimensions.

Finally, let us give quantitative data concerning computational load for the linear search and the proposed clustering technique. Table 3 presents the results for three images - Lenna, Mandrill and Peppers. The PSNR values are given for the range block size 8x8, domain block size 16x16. The number of domain blocks is 62001. The optimal number of clusters for the considered case is equal to 62, but for suitability of the clustering algorithm operation it was selected to be 64. Three values of CR have been considered – 5.3, 21.3, and 85.3. We do not present here the time required for clustering expressed in seconds since the CPU depends upon many factors like computer type and rate, etc. That is why the ratio T_s/T_c is given where T_s and T_c are the time interval durations for making up the linear (full) search and search with clustering.

As seen from Table 3, for all the considered cases the difference in PSNR for compression based on linear search or clustering is not large. It does not extend 0.4 dB for small CR and reduces for larger CR values. Of course, the PSNR also depends upon image type and it is the worst (in case of equal CR) for the test image “Mandrill” which contains many textured regions. The most impressive result is the considerable speeding up the computations due to clustering. The benefit is especially large for CR values within the range of 5...20, where the computation time can be decreased by 30...60 times. Just this range of CR values is typical for many practical applications. For larger CR the ratio T_s/T_c decreases but still remains large enough.

CONCLUSIONS

It is shown that due to clustering the computational complexity of the fractal compression can be significantly decreased. The median split method proposed for performing the clustering has been demonstrated to provide an appropriately small loss in PSNR with respect to the linear search fractal compression techniques for wide range of CR and different test images.

REFERENCES

[1] *Fractal Image Encoding and Analysis*, Edited by Yuval Fisher, Berlin Heidelberg, Germany: Springer Verlag, 1998.

- [2] *Wavelet Image and Video Compression*, Edited by Pankaj N. Topiwala, Boston, USA: Kluwer Academic Publishers, 1998.
- [3] B. Wohlberg, G. Jager, "A Review of Fractal Coding Literature", *IEEE Trans. On Image Processing*, Vol. 8, pp. 1716-1729, Dec. 1999.
- [4] C. J. Wein, I.F. Blake, "On the Performance of Fractal Compression with Clustering", *IEEE Trans. On Image Processing*, Vol. 5, pp. 522-526, March 1996.

Table 1. Comparison of performance of different compression methods for test image Lenna, CR=21.3.

Cluster number	PSNR for linear search dB	PSNR for Clustering without Reassigning, dB	PSNR loss because of clustering, dB	PSNR for clustering with reassigning, dB	PSNR loss because of clustering, dB	Loss reduction due to reassigning, dB	Percentage of reassigned domains blocks
32	30.28	29.81	0.47	30.12	0.16	0.31	61.1%
64	30.28	29.75	0.53	30.04	0.24	0.29	65.7%
128	30.28	29.67	0.61	30.01	0.27	0.34	69.7%
256	30.28	29.58	0.7	29.9	0.38	0.32	72.7%

Table 2. Comparison of PSNR and computation time for two considered techniques of clustering.

Domain block number	PSNR, dB (Mean is not subtracted)	T_w/T	PSNR, dB (mean is subtracted)	T_s/T
62001	30.04	1.154	30.06	1.091
50000	30.00	1.126	29.98	1.053
38000	29.93	1.132	29.92	1.039
26000	29.80	1.097	29.80	1.045
20000	29.71	1.084	29.70	1.025
14000	29.57	1.101	29.54	1.006

Table 3. Comparison of PSNR and computation time for compression techniques based on linear search and clustering for different test images.

Test image	CR	PSNR, dB, linear search	Number of clusters	PSNR, dB, with clustering	Difference in PSNR, dB	T_l/T_c
Lenna	5.3	36.34	128	35.95	-0.4	58.5
	21.3	30.28	64	30.06	-0.22	28.4
	85.3	26.00	32	25.80	-0.2	13.8
Mandrill	5.3	26.72	128	26.33	-0.4	57.6
	21.3	21.78	64	21.65	-0.13	28.0
	85.3	19.64	32	19.79	+0.15	14.9
Peppers	5.3	35.54	128	35.30	-0.24	59.0
	21.3	30.44	64	30.22	-0.22	29.4
	85.3	25.53	32	25.37	-0.16	14.8

MYRIAD FILTER PROPERTIES AND PARAMETER SELECTION

Sergey K. Abramov¹, Vladimir V. Lukin¹, Jaakko T Astola²

¹Dept 504, State Aerospace University (Kharkov Aviation Institute)
17 Chkalova Street, 61070, Kharkov, UKRAINE,
Tel/fax + 38 0572 441186, E-mail: lukin@mmds.kharkov.ua

²Tampere International Center for Signal Processing, Tampere University of Technology
P.O.Box-553, FIN-33101, Tampere, FINLAND, Tel. +358 3 365 2923,
Fax +358 3 365 3857, E-mail jta@cs.tut.fi

ABSTRACT

Peculiar properties of myriad filters in case of their application to 1-D signal processing in mixed noise environment are considered. It is shown that the filter output bias and the filter noise suppression effectiveness considerably depend upon filter parameter k , signal behavior and noise characteristics. Some recommendations concerning parameter k selection are given.

1. INTRODUCTION

In processing the signals corrupted by mixed, i.e. additive and impulsive, noise the nonlinear filters have shown themselves to be a useful and efficient tool since they are able simultaneously to remove spikes, suppress additive noise component and preserve signal abrupt changes well enough [1], [2].

A new and recently introduced class of nonlinear data processing algorithms is a myriad filter and its modifications like weighted myriad filters [3], [4]. They have been proposed and designed as nonlinear filters for robust non-gaussian signal processing in impulsive noise environment, in particular, in case of α -stable distributions of noise having heavier tails than gaussian and, for some cases, Laplacian ones.

The performance characteristics of the standard myriad filter strongly depend on the so-called linearity parameter k that controls the outlier-rejection capability of an estimator. At the same time, this parameter determines the dynamic (output bias) and statistical characteristics of the myriad filter. Certainly, all these characteristics also depend upon a scanning window size.

Some of these properties have not been exploited in [3], [4] and for their analysis we apply approaches described in our papers [5], [6]. It is demonstrated, in particular, that myriad filters with rather small k (just under this condition an appropriate robustness is provided) distort piece-wise fragments of signal in peculiar manner and even "amplify" noise for rapidly increas-

ing/decreasing signals. These drawbacks are undesirable and should be avoided. To get around this problem is possible due to proper, compromise, selection of parameter k . In order to perform it we present below the computer simulation results and give the corresponding recommendations based on test signal analysis.

2. MYRIAD FILTER DEFINITION

A sample myriad is an M-estimator of location. Given a set of samples $\{y(i)\}, i=1, \dots, N$, an M-estimator of location is described by expression

$$\hat{\theta}^{\Delta} = \arg \min_{\theta} \sum_{i=1}^N \rho(y(i) - \theta), \text{ where } \rho(\cdot) \text{ is the cost}$$

function. The sample myriad is defined using the cost function $\rho(y) = \log(k^2 + y^2)$. Therefore, for standard myriad filter output the used cost function is

$$\sum_{i=1}^N \log(k^2 + (y(i) - \theta)^2) \text{ where } \{y(i)\} \text{ is formed by}$$

sample values within the scanning window of size N . In other words, the output $x(j)$ of the myriad filter applied to process $\{y(j)\}, j=1, \dots, J, J \gg N$ is written as follows

$$x(j) = \arg \min_{\theta} \sum_{i=j-(N-1)/2}^{j+(N-1)/2} \log(k^2 + (y(i) - \theta)^2). \quad (1)$$

Another, sometimes more convenient, way of the myriad filter output definition is expressed as

$$x(j) = \arg \min_{\theta} \prod_{i=j-(N-1)/2}^{j+(N-1)/2} (1 + (y(i) - \theta)^2 / k^2). \quad (2)$$

It can be easily shown that the sample myriad reduces to the sample mean as $k \rightarrow \infty$. At the same time, as $k \rightarrow 0$ the sample myriad tends to the distribution mode and this is a rather important property. That is why

we used earlier the terms "rather small" and "rather large" values of parameter k . However, one should keep in mind that these relative definitions depend upon signal scale.

Other details concerning the myriad and weighted myriad filter peculiarities can be found in [3], [4]. Our purpose is to attract the reader's attention to specific aspects that could be important in practical applications.

3. CONSIDERED SIGNAL/NOISE MODEL

Let us consider the following signal/noise model of a 1-D sampled data sequence

$$y(j) = \begin{cases} S(j) + n_a(j), & \text{with probability } 1 - P_{imp} \\ S(j) + n_{imp}(j), & \text{with probability } P_{imp}, \end{cases} \quad (3)$$

where $S(j)$ denotes the true signal value of the j -th sample; $n_a(j)$ is the zero mean gaussian additive noise with the variance σ_a^2 ; $n_{imp}(j)$ defines the amplitude of impulsive noise that occurs with the probability P_{imp} and it is supposed that $|n_{imp}| > \sigma_a$.

The considered test signal is presented in Fig.1, test. The behavior of the myriad filter output was studied for different fragments of this signal: 1) a constant signal (for example, the fragment with indices from 10 to 40); 2) a step edge (indices 40-60); 3) piecewise linear curves (indices 90-110 and 190-210); 4) linearly increasing and decreasing signals (110-140 and 160-190); 5) peak maximum (140-160); 6) polynomial maximum (265-285); 7) a specific fragment (240-260). In other words, this set of test fragments covers a rather wide variety of situations taking place in practice. Thus, the analysis performed in this paper characterizes the myriad filter output peculiarities from quite different points of view.

4. OUTPUT BIAS

Myriad filter mean output considerably depends upon noise characteristics even if the noise is gaussian. This property is, in general, typical for many nonlinear filters [1], [2], [5]. However, since for rather large k the myriad filter characteristics approximate the standard mean filter ones, it is not worth considering the myriad filter bias for this case. The parameter k selection to be rather small when essentially nonlinear and robust properties of the myriad filter are provided is of more theoretical and practical interest.

The studies done have shown that in case of small k , considerably less than the step edge height Δh , the myriad filter preserves it better than a standard median filter

with the same N . This conclusion is valid for wide range of additive noise variances $\sigma_a \ll \Delta h$.

In general, the properties of all nonlinear filters also depend upon scanning window size N . Below we analyze only the case $N=9$, however, the character of main dependences is similar for other N within the limits from 5 to 11.

However, very peculiar effects take place in the neighborhoods of indices 100, 200, 250, 300, i.e. for piecewise curve junction points. These effects are illustrated in Fig. 2. When k is rather small (0.01) the myriad filter output bias derived as $\Delta_{myr}(j) = E[x(j)] - S(j)$ can have sign that is opposite to the output bias sign in case of large k . For instance, for the neighborhood of the sample with index 100 the myriad filter output bias is positive but very small for $j < 100$, and for $j = 101, 102$, and 103 it becomes negative (if $k = 0.01$). Finally, with further j increasing the myriad filter mean $E[x(j)]$ quickly "jumps" and becomes approximately equal to $S(j)$ (See Fig. 2,a). Similar effects are observed for $k = 0.01$ in the neighborhoods of sample indices 200, 250 (See Fig.2,b), and 300. At the same time, for $k = 1.0$ the myriad filter output behavior is much more typical and it is quite similar to many other nonlinear (for example, α -trimmed) and linear filters [5] (Consider curves for $k = 1.0$ in Fig.2,a and 2,b). All these effects are explained by the mode selection property of myriad filters with small k .

The Table 1 presents the numerical simulation results for the test signal and different noise characteristics. The values of mean square errors (MSE) have been evaluated for the entire signal (χ_l) and locally for its fragments. For example, χ_{10-40} is the local MSE evaluated for the signal fragment from the index 10 to the index 40 and the MSE χ_{40-60} just corresponds to the step edge neighborhood. The MSEs do not characterize the output bias directly since they also take into account the variance of noise remained after filtering. However, the contribution of output bias into local MSEs is considerable in step edge and junction point neighborhoods. That is why the MSE values give reliable information for analysis of filter output bias for many fragments except constant, linearly increasing/decreasing signal (where bias is zero for case of spike absence) and polynomial maximum fragments. In general, polynomial maximum is preserved by myriad filter well enough if k is not rather large.

The dynamic errors (output bias) in the neighborhood of peak extremum (index 150) depend upon k and usually they slightly increase with parameter k increasing. It should be also mentioned that except of output bias the considerable growth of residual noise variance can be observed in the neighborhoods of junction points (indices 100, 150, 200, 250, 300) if k is rather small.

And this is one more drawback of the myriad filters. As it can be seen from Table 1 for *Case A* (small variance of gaussian noise and spikes are absent), the smallest MSE values in the step edge neighborhood are provided by the myriad filter with $k=0.01$. At the same time for other junction point neighborhoods the better results are observed for myriad filters with $k=1.0$ and $k=10.0$. This means that a trade-off between noise suppression efficiency and detail preservation ability of the myriad filter should be found and it is reachable by means of parameter k selection (variation).

5. NOISE SUPPRESSION EFFICIENCY AND ROBUST PROPERTIES

In case of rather small k the peculiar effects can be observed for constant and linearly increasing/decreasing signal (LIDS) fragments. The noise suppression efficiency of nonlinear filter can be assumed sufficient if it does not differ a lot from the corresponding efficiency of the standard mean filter with the same N [6]. For example, for the best nonlinear noise suppressing filters like Wilcoxon and Hodges-Lehman ones the residual noise variance is not larger than by 30% in comparison to mean filter for LIDS [6]. The myriad filter with rather small k satisfies this requirement for constant signal (analyze χ_{10-40} in Table 1 for *Case A*). But note that in this situation the value of k is comparable with σ_a . If the ratio σ_a/k increases the difference between noise suppression efficiency of the myriad and mean filters for constant signal radically becomes larger. This means that the parameter k selection can be scale dependent and for constant signal the sample scale is determined by noise standard deviation σ_a (if the spikes are absent).

Another tendency is observed for LIDSs. Results in Table 1 (consider $\chi_{110-140}$ and $\chi_{160-190}$ for *Case A*) show that the variance of "residual" noise for $k=0.01$ is larger than even σ_a^2 . That is the myriad filter with rather small k can "amplify" noise. And this is a very serious drawback to be avoided. Now let us take into account that the sample scale in this situation is mainly determined not by noise standard deviation but by parameter $\Delta S_{sc} = N\Delta S$ where $\Delta S = |S(j+1) - S(j)|$. Thorough studies have demonstrated that by selecting $k > 0.3\Delta S_{sc}$ it is possible to appropriate noise suppression efficiency of the myriad filter for LIDS. And the selection $k=0.01$ does not satisfy this inequality for LIDS of the considered test signals. This is also the reason why large residual noise variances are observed for $k=0.01$ in the neighborhoods of samples with indices 150 (peak maximum), 250 and 300 (junction points of polynomial maximum with constant signals). Therefore, the conclu-

sion is that for almost all the signal fragments except step edge neighborhood the proper selection of parameter k can be done on basis of local estimation of sample scale. But one should keep in mind that k increasing leads to worse robustness with respect to spikes and also results in larger output bias.

To confirm the fact that the myriad filter robustness sufficiently depends upon k let us briefly analyze the numerical simulation data for *Case B* (spike presence) in Table 1. As seen, the myriad filter with $k=0.01$ outperforms the myriad filters with $k=1.0$ and 10.0 for some signal fragments in the sense of smaller local MSE. In particular, this takes place for constant signal fragment, step edge neighborhood and even the fragment with indices from 90 to 110 (first piecewise curve junction neighborhood). The main reason are the good robust properties of the myriad filter with $k=0.01$. However, the best, i.e. the smallest MSEs χ_i for entire signal are provided by the myriad filter with $k=1.0$ for both *Case A* and *Case B*. These interesting results serve as a motivation for locally adaptive approach application to signal processing on basis of myriad filters with varied parameters. Several approaches to providing such filter robustness are worth exploring. One of them is the use of robust estimators of scale.

CONCLUSIONS

The definition of the myriad filter is given and its advantages and drawbacks are considered. The basic characteristics of the myriad filter can be tuned by selecting the only parameter k but it should be a compromise decision depending upon the required trade-off between filter robustness, noise suppression efficiency and output bias. Numerical simulation results and performed analysis results give a starting point for further studies.

REFERENCES

- [1] J. T. Astola, P. Kuosmanen, *Fundamentals of Nonlinear Digital Filtering*, New York, NY: CRC Press, Boca Raton, 1997.
- [2] I. Pitas, A. N. Venetsanopoulos, *Nonlinear Digital Filters: Principles and Applications*, New York: Kluwer Academic Publishers, 1990.
- [3] S. Kalluri, G. Arce, "Adaptive weighted myriad filter algorithms for robust signal processing in α -stable noise environments", *IEEE Trans. On Signal Processing*, Vol. 46, pp. 322-334, Feb. 1998.
- [4] S. Kalluri, G. Arce, "Fast algorithms for weighted myriad computation by fixed point search", *IEEE Trans. On Signal Processing*, Vol. 48, pp. 159-171, Jan. 2000.
- [5] S. Abramov, N. Tulyakova, V. Lukin, V. Melnik, "Dynamic characteristics of some nonlinear 1-D signals", *Proceedings of 1999 Finnish Signal Process-*

Table 1. The MSEs for the considered filters

	χ_1	χ_{10-40}	χ_{40-60}	χ_{90-110}	$\chi_{110-140}$	$\chi_{140-160}$	$\chi_{160-190}$	$\chi_{190-210}$	$\chi_{240-260}$	$\chi_{265-285}$
Case A: $\sigma_a^2=0.001$; $P_{imp}=0.00$, $n_{imp}=0.00$										
Noisy	0.00099	0.00102	0.00100	0.00097	0.00098	0.00094	0.00101	0.00095	0.00101	0.00098
$K=0.01$	0.00742	0.00013	0.00015	0.00058	0.00116	0.00244	0.00276	0.00152	0.05298	0.00076
$K=1.0$	0.00357	0.00011	0.02462	0.00030	0.00011	0.00122	0.00011	0.00051	0.01457	0.00042
$K=10.0$	0.00631	0.00011	0.03525	0.00029	0.00011	0.00125	0.00011	0.00051	0.03135	0.00043
Case B: $\sigma_a^2=0.01$; $P_{imp}=0.03$, $n_{imp}=1.00$										
Noisy	0.03961	0.04662	0.03825	0.04058	0.03804	0.03711	0.04083	0.03457	0.03641	0.03998
$K=0.01$	0.01097	0.00218	0.01291	0.00316	0.00435	0.00637	0.00773	0.00506	0.05543	0.00365
$K=1.0$	0.00670	0.00274	0.02879	0.00332	0.00288	0.00345	0.00308	0.00373	0.02041	0.00257
$K=10.0$	0.01134	0.00515	0.04034	0.00541	0.00501	0.00551	0.00508	0.00596	0.03738	0.00455

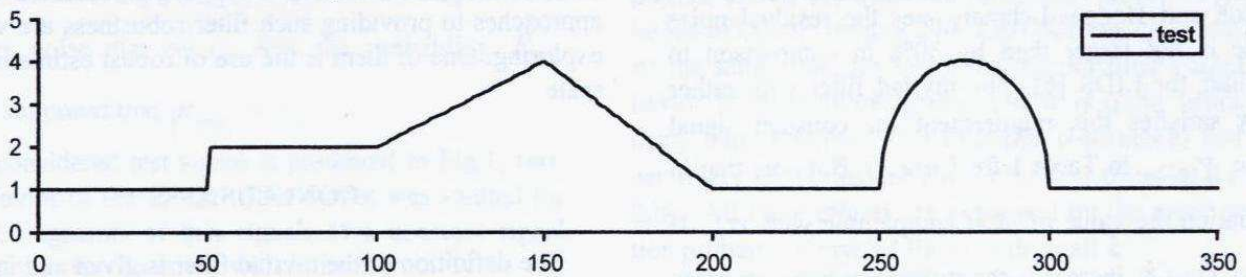
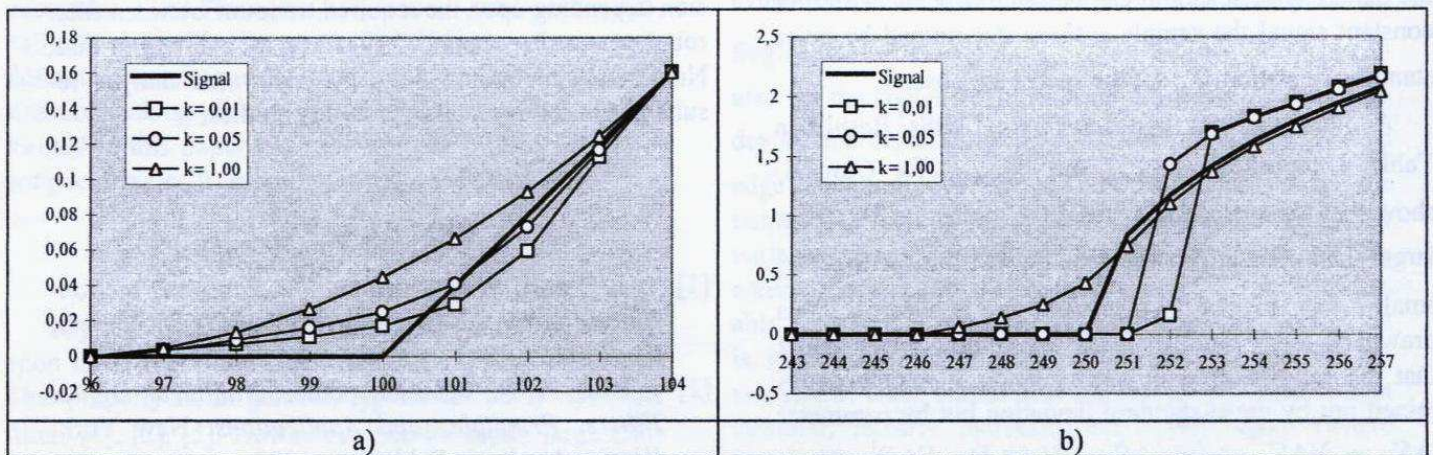


Figure 1. The considered noise-free test signal.



Figures 2. Myriad filter mean outputs for piecewise curve fragments (only gaussian noise, $k=0.01$, 0.05 and 1.0): a) neighborhood of sample 100; b) neighborhood of sample 250.

Методи стискання даних за допомогою перетворення сплесків

Р.В.Лотоцький

Інститут кібернетики ім. Глушкова НАН України, т.(044)211-56-54, roman_l@i.com.ua,

Розглядається застосування перетворення сплесків для однієї з найважливіших проблем обробки інформації – стискання графічних зображень. Наводяться основні факти теорії сплесків в об'ємі, достатньому для конструктивного опису побудованого алгоритму. Результати роботи розробленого програмного комплексу проаналізовані на прикладах стискання зображень реальних відбитків пальців. Обговорюються основні підходи до удосконалення розглянутого методу стискання.

1. Вступ

Методи стискання даних можна розділити на два класи: стискання із втратами і без втрат. Алгоритми з другого класу зазвичай застосовуються для перетворення текстової інформації, тому що в цьому випадку потрібне точне відновлення даних. На відміну від текстової інформації, для графічних напівтонових зображень допускаються втрати при стисканні. У цьому випадку відновлене зображення не буде точною копією оригіналу, але часто така умова і не потрібна, оскільки людське око не сприймає невеликі зміни окремих пікселів. Застосування алгоритмів із втратами дозволяє істотно підвищити коефіцієнт стискання в порівнянні з алгоритмами без втрат. У даній роботі розглядається метод стискання з втратами графічних зображень, основним етапом якого є перетворення сплесків (Wavelet). Результати роботи реалізованого алгоритму показано на прикладах обробки реальних фотографій відбитків пальців.

2. Основні відомості з теорії перетворення сплесків

2.1. Неперервне перетворення сплесків. Неперервне перетворення сплесків визначається таким чином:

$$WF(a,b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} f(t) \overline{\psi\left(\frac{t-b}{a}\right)} dt = \int_{-\infty}^{\infty} f(t) \overline{\psi_{a,b}} dt;$$

Тут, функція ψ задовільняє наступним умовам:

$$\psi \in L_2(\mathbb{R}), \int_{-\infty}^{\infty} \frac{|\hat{\psi}(\xi)|^2}{|\xi|} d\xi < \infty. \quad (1)$$

Тут і далі $\hat{\psi}$ - перетворення Фур'є функції ψ . Базисна система сплесків $\{\psi_{a,b}\}$ складається з усіх комбінацій зсувів і стиснень функції ψ , що у термінах теорії сплесків називається "утворюючим сплеском" (mother wavelet).

$$\psi_{a,b} = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right), \quad (2)$$

Зворотнє перетворення:

$$f(t) = C_{\psi}^{-1} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} WF(a,b) \psi_{a,b}(t) \frac{dad b}{a^2}, \quad (3)$$

$$\text{де } C_{\psi} = 2\pi \int_{-\infty}^{\infty} \frac{|\hat{\psi}(\xi)|^2}{|\xi|} d\xi.$$

2.2. Кратно-масштабний аналіз. Кратно-масштабний аналіз є математичним апаратом для конструктивного опису і побудови різноманітних базисів функцій сплесків. Вперше це поняття було введено Маллатом [1]. Основні положення теорії кратно-масштабного аналізу можна знайти також у [2,3].

Кратно-масштабний аналіз на $L_2(\mathbb{R})$ визначається як послідовність вкладених підпросторів V_j на $L_2(\mathbb{R})$ з такими властивостями:

- 1) $V_j \subset V_{j+1}, j \in \mathbb{Z}$;
- 2) $v(x) \in V_j \Leftrightarrow v(2x) \in V_{j+1}$;
- 3) $v(x) \in V_0 \Leftrightarrow v(x+1) \in V_0$;
- 4) $\overline{\bigcup_{j=-\infty}^{+\infty} V_j} = L_2(\mathbb{R}), \quad \bigcap_{j=-\infty}^{+\infty} V_j = \emptyset,$ де $\overline{\bigcup_{j=-\infty}^{+\infty} V_j}$ -- замикання підпростору $\bigcup_{j=-\infty}^{+\infty} V_j$.
- 5) Існує функція ϕ така, що система функцій $\{\phi(x-k) | k \in \mathbb{Z}\}$ утворює ортонормований базис простору V_0 . Функція ϕ називається масштабуючою функцією.

Із властивостей 2) і 5) випливає, що множина функцій $\varphi_k^j(x) = \sqrt{2^j} \varphi(2^j x - k)$, $j \in \mathbf{Z}$, $k \in \mathbf{Z}$,

утворює ортонормований базис підпростору V_j .

Оскільки $V_0 \subset V_1$, то будь-яку функцію з V_0 можна записати як лінійну комбінацію базисних функцій $\sqrt{2}\varphi(2x - k)$. Отже, маємо

$$\varphi(x) = \sum_k h(k) \sqrt{2} \varphi(2x - k),$$

де коефіцієнти $h(k)$ внаслідок ортогональності $\varphi(x)$ визначаються таким чином:

$$h(k) = (\varphi(x), \sqrt{2}\varphi(2x - k)). \quad (4)$$

Нехай W_j -- ортогональне доповнення підпростору V_j до V_{j+1} , $V_{j+1} = V_j \oplus W_j$. З останньої рівності і властивості 1) можна записати

$$\forall j, j_1, j < j_1: V_{j_1} = V_j \oplus \bigoplus_{k=0}^{j_1-j-1} W_{j_1-k}.$$

Таким чином, враховуючи, що $\bigcup_{j=-\infty}^{+\infty} V_j = L_2(\mathbf{R})$,

маємо $\bigoplus_{j \in \mathbf{Z}} W_j = L_2(\mathbf{R})$.

Головною властивістю кратно-масштабного аналізу є той факт, що якщо послідовність вкладених підпросторів V_j задовольняє умовам 1)-5), то існує можливість побудови такого ортонормованого базису $\{\psi_k^j; j, k \in \mathbf{Z}\}$ простору

$L_2(\mathbf{R})$, $\psi_k^j \in W_j$, що $\forall f \in L_2(\mathbf{R})$:

$$P_{j+1}(f) = P_j(f) + \sum_k (f, \psi_k^j) \psi_k^j,$$

де $P_j(f)$ -- ортогональна проекція функції f на V_j .

Базис $\{\psi_k^j; j, k \in \mathbf{Z}\}$ називається базисом сплесків. Так як $W_j \subset V_{j+1}$, то має місце таке подання функції $\psi^j(x)$:

$$\psi^j(x) = \sum_k g(k) \varphi_k^{j+1}(x). \quad \text{Доведено [2], що}$$

умова ортонормованості базису $\{\psi_k^j; j, k \in \mathbf{Z}\}$ виконується при наступному виборі коефіцієнтів $g(k)$: $g(k) = (-1)^k h(-k+1)$. Таким чином,

формули залежності $\psi^j(x)$ від $\{\varphi_k^{j+1}(x)\}$ і $\varphi^j(x)$ від $\{\varphi_k^{j+1}(x)\}$ мають такий вид:

$$\varphi^j(x) = \sum_k h(k) \varphi_k^{j+1}(x), \quad (5)$$

$$\psi^j(x) = \sum_k (-1)^k h(-k+1) \varphi_k^{j+1}(x). \quad (6)$$

2.3. Швидке перетворення сплесків.

Кратно-масштабний аналіз дозволяє побудувати швидкий алгоритм перетворення сплесків. Так як $V_{j+1} = V_j \oplus W_j$, то функцію $v_{j+1} \in V_{j+1}$ можна записати

$$v_{j+1} = \sum_k \lambda_k^{j+1} \varphi_k^{j+1}(x) = v_j(x) + w_j(x) = \sum_l \lambda_l^j \varphi_l^j(x) + \sum_l \gamma_l^j \psi_l^j(x), \quad (7)$$

де $v_j \in V_j$ і $w_j \in W_j$.

Встановимо зв'язок між коефіцієнтами λ_k^{j+1} , λ_l^j і λ_k^{j+1} , γ_l^j . З формули (5) випливає:

$$\varphi_k^j(x) = 2^{2^j} \varphi(2^j x - k) = 2^{2^j} \sum_l h(l) \varphi(2^{j+1} x - 2k - l) = \sum_l h(l) \varphi_{2k+l}^{j+1}(x) = \sum_l h(l - 2k) \varphi_l^{j+1}(x) \quad (8)$$

Аналогічно:

$$\psi_k^j(x) = \sum_l g(l - 2k) \varphi_l^{j+1}(x) \quad (9)$$

Звідси маємо

$$\lambda_k^j = \sum_l h(l - 2k) \lambda_l^{j+1}, \quad (10)$$

$$\gamma_k^j = \sum_l g(l - 2k) \lambda_l^{j+1}. \quad (11)$$

Обернене перетворення отримуємо аналогічно. З (7)-(9) випливає, що:

$$\lambda_k^{j+1} = \sum_l \lambda_l^j (\varphi_l^j, \varphi_k^{j+1}) + \sum_l \gamma_l^j (\psi_l^j, \varphi_k^{j+1}) = \sum_l h(l - 2k) \lambda_l^j + \sum_l g(l - 2k) \gamma_l^j. \quad (12)$$

Таким чином, якщо відомі коефіцієнти розкладу даної функції по базисних функціях у деякому підпросторі V_j , то, скориставшись (10), (11), можна знайти розклад цієї функції також в усіх підпросторах W_J , $J < j$, а скориставшись (12), можна відновити вихідний розклад.

Формули (10), (11) можна записати в більш компактній і зручній для практичного застосування формі, використовуючи так зване операторне подання фільтрів.

Отже, нехай вектор $a = (a_l)$, $l \in [0, N]$ складається з коефіцієнтів розкладу деякої функції

по базису підпростору V_j . Визначимо оператори H і G :

$$(Ha)_k = \sum_{l=0}^N h(l-2k)a_l, \quad (13)$$

$$(Ga)_k = \sum_{l=0}^N g(l-2k)a_l, \quad \forall k \in \left[0, \left[\frac{N}{2}\right]\right], \quad (14)$$

[•] - ціла частина числа. Ці оператори відповідають перетворенням (10), (11) і виконують один крок у перетворенні сплесків. Відповідні оператори для оберненого перетворення (12) мають такий вигляд:

$$(H^*a)_l = \sum_{k=0}^{\left[\frac{N}{2}\right]} h(l-2k)a_k, \quad (15)$$

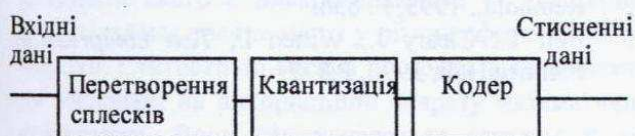
$$(G^*a)_l = \sum_{k=0}^{\left[\frac{N}{2}\right]} g(l-2k)a_k, \quad \forall l \in [0, N]. \quad (16)$$

3. Алгоритм стискання зображень методом сплесків.

На сьогодні найбільш відомим алгоритмом стискання графічних файлів з втратами є алгоритм JPEG [4], в якому використовується косинусоїдалне перетворення Фур'є (КПФ). Цей алгоритм дозволяє стискати графічні зображення в 25-30 разів без суттєвого погіршення якості. Проте якщо зміна інтенсивності пікселів зображення не є плавною (як, наприклад, у відбитках пальців), то коефіцієнт стиску різко падає. Пропонований алгоритм має структуру, подібну алгоритму JPEG, але використовує замість КПФ перетворення сплесків.

У наступних викладах будемо вважати, що зображення є двомірною дискретною функцією, $f(x, y) \in [0, 255]$ значеннями якої є 256 відтінків сірого (0 -- чорний, 255 -- білий)

Процес стискання складається з таких трьох кроків: перетворення сплесків, квантизація, кодування (Рис. 3).



Вхідне зображення розбивається на блоки 8×8 пікселів. Наведені вище операції застосовуються окремо до кожного блоку.

На першому кроці виконується перетворення сплесків. Вище було розглянуто одномірне перетворення сплесків для векторів. У двомірному випадку для перетворення матриці даних застосовуємо одномірне перетворення для кожного рядка, і для отриманої в результаті матриці знову

застосовуємо одномірне перетворення для кожного стовпчика.

Оскільки різні коефіцієнти перетворення по-різному впливають на відновлення вихідного зображення, то, для підвищення ступеня стискання, отримані коефіцієнти w_i квантизуються, тобто діляться на коефіцієнти квантизації: $k_i \cdot r_i = w_i / k_i$. Потім, округлені до найближчого цілого, вони надходять на вхід ентропічного кодера.

При підборі правильної таблиці квантизації $Q = (k_{i,j})$ необхідно враховувати вплив кожного коефіцієнта на відновлення зображення. У даній роботі пропонується такий алгоритм для знаходження матриці квантизації. Нехай E - деякий рівень похибки, тобто $E = \|f(x, y) - \tilde{f}(x, y)\|$,

де $\tilde{f}(x, y)$ -- відновлене зображення. Для кожного коефіцієнта перетворення знаходимо таке максимальне значення k_j , що

$$\|f(x, y) - \tilde{f}(x, y)\| \leq E. \text{ Нехай } k_{\min} = \min_j k_j \text{ і}$$

$$q_j = \frac{k_j}{k_{\min}}. \text{ Тоді шукана матриця має такий}$$

вигляд: $Q = (q_j k_{\min})_j$. За допомогою коефіцієнта k_{\min} будемо регулювати коефіцієнт стискання - при більшому k_{\min} стискання буде більшим. Так як не вигідно знаходити матрицю квантизації для кожного зображення окремо в зв'язку із значним часом роботи алгоритму, то достатньо побудувати одну матрицю для деякого класу зображень. Такими класами можуть бути, наприклад, відбитки пальців, фотографії обличчя людини.

На заключному етапі алгоритму стискання квантизовані коефіцієнти кодуються за допомогою одного з алгоритмів ентропійного кодування. У даній роботі використовується динамічний алгоритм кодування Хаффмана. Цей алгоритм добре відомий у теорії кодування ([4, 5]).

3.1. Особливості розробленої програми. Розроблена програма дозволяє стискати зображення з градацією сірого. У якості базису сплесків використовується система функцій Добеші [2].

Таблиця квантизації, побудована для класу зображень відбитків пальців, має такий вид:

5k, 6k, 7k, 7k, 8k, 10k, 10k, 8k,
6k, 7k, 7k, 7k, 10k, 10k, 10k, 9k,
7k, 7k, 8k, 7k, 11k, 11k, 10k, 10k,
7k, 7k, 8k, 7k, 10k, 10k, 10k, 10k,
8k, 9k, 10k, 10k, 11k, 11k, 11k, 10k,
9k, 10k, 10k, 10k, 11k, 11k, 11k, 10k,
9k, 9k, 10k, 10k, 11k, 11k, 11k, 11k,

3.2. Отримані результати. Результати роботи програми будемо розглядати на прикладі фотографії відбитка пальця. У наступній таблиці наведені деякі значення "відстані" між оригінальним зображенням і зображенням, відновленим після стиснення. "Відстань" обчислюється по формулі середньоквадратичного значення різниць між інтенсивностями відповідних пікселів двох зображень:

$$D = \sqrt{\frac{\sum_{i=0}^{width-1} \sum_{j=0}^{height-1} (f(i,j) - \tilde{f}(i,j))^2}{width * height}}$$

де $f(i,j)$ і $\tilde{f}(i,j)$ -- функції інтенсивності пікселів відповідно оригінального і відновленого зображень, $width \times height$ -- розміри зображення (Таблиця).

Таблиця

Коефіцієнт стиснення	Відстань
3	2.93
5	4.15
10	6.84
15	8.10
20	9.51
25	11.53
30	12.59
35	13.52
40	14.19
45	15.15

Нижче приведено результати роботи програми на прикладі фотографії відбитку пальця

Оригінальне зображення: Коефіцієнт стиснення 10:



Коефіцієнт стиснення 20:

Коефіцієнт стиснення 40:



ВИСНОВКИ

Основним вдосконаленням аналізованого методу є, звичайно, використання кращих систем сплесків, тобто систем, що краще відбивають специфіку вхідної функції, в даному випадку зображення. На сьогодні найкращими базисами сплесків, що застосовуються для стиснення зображень, є сплески Добеші [2]. Кращі результати можна одержати, використовуючи для перетворення суперпозицію різних базисних функцій сплесків. Також важливим чинником, що дуже сильно впливає на коефіцієнт стиску, є правильний вибір таблиці квантизації. У даній роботі при знаходженні коефіцієнтів квантизації враховується вплив кожного коефіцієнта окремо, хоча в загальному випадку втрати при квантизації двох коефіцієнтів не будуть рівні сумі втрат для кожного коефіцієнта окремо. Найкращим методом знаходження цих коефіцієнтів буде використання класичних методів оптимізації, але час роботи таких алгоритмів може бути дуже великим.

Література

1. Mallat S. *Multiresolution approximation and wavelets*. // Trans. AMS – 1989 --315, -- P. 69-88.
2. Daubechies I., *Ten lectures on wavelets*. // SIAM Philadelphia, 1992. -- P.377
3. Новиков И.Я., Стечкин С.Б. *Основы теории всплесков* // Успехи мат. наук. – 1998 -- 53. № 6 (324) -- С. 53-128
4. Pennebaker W.B., Mitchell J.L., *JPEG Still Image Data Compression Standard*. Van Nostrand Reinhold., 1993, P. 656.
5. Bell T., Cleary J., Witten I., *Text compression*. Prentice Hall, 1990, P.337

COMMON APPROACH TO SOLUTION OF SOME PROBLEMS OF DISCRIMINANT AND CLUSTER ANALYSIS

O.V.Babak¹, A. S. Gasanov¹, T.A. Babayev², J. V. Backlan³

1. International Research and Training Centre of Information Technologies and Systems,
40, prospect Academica Glushkova, 03680, Kiev, Ukraine.

Ph.: (38044) 266-41-87 fax: (38044) 266-15-70,

E-mail: vig@tel.dlab.kiev.ua

2. Baku Scientific and Training Centre

11, I. Idayat-Zade, 370154, Baku, Azerbaijan,

Ph.: (99412) 66-38-81, fax: (99412) 66-39-95

E-mail: tbabayev@azeri.com

3. Education-Scientific complex "Institute for Applied System Analysis",

37, Peremoghy ave, 01056, Kiev, Ukraine.

Ph: (38044) 243-59-20, fax: (38044) 243-59-20

E-mail: kopernik2000@mail.ru

Abstract In this work a common approach is proposed, that simplifies solution of some problems of the discriminant and cluster analysis. The approach is based on restoration of an objectively existing relation, hidden in input data. The relation can be synthesized as a multiplicative function in accordance with obtained information on direction of gradient components of linear model of object classification. The revealed relation projects input information onto an axis and makes it possible to get an enough simple decision rule of dividing objects onto classes, or to establish, that in a given class of uniform objects there exist subclasses of objects, similar in a sense.

ВСТУП

Статистичний підхід щодо розпізнавання образів, методами якого є дискримінантний та кластерний аналіз, займає значне місце у різних сферах науки та техніки. Статистичні моделі алгоритмів розпізнавання засновані на використанні апарату математичної статистики. Вони застосовуються загалом у тих випадках, коли відомі або можуть бути просто визначені ймовірні характеристики класів, що, наприклад, відповідають функції розподілу.

З математичної точки зору засоби розв'язання задач у вказаних двох галузях розпізнавання образів мають істотні відмінності. Тому створення єдиного підходу до їх розв'язання навіть щодо окремих випадків, викликає певний практичний та теоретичний інтерес. Така можливість виникає, якщо

у вихідній інформації об'єктивно існують приховані залежності одного характеру, виявлення яких дозволяє здобути бажаний результат. Наслідком цього є істотне спрощення розв'язання задачі, що одержується за допомогою одного і того ж алгоритмічного та програмового забезпечення.

Методологія розпізнавання при цьому виглядає приблизно так. Допустимо, наприклад, що є дані, які отримано в результаті фізичного або імітаційного експерименту. Ці дані в деякому дуже обмеженому змісті, що характеризують досліджуваний об'єкт або явище, необхідно спробувати звести разом для того, щоб встановити, які закономірності відбуваються в наявному матеріалі. Для цього висувається деяка робоча гіпотеза, що породжується евристикою, якій придається математичний образ – потім покладений в основу моделі розпізнавання тієї чи іншої задачі дискримінантного або кластерного аналізу.

1. ЕЛЕМЕНТИ ТЕОРІЇ ТА ОСОБЛИВОСТІ АЛГОРИТМІВ

Перш ніж перейти до особливостей побудови єдиного підходу до розв'язування деяких завдань дискримінантного та кластерного аналізу відзначимо:

При дискримінантному аналізі, у разі розпізнавання, наприклад, двох класів об'єктів, за наявними даними або відновлюються, або виявляються дискримінантні залежності y^* від n – числа ознак (змінних) x . Доречно відзначити, що тут відновлення та виявлення – це принципово різні задачі. Якщо відновлення – це вибирання за заданим

набором ознак її структури з певного класу залежностей [1], то при виявленні використовується не весь набір ознак, а лише кращі з них або їх комбінації. Особливо важливим є виявлення комбінацій ознак, тому що вже давно спостережено, що часто “розрізювання” приховане не у самих ознаках, а в їх сполученнях [2]. Таким чином, здійснюється перехід до нового, лінійного за параметрами й нелінійного за змінними простору, вимірність якого завдяки редуціюванню вхідної інформації можна істотно знизити (аж до одновимірного).

Під кластеризацією об'єктів зазвичай вважається виявлення у певній матриці даних звичайного, об'єктивно існуючого порядку, завдяки чому стає можливим виділення кластерів – деяких підмножин множин, що досліджуються.

У праці [2] наведено основні різновиди кластерного аналізу, які зустрічаються найчастіше. Найбільш розповсюджений метод виявлення кластерів пов'язаний із завданням на багатьох спостереженнях належної метрики, наприклад, евклідової та вирахуванням за допомогою її відстаней між усіма точками вказаної множини.

Доречно відмітити, що особливістю процесу виділення кластерів є те, що він може розглядатися й як геометричне завдання про виявлення у певному просторі “щільних” накопичень точок. Вказана особливість обумовлює можливість розвинути зовсім специфічні методи її вирішення. Один з таких методів, що дозволяє не лише спростити рішення, але й візуально “побачити” підсумки кластеризації розглядається у поданій роботі. Ідея запропонованого методу базується на редуції інформації, в результаті чого вихідну, як правило, складну багатомірну задачу перетворюють у більш просту, наприклад, трьохмірну, двомірну або одномірну. При цьому існують два принципово різних напрямки скорочення розмірності простору наявних даних.

В основі першого напрямку лежить стиснення набору вихідних факторів (незалежних змінних) максимум до трьох за рахунок відсіювання решти, якщо вдається встановити, що останні практично не впливають на виділення кластерів. Вказана процедура носить назву цілеспрямованого проектування й, зазвичай, опирається на метод головних компонент [5], технічна реалізація якого, відрізняючись складністю, і є можливою лише тоді, коли фактори корельовані між собою. Другий напрямок, що розвивається нами, базується на пошукові деякої компактної функціональної залежності, що об'єднує усі значні і некорельовані (корельовані) між собою фактори, число яких може бути більше трьох, у вигляді так званої узагальненої змінної (поняття введено у працях О. Г. Івахненко [4]), завдяки чому стає можливим перетворення вихідної багатомірної задачі в зручну одномірну.

Звичайно, відображення багатомірного простору на одномірне теоретично не може зменшити мінімально досяжний рівень помилки виявлення кластерів. Однак, з успіхом можна вважати, що якщо дані, які породжуються певним джерелом, підпорядковуються нормальному закону розподілу, то ці жертви будуть мінімальними [7]. Відрізняючись, як буде вказано нижче, простотою технічної реалізації, другий напрямок є плідним для виділення кластерів, практично некорельованих між собою значущих факторів.

Оскільки при кластерному аналізі об'єкти поділяються на групи за їх подібністю у певному понятті, що можна інтерпретувати й так: дані, що належать до певного класу однорідних об'єктів поділяються на підкласи. Саме таке розуміння розбиття класу подібних об'єктів на підкласи разом з проектуванням вхідної інформації на одновимірний простір створює базу єдиного підходу до розв'язання деяких задач дискримінантного та кластерного аналізу.

Виникає питання: як знайти оптимальні у певному сенсі сполучення ознак, не вдаючись до комбінаторного перебору, що потребує великого обсягу обчислень, не гарантуючи вірності вибору того чи іншого критерію. Відповідь на це питання базується на таких положеннях:

1. Більшість явищ довкілля і відповідно їх моделей нелінійні. При цьому найважливіша інформація про характер структури такої моделі є в лінійній за параметрами й змінними моделі явища, що може бути представлена в виді (1) в якій є інформація про складові градієнта функції відгуку:

$$\hat{y} = \sum_{i=1}^n a_i x_i, \quad (1)$$

де \hat{y} - прогнозоване значення функції відгуку у, a_i - оцінки коефіцієнтів, x_i - число ознак (змінних).

2. Більшість моделей класичної фізики є мультиплікативні і можуть бути представлені у виді (2):

$$y = k \prod_{i=1}^n x_i^{p_i}, \quad (2)$$

де $k > 0$ – коефіцієнт (раціональне число), p_i – показник степені позитивне (негативне) раціональне число.

3. Розв'язання деяких задач дискримінантного та кластерного аналізу зводиться до виявлення прихованих у вихідних даних закономірностей, що об'єктивно існують, у вигляді мультиплікативних моделей типу (2).

В роботі [3] показано: якщо у лінійній моделі (1) коефіцієнти a є оцінки напрямку складових градієнту функції u , то можна синтезувати узагальнену змінну.

$$v = \prod_{i=1}^n x_i^{p_i}, \quad (3)$$

де p_i дорівнюють ± 1 , залежно від знаку відповідної компоненти градієнта.

Функції (1) і (3) у певному змісті подібні. Подібність їх є у тому, що із зменшенням (зростанням) будь-якої змінної відповідно змінюватимуться (зростатимуть або зменшуватимуться) значення u та v . Саме це, зберігаючи фізичні особливості явища при переході від його лінійної моделі до нелінійної, забезпечує вибір сполучень ознак, не удаючись до комбінаторного перебору. Цей результат закладено в основу єдиного підходу до розв'язання деяких задач дискримінантного та кластерного аналізу. Схема алгоритмів їх рішення включає такі етапи:

1. За даними навчальної вибірки, де для позначення належності об'єктів до двох класів або одного, використовується, відповідно, значення ± 1 або $+1$, відновляється лінійна функція (1) у результаті чого знаходяться оцінки коефіцієнтів a , що є оцінками напрямку складових її градієнту.

2. Знаючи напрямок складових градієнту функції відгуку, синтезується узагальнена змінна v (3).

3. Значення v_j , $j = \overline{1, l}$, де l – довжина навчальної вибірки, якщо це необхідно – масштабуються і потім розміщуються на одній координатній осі.

4. Характер розміщення v_j аналізується в залежності від роду задачі.

Примітка. За обчислення v можуть бути використані, як нормовані, так і ненормовані, значення незалежних змінних. При цій ситуації, значення, що включають їхнє нульове не розглядаються, оскільки в даному випадку класифікація стає неможливою.

Далі розглядатимуться особливості алгоритмів розв'язання таких задач.

2. ЗАДАЧА ДИСКРИМІНАНТНОГО АНАЛІЗУ

Нехай задано навчаючу вибірку довжиною l

$$\{x_{ij}, y_j^*\}, \quad i = \overline{1, n}, \quad j = \overline{1, l}, \quad (4)$$

причому $\forall x_{ij} \neq 0$, де y_j^* приймає значення ± 1 в залежності від належності класу A або B елементів вибірки за якою відбудовується лінійна дискримінантна функція u . При цьому стають відомими оцінки напрямку складових її градієнту і

можливий синтез v [3]. Оскільки значення v_j , $j = \overline{1, l}$, розташовані на одній координатній осі, нескладно встановити межі належності їх до класів A і B та одержати у даному простому випадку розв'язальне правило.

При побудові його необхідно врахувати важливе поняття дискримінантного аналізу – “відмова від розпізнавання” [2]. Сутність його полягає в тому, що в сумнівних випадках (існує область, що включає об'єкти як класу A , так і класу B), або при поганій згоді (існує область в якій об'єкти розташовані далеко від середніх значень класів) класифікація неможлива. Отже, при необхідності в таких випадках вона повинна здійснюватися іншими засобами, зокрема, наприклад, виходячи з евристичних розумінь. Природно, використовуючи відмову від класифікації, можна різко підвищити її якість при зменшенні числа правильних класифікаційних рішень. Таким чином, з урахуванням сумнівних випадків (область з межами v_a і v_b , $v_a < v_b$, де a, b деякі числа з вибірки $\overline{1, l}$) і при поганій згоді (область з межами $-\infty$, v_c і v_d , $+\infty$ де $v_c < v_a$ і $v_b < v_d$ і c, d теж деякі числа з вибірки $\overline{1, l}$) вирішальне правило матиме вигляд:

$$\begin{aligned} v_c < v < v_a, & \quad v \in \text{класу } A, \\ v_b < v < v_d, & \quad v \in \text{класу } B. \end{aligned}$$

При цьому оцінка вирішальних правил повинна бути виконана на основі оцінювання класифікаційної помилки. Найпростіша можливість полягає у застосуванні його до об'єктів навчання і підрахунку числа невірних рішень. Інша можливість більш суворої оцінки пов'язана з поділом множини об'єктів на навчальне і контрольне. При цьому навчальна множина використовується при побудові вирішального правила, а контрольна – для оцінки його якості. Найбільш сильний результат може бути отриманий із застосуванням методу контролю, що стежить і який передбачає вилучення з навчальної множини лише одного об'єкта, який потім використовується для оцінки якості вирішального правила [1].

3. ЗАДАЧА КЛАСТЕРНОГО АНАЛІЗУ

Нехай задано навчаючу вибірку

$$\{x_{ij}\}, \quad i = \overline{1, n}, \quad j = \overline{1, l},$$

при чому $\forall x_{ij} \neq 0$, що репрезентує клас однорідних об'єктів. Аналогічно до вибірки дискримінантного аналізу її можна записати як (4) з

тією різницею, що усі v прийматимуть значення $+1$ (тільки один клас). Далі, як і раніше, відбудовується лінійна функція (1) і синтезується v . У зв'язку з тим, що v_j розміщені на одній осі, легко побачити чи є на ній "густі" скупчення точок (підкласи подібних об'єктів або кластери). Необхідно відмітити, що проблема остаточного виявлення кластерів зовсім не проста, оскільки оцінка міри подібності вельми суб'єктивна.

Пояснюється це тим, що в кластерному аналізі не існує однозначного кількісного критерію подібного помилці класифікації в дискримінантному аналізі. Такий критерій не можна сформулювати, оскільки в різноманітних прикладних задачах різноманітними можуть бути і цілі аналізу, спрямовані на одержання нової інформації. У аналізованому випадку усе ж доцільно ввести ряд понять, що дозволяють якоюсь мірою порівняти і сприяти виділенню кластерів. До таких понять ставляться:

по-перше, довжина кластера L визначається наступним способом:

$$v_{j \min} = \min\{v_j; \overline{1, k}\},$$

$$v_{j \max} = \max\{v_j; \overline{1, k}\},$$

$$L = (v_{j \max} - v_{j \min}),$$

де k - число кластерів;

по-друге, щільність кластера:

$$d_j = \frac{m_j}{L_j}, \quad j \in \overline{1, k},$$

де m - число елементів j -го кластера.

З огляду на те, що кластер є утворенням, де внутрішні структурні зв'язки в декілька разів більші порівняно з зовнішніми зв'язками одного кластера з іншим, для виділення кластерів також корисно ввести поняття міри близькості елементів кожного кластера у вигляді

$$\forall (v_{i+1} - v_i) \leq \varepsilon,$$

де i - елемент кластера, ε - деяке чисельне значення, що може бути задане відповідно до характеру тієї чи іншої задачі кластеризації.

ВИСНОВКИ

У відповідності з викладеним можна зробити наступні висновки.

1. Запропоновано єдиний підхід до вирішення деяких задач дискримінантного та кластерного аналізу, особливостями яких є наявність у вхідних даних прихованих об'єктивно існуючих залежностей, які можна відобразити у вигляді мультиплікативних функцій [3].

2. В основі єдиного підходу до вирішення деяких задач дискримінантного і кластерного аналізу лежить у першому випадку розбиття багатьох різного роду об'єктів на класи, а у другому випадку розбиття багатьох подібних об'єктів (класу об'єктів) на підкласи. Ця обставина дозволяє розглядати вищевказані задачі з однієї і тієї ж математичної точки зору, що спирається на побудову лінійних моделей, які утримують інформацію про спрямування складових градієнта лінійних функцій.

3. Наявність інформації про спрямування складових градієнта лінійної моделі при рішенні деяких задач дискримінантного і кластерного аналізу дозволяє здійснити синтез узагальненої змінної у вигляді мультиплікаційної функції, зберігаючи фізичні особливості явища.

4. Значення мультиплікаційної функції у задачах дискримінантного і кластерного аналізу можуть розташовуватися на одній координатній осі, що суттєво спрощує процедуру побудови алгоритмів класифікації та кластеризації і робить її доступною для широкого кола користувачів.

ЛІТЕРАТУРА

1. Алгоритмы и программы восстановления зависимостей (под ред. В. Н. Вапника). - М.: Наука. - 1984. - 816 с.
2. Распознавание образов: состояние и перспективы. Пер. с англ. (К. Верхаген, Р. Дейн и др.). - М.: Радио и связь. - 1985. - 104 с.
3. Бабак О. В. Об одном подходе к решению задачи восстановления зависимости в классе кусочно-линейных функций // Проблемы управления и информатики. - 1995. - №6. - С. 134-141.
4. Ивахненко А. Г., Мюллер Й. А. Самоорганизация прогнозирующих моделей. - К.: Техника, 1985; Берлин: ФЭБ Ферлаг Техник, 1984. - 233 с.
5. Прикладная статистика: Классификация и снижение размерности: Справ. Изд./С.А. Айвазян и др., под ред. С. А. Айвазяна. - М.: Финансы и статистика, - 1989. - 426 с.
6. Распознавание, классификация, прогноз/ Отв. ред. Ю. И. Журавлев. - М.: Наука, 1989. - 280 с.
7. Дуда Р., Харт П. Распознавание образов и сцен. - М.: Мир, 1976. - 511 с.8