

Przemysław SKUROWSKI
Politechnika Śląska, Instytut Informatyki

ANALIZA MOŻLIWOŚCI WYKORZYSTANIA LINIOWYCH PRZEKSZTAŁCENÍ GEOMETRYCZNYCH W KOMPRESJI SEKWENCJI WIZYJNYCH

Streszczenie. Standardowe metody kompresji sekwencji wizyjnych bazują m.in. na predykcji obrazu na podstawie wektorów przesunięć. Istnieje możliwość rozszerzenia metody i stworzenia algorytmu rekonstruującego obraz, używającego wektorów przesunięć, kąta obrotu i współczynnika przeskalowania. Do ich poszukiwania i pasowania bloków zaproponowano użycie transformaty Mellina i Fouriera-Mellina. Funkcjonowanie i zbieżność rozwiązania zweryfikowano względem poszukiwania podobieństwa obrazów na zestawie obrazów testowych.

Słowa kluczowe: kompresja wideo, pasowanie bloków, estymacja ruchu

APPLICABILITY ANALYSIS OF GEOMETRIC LINEAR TRANSFORMATIONS FOR VIDEO SEQUENCES COMPRESSION

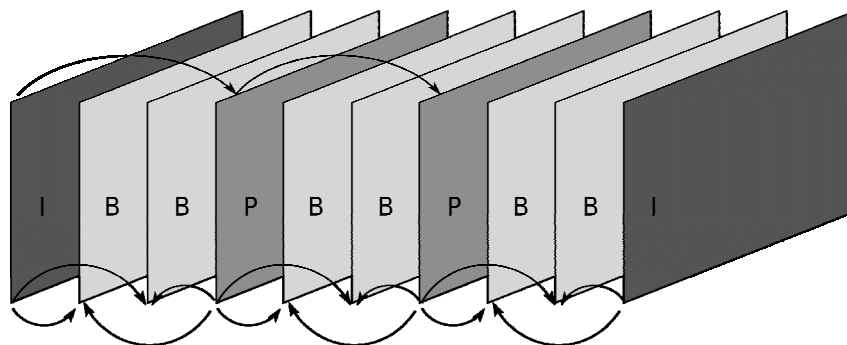
Summary. Standard video compression algorithms are based on motion vector prediction of images. There is potential possibility for enhancement of such an approach and creating an algorithm that predict images using motion vector rotation angle and scale factor as well. As a tool for searching these parameters and block matching there were Mellin and Fourier-Mellin transform proposed. The usability of proposed solution and the convergence of searching was verified against test images.

Keywords: video compression, block matching, motion estimation

1. Kompresja sekwencji wizyjnych

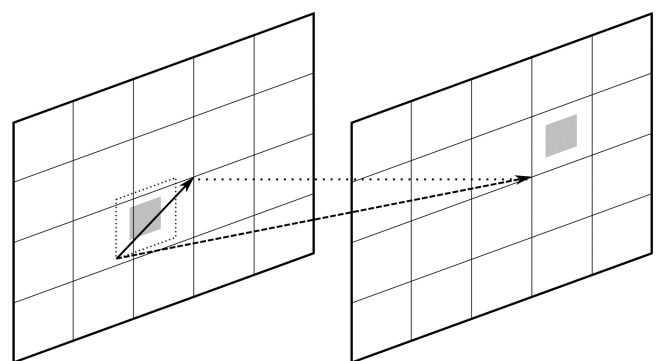
Większość współczesnych koderów sekwencji wizyjnych realizuje pewien schemat. W sekwencji w pewnym ustalonym porządku [2,4] (np. tzw. GOP –ang. *Group of Pictures*) występują obrazy dwóch rodzajów – kodowane wewnątrz- i międzyobrazowo. Obrazy kodo-

wane wewnątrzobrazowo (**I** ang. *intrapicture*) są bytami samodzielnymi i niosą całość informacji potrzebną do rekonstrukcji obrazu oryginalnego z zadaną dokładnością – wykorzystuje się tutaj najczęściej kompresję stratną przy użyciu transformacji sumacyjnych jak DCT, których widmo jest następnie kwantowane. Obrazy te stanowią swego rodzaju punkty odniesienia i powinny się powtarzać w sekwencji regularnie jako że ich jakość jest najwyższa i nie kumulują się w nich błędy predykcji – odbywa się to kosztem objętości, obrazy te w sekwencjach wizyjnych mają największy rozmiar.



Rys. 1. Przykładowy schemat GOP
Fig. 1. Exemplary GOP scheme

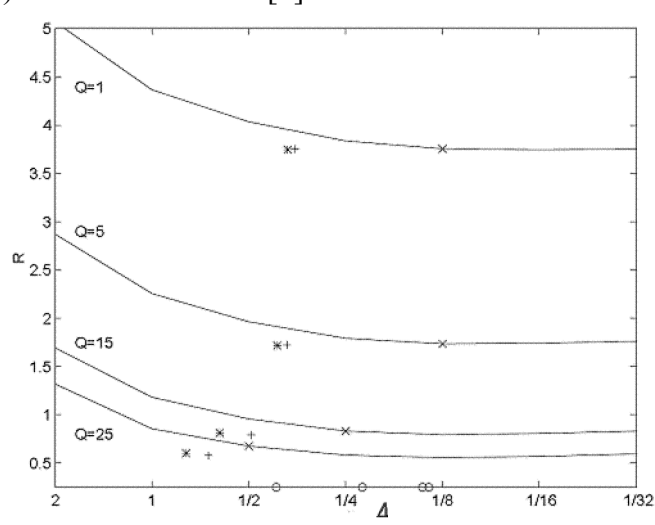
Obrazy kodowane międzyobrazowo (ang. *interpicture*) wymagają do rekonstrukcji zarówno informacji zawartych w nich samych, jak i treści innych obrazów – zwanych obrazami odniesienia – na podstawie których obraz jest odtwarzany. Na rys. 1 przedstawiono przykładowy schemat GOP, który obowiązuje w kompresji MPEG1/2 wraz z powiązaniem obrazów wynikowych i ich obrazów odniesienia. Predykcja ta może się odbywać zarówno jednokierunkowo „wstecz” (obrazy **P**), jak i dwukierunkowo (obrazy **B**), gdy obraz rekonstruowany jest na podstawie obrazów przeszłych i przyszłych.



obraz odniesienia obraz rekonstruowany
Rys. 2. Predykcja obrazu wektorem przesunięć
Fig. 2. Motion vector image prediction

Najpowszechniej spotykaną techniką rekonstrukcji jest tzw. predykcja przy użyciu wektorów przesunięć (MV ang. *motion vectors*), gdzie fragment obrazu (makroblok o rozmiarach

16×16) odniesienia wskazany wektorem przesunięcia jest umiejscawiany w rekonstruowanym obrazie. Rekonstrukcja taka pociąga za sobą sporą dozę niedokładności, stąd pojawia się konieczność zachowania obrazu różnic (tzw. obraz residualny), czyli różnicy pomiędzy uzyskanym obrazem przy użyciu wektorów przesunięć a docelowym obrazem. Obrazy różnic najczęściej kodowane są wewnątrzobrazowo, co powoduje znaczne zwiększanie rozmiaru skompresowanego obrazu, zatem im precyzyjniejsza będzie predykcja przy użyciu wektora przesunięć, tym efektywniej będzie można skompresować obraz [7] (rys. 2). W celu poprawy jakości takiej predykcji w kolejnych standardach zwiększano dokładność wektorów przesunięć od 1-pikselowej (ang. *integer pel accuracy*) w kodowaniu h.261 do ¼ piksela (ang. *quarter pel accuracy*) w kodeku h.263v2 [9].



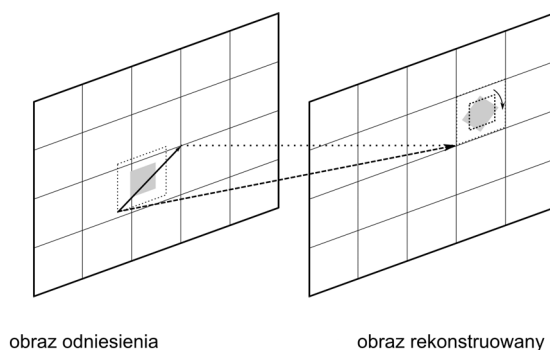
Rys. 3. Bitrate (R) w zależności od dokładności predykcji (Δ) i kroku kwantyzacji (Q) obrazu różnic
Fig. 3. Bitrate (R) in relation to prediction precision (Δ) and quantization step (Q) of difference image

2. Dyskusja problemu

Opisany w p. 1 sposób rekonstrukcji sprawdza się w szczególności, gdy zmiany w sekwencji mają charakter linearny, jednakże niejednokrotnie w sekwencjach występują nie tylko przesunięcia, ale również inne zmiany, jak obrót, przeskalowanie, zmiana perspektywy etc. Można zatem domniemywać, że dokładność predykcji może znacząco wzrosnąć, jeżeli uwzględnić w niej inne przekształcenia geometryczne opisywane przez przekształcenia afiniczne dane wzorem:

$$\begin{bmatrix} x' \\ y' \\ c \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}, \quad (1)$$

gdzie: x, y – współrzędne punktu, x', y' – nowe współrzędne punktu, a_{ij} – współczynniki przekształcenia, c – zmienna pomijana.



Rys. 4. Predykcja obrazu wektorem przesunięć i obrotem
Fig. 4. Image prediction using motion vector and rotation

Teoretycznie można by dla każdego makrobloku dobrać macierz przekształcenia, np. przez poszukiwanie jej współczynników którymś z ogólnych algorytmów minimalizujących funkcję błędu predykcji, niemniej takie podejście wydaje się być całkowicie nieracjonalne z punktu widzenia złożoności obliczeniowej problemu. Złożoność obliczeniowa algorytmów optymalizacyjnych jest duża, gdyby przeprowadzać poszukiwanie minimum błędu predykcji dla obrazu telewizyjnego SDTV przy 24 klatkach na sekundę o rozdzielczości 720×576 (2070 makrobloków), to otrzymalibyśmy 49680 zadań optymalizacyjnych do przeprowadzenia dla skompresowania sekundy filmu. Kompresja wideo w jakości HDTV byłaby odpowiednio bardziej złożona, co praktycznie dyskwalifikuje takie podejście.

W związku z powyższym, nasuwa się pytanie o to, jak można poszukać w rozsądnym czasie przynajmniej części parametrów, tak aby poprawić jakość predykcji. Narzędziem, które pozwala odpowiedzieć na to pytanie, wydaje się być transformata Mellina, ewentualnie transformata Fouriera-Mellina wykorzystywana w rozpoznawaniu obrazów w celu uzyskania niezmiennej względem przesunięcia, obrotu i przeskalowania reprezentacji obrazu.

3. Transformata Mellina i Fouriera-Mellina

Transformata Mellina [10] dana jest wzorem:

$$M\{f(t)\} = F_M(s) = \int_0^{\infty} t^{s-1} f(t) dt, \quad (2)$$

gdzie: s jest zmienną zespoloną.

Jak wykazali Robbins i Huang [8], można obliczyć ją przy użyciu transformaty Fouriera. Rozpatrując transformatę na osi urojonej $s = -j\omega$, gdzie $\omega = 2\pi f$ otrzymujemy:

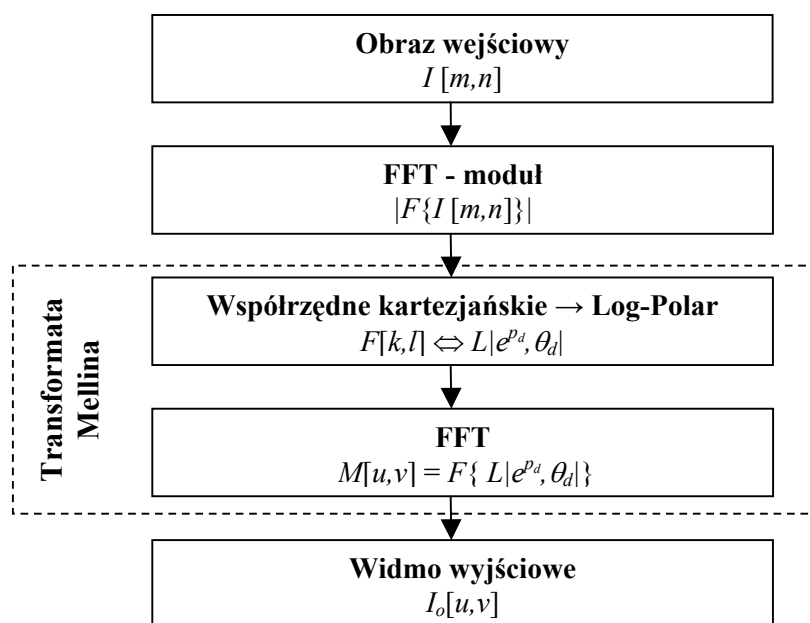
$$F_M(\omega) = \int_0^{\infty} t^{-j\omega-1} f(t) dt, \quad (3)$$

gdzie: s jest zmienną zespoloną. Postać fourierowska transformaty uzyskiwana jest następująco: podstawiając $\tau = \ln t$ lub $t = e^\tau$, co za tym idzie $dt = e^\tau d\tau$, otrzymujemy (uwaga na zmianę granic całkowania $\ln 0 = -\infty$):

$$F_M(\omega) = \int_{-\infty}^{\infty} f(e^\tau)(e^\tau)^{-j\omega-1} e^\tau d\tau = \int_{-\infty}^{\infty} f(e^\tau) e^{-j\omega\tau} d\tau, \quad (4)$$

czyli transformatę Fouriera, przeprowadzaną na ‘zmaszczonym’ (ang. *warped*) parametrze $\tau = \ln t$.

Analogicznie przypadek ten rozszerzalny jest na dwa wymiary. Jedną z konsekwencji użycia logarytmu na parametrze jest zamiana iloczynu na sumę, czyli skali na przesunięcie w fazie. Pozwala to na poszukiwanie i porównywanie cech niezmienniczych względem przeskalowania.



Rys. 5. Schemat obliczania transformaty Fouriera-Mellina
Fig. 5. Fourier-Mellin transform computation scheme

W kwestii nazewniczej dotyczącej transformaty Fouriera-Mellina (F-M) panuje pewna dowolność. Różni autorzy interpretują pojęcie transformaty F-M w dwojaki sposób. Jako samodzielną [1] transformatę Mellina (transformata Fouriera logarytmicznego promienia r) wraz z transformatą Fouriera po kącie (θ):

$$M\{f(x,y)\}(k,v) = \int_0^{\infty} \int_0^{2\pi} f(r,\theta) r^{1-iv} \underbrace{e^{-ik\theta} d\theta}_{\text{cz. Mellina}} \underbrace{dr}_{\text{cz. fourierowska}}. \quad (5)$$

W ujęciu prezentowanym w pracach innych autorów (np. [5]), dotyczących zastosowania transformaty Fouriera-Mellina, jako narzędzia do rozpoznawania obrazów niewrażliwego względem obrotu, skalowania i przesunięcia, jest ona „użytkową” kombinacją klasycznej transformaty Fouriera, zapewniającej niewrażliwość na przesunięcie, wraz z dwuwymiarową transformatą Mellina (ew. Fouriera-Mellina) obliczaną na podstawie (5) i najczęściej implementowaną przez transformatę Fouriera na obrazie i we współrzędnych biegunowo logarytmicznych (ang. *log-polar*). Schemat obliczania transformaty F-M w tym rozumieniu przedstawiono na rysunku 5. Taka procedura postępowania pozwala na uzyskanie w pełni niezmienniej reprezentacji obrazu względem przesunięcia, skalowania i obrotu. Kolejno:

- a) transformata Fouriera zamienia przesunięcie w zmianę fazy: $F\{f(t - D)\} = e^{-j\omega D} F\{f(t)\}$,
- b) transformacja log-polar z własności logarytmu zamienia przeskalowanie (iloczyn) na przesunięcie (sumę) oraz obrót na przesunięcie,
- c) kolejna transformata Fouriera lokuje te zmiany (punkt b) jako przesunięcia w fazie.

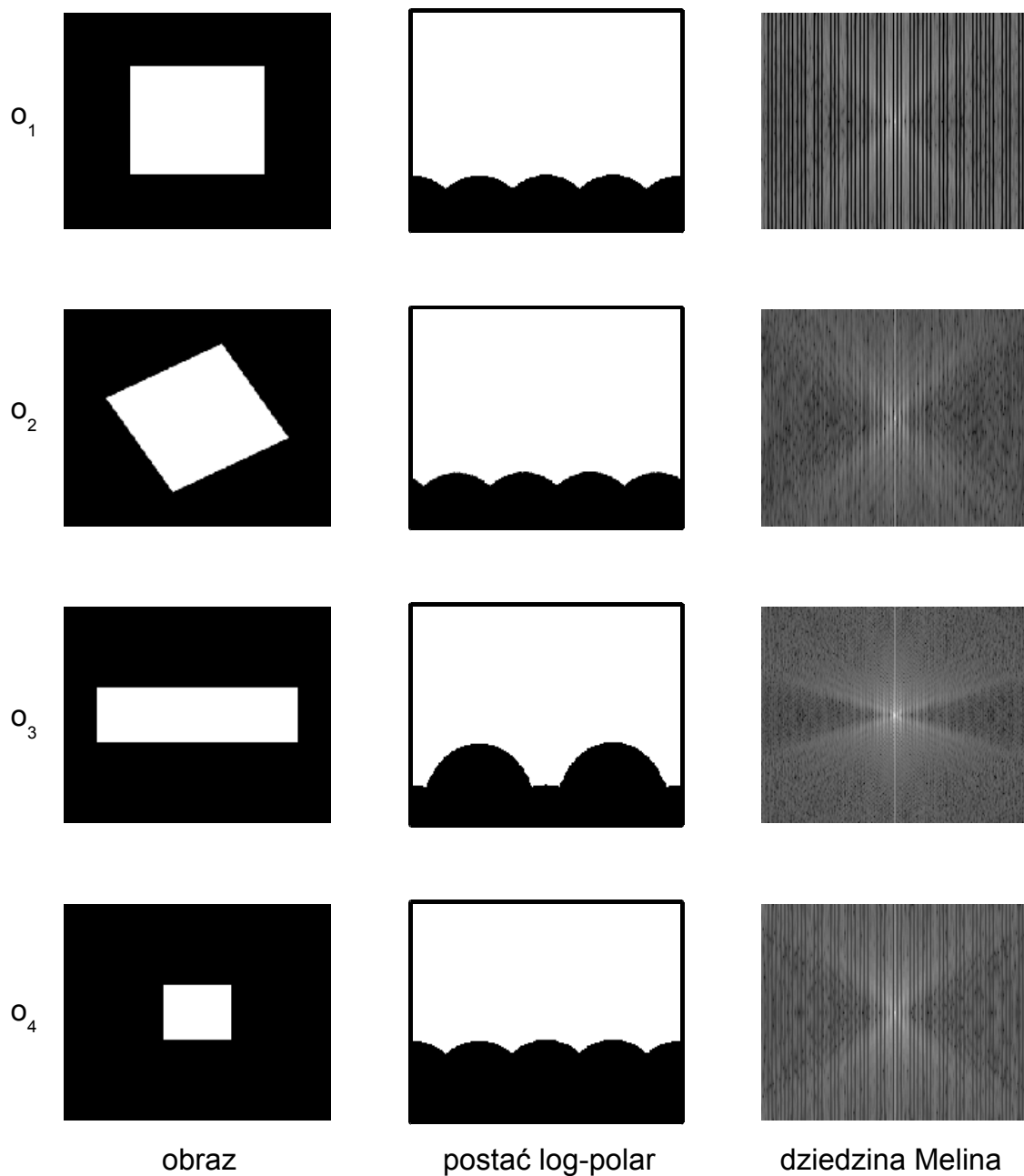
Autor przychyliła się również do tego drugiego podejścia, jako bardziej czytelniejszego. W związku z tym w dalszych punktach będzie stosowana ta konwencja.

4. Badanie możliwości wykorzystania transformaty Mellina

Dla czterech obrazów testowych w skali szarości (znormalizowanych w zakresie [0,1] o rozmiarach 256×256, przedstawionych na rys. 6) przeprowadzono transformatę Mellina. Kolejno powtórzono procedurę dla powyższych obrazów obciążonych każdy z osobną indywidualną realizacją szumu białego o amplitudzie równej 1/3 zakresu wartości obrazu (rys. 7). Następnie zaś oszacowano różnicę pomiędzy obrazami przy użyciu miary MAE (ang. *mean absolute error*) zwanej też MAD (ang. *mean absolute difference*) [2] stosowanej powszechnie w estymacji wektorów przesunięć, a opisanej zależnością:

$$\text{MAE}(p, q) = \frac{1}{MN} \sum_{m=1}^M \sum_{n=1}^N \left| |M\{p\}(m, n)| - |M\{q\}(m, n)| \right|, \quad (6)$$

gdzie: p, q – porównywane obrazy, $M\{\cdot\}$ transformata Mellina, m, n – współrzędne poszczególnych składowych.



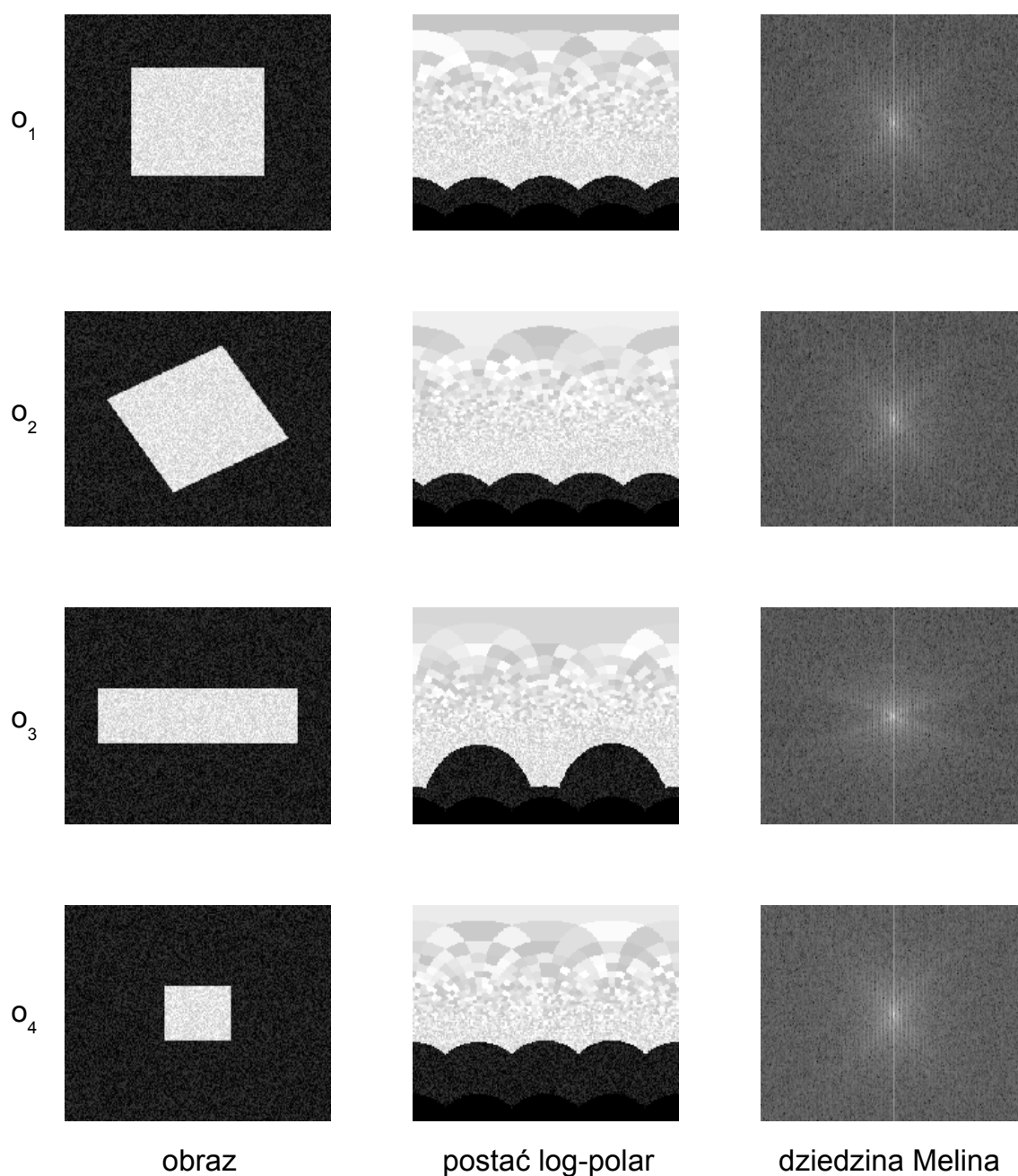
Rys. 6. Transformata Mellina dla czterech obrazów testowych
 Fig. 6. Mellin transform for four test images

Uzyskane wyniki (tab. 1) wskazują, że transformata Mellina „rozpoznaje” kształt obiektu, a zatem jest w stanie ocenić, czy blok porównywany odpowiada blokowi odniesienia.

Tabela 1

Różnice pomiędzy obrazami testowymi w dziedzinie Mellina

	obrazy niezasumiane	obrazy zasumiane
$MAE(o_1, o_2)$	3.54	10.08
$MAE(o_1, o_3)$	10.54	13.02
$MAE(o_1, o_4)$	4.98	11.07
$MAE(o_2, o_4)$	5.22	10.92



Rys. 7. Transformata Mellina dla czterech zaszumianych obrazów testowych
 Fig. 7. Mellin transform for four noisy test images

5. Problem zbieżności do minimum błęd MAE w dziedzinie Mellina

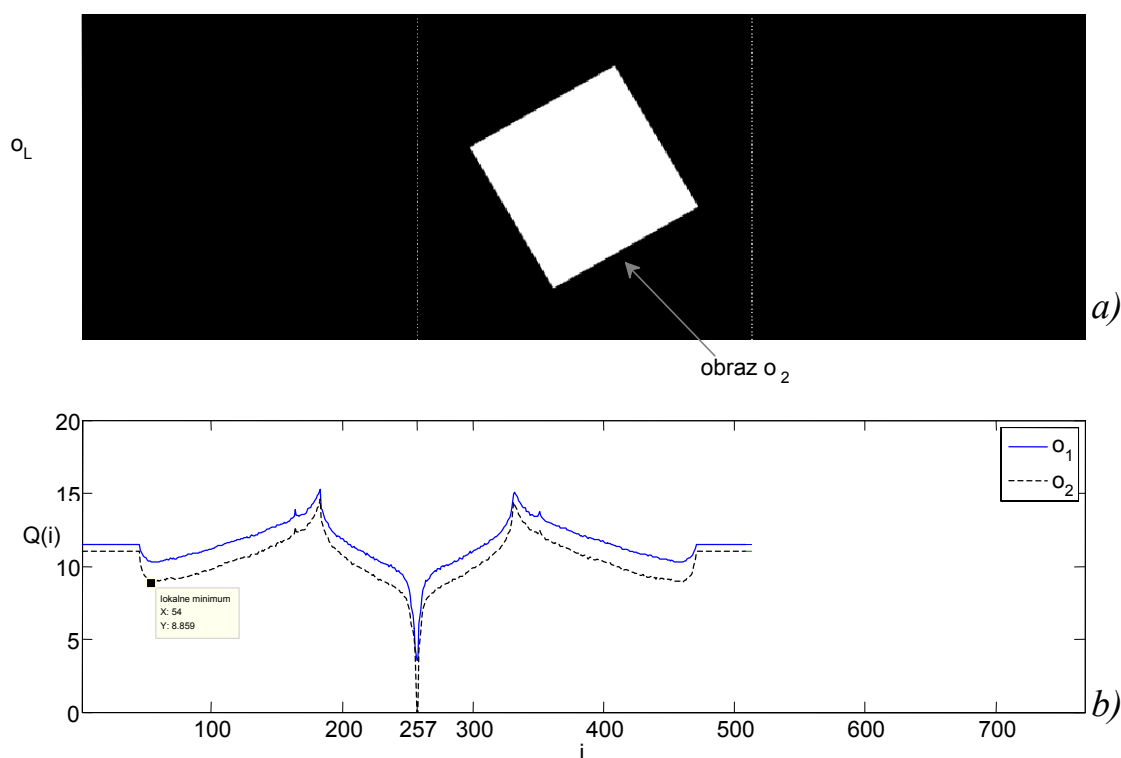
Następnym zagadnieniem do rozważenia jest pytanie, czy w trakcie poszukiwań wzorca w miarę zbliżania się do niego błąd MAE będzie jednostajnie malał, czy pojawią się lokalne minima, w których algorytmy poszukujące mogłyby się zatrzymać – może to być o tyle nieefektywne, że najczęściej są to proste algorytmy ukierunkowane na wydajność (algorytmy

trzech i czterech kroków, poszukiwanie logarytmiczne), a nie na dokładność znajdowania minimum globalnego, zatem „pułapka” w postaci minimum lokalnego może okazać się dla nich problemem. Zadanie to można opisać jako problem optymalizacyjny:

$$Q(i, j) = \text{MAE}(o(m, n), o_{ref}(m - i, n - j)) \rightarrow \min, \quad -p \leq i, j \leq p, \quad (7)$$

gdzie: $o(\cdot)$ – obraz, dla którego poszukujemy bloków w obrazie odniesienia, $o_{ref}(\cdot)$ – obraz odniesienia, p – zakres poszukiwań, i, j – przesunięcie.

W celu zbadania tego zagadnienia przygotowano obraz testowy o_L (przedstawiony na rys. 8a) o rozmiarach 768×256 , gdzie w pusty obraz wklejono obraz o_2 , który był iteracyjnie przeglądany (w jednym wymiarze i , przy ustalonym $j=0$) poszukiwaniem zupełnym ES (ang. *exhaustive search*) w nim wzorców o_1 i o_2 . Błąd MAE był wyliczany w dziedzinie Mellina i rejestrowany jako $Q(i)$, a został przedstawiony na rys. 8b.



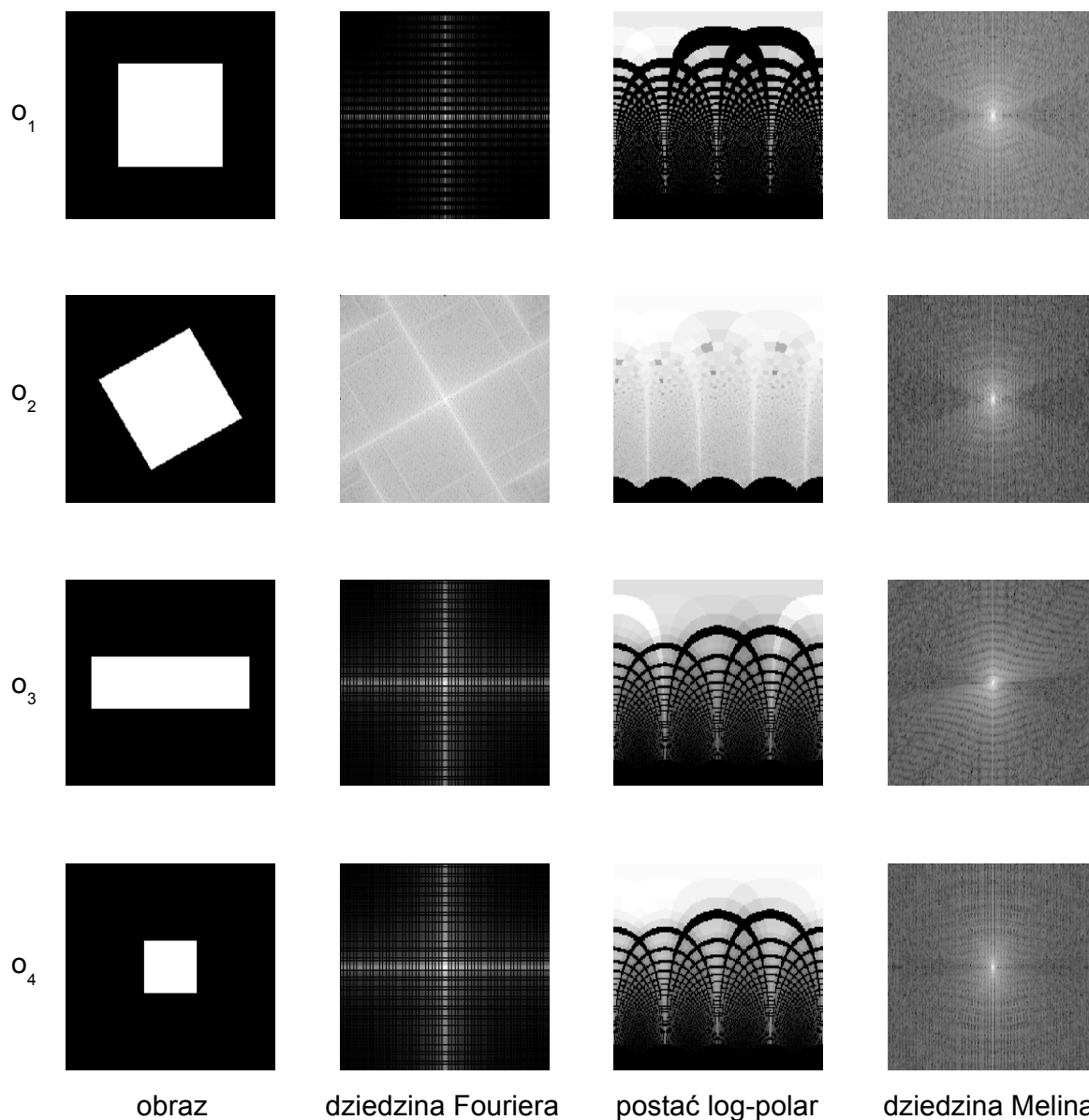
Rys. 8. Badanie zbieżności błędu MAE do obrazu oryginalnego przy przesunięciu i (perfekcyjne dopasowanie przy $i=257$)

Fig. 8. Testing of MAE error convergence to original image with shift i (perfect matching at $i=257$)

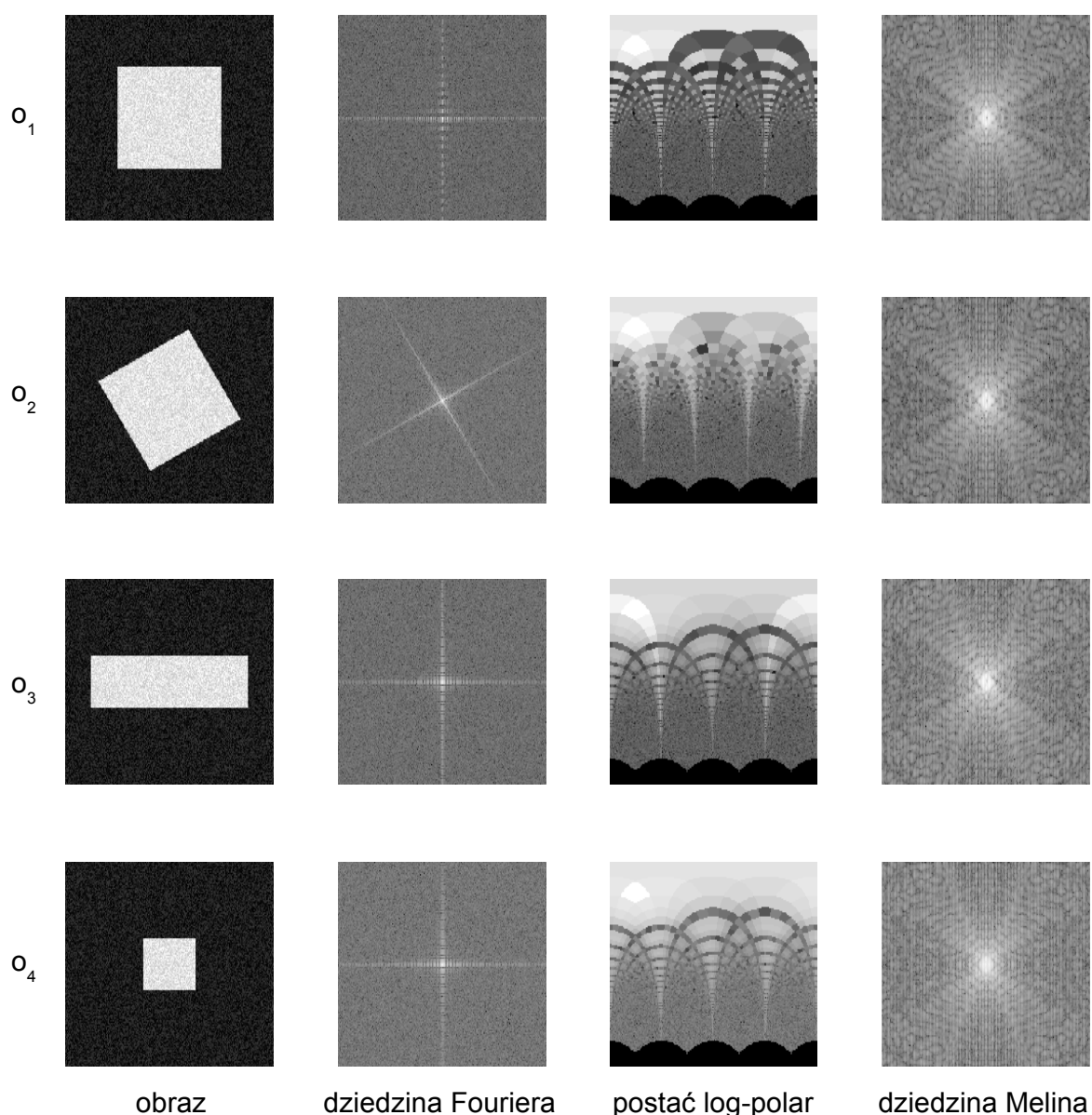
Można zaobserwować, że pojawia się minimum lokalne, a co za tym idzie występuje wzrost błędu w miarę zbliżania się do najlepszego dopasowania. Może to nastroczać pewnych trudności, np. algorytmom poszukującym wykorzystującym krok logarytmiczny.

6. Badanie możliwości wykorzystania transformaty Fouriera-Mellina

Analogicznie do rozważań w podrozdziale 4 przeanalizowano wykorzystanie transformaty Fouriera-Mellina. Dla tego samego zestawu czterech obrazów testowych (czystych i zaszumianych) zbadano różnicę mierzoną miarą MAE w dziedzinie transformaty F-M. Przeprowadzone obliczenia zilustrowano na rys. 9 i rys. 10.



Rys. 9. Transformata Mellina dla czterech obrazów testowych
 Fig. 9. Mellin transform for four test images



Rys. 10. Transformata Fouriera-Mellina dla czterech obrazów testowych
 Fig. 10. Fourier-Mellin transform for four test images

Uzyskane wyniki liczbowe zebrane są w tabeli 2.

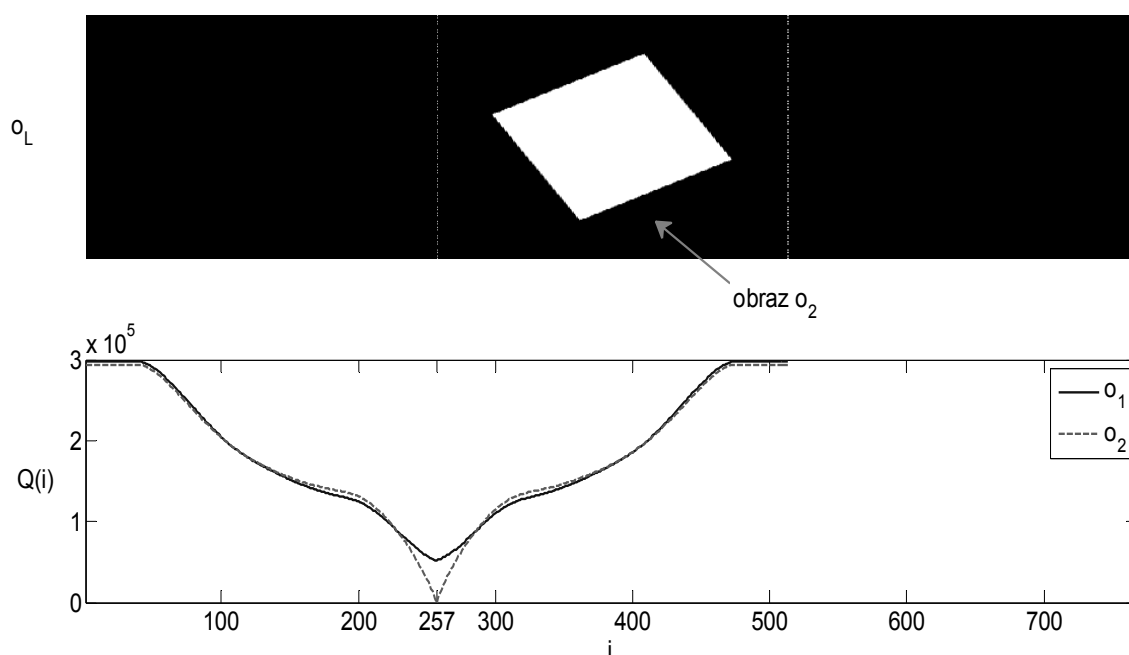
Tabela 2

Różnice pomiędzy obrazami testowymi w dziedzinie F-M

	obrazy niezaszumiane	obrazy zaszumiane
$MAE(o_1, o_2)$	$3.88 \cdot 10^4$	$5.37 \cdot 10^4$
$MAE(o_1, o_3)$	$6.34 \cdot 10^4$	$1.09 \cdot 10^5$
$MAE(o_1, o_4)$	$1.11 \cdot 10^5$	$1.92 \cdot 10^5$
$MAE(o_2, o_4)$	$9.39 \cdot 10^4$	$1.88 \cdot 10^5$

7. Problem zbieżności do minimum błędu MAE w dziedzinie F-M

Finalnie konieczna była weryfikacja zbieżności błędu MAE w dziedzinie F-M. W celu zbadania tego zagadnienia powtórzono eksperyment z p. 5. Przygotowano obraz testowy o_L (przedstawiony na rys. 11a) o rozmiarach 768×256 , gdzie w pusty obraz wklejono obraz o_2 , który był jak uprzednio iteracyjnie przemieszczany (w jednym wymiarze i , przy ustalonym $j=0$) poszukiwaniem zupełnym ES (ang. *exhaustive search*) w celu znalezienia w nim wzorców o_1 i o_2 , (jak na rys. 9). Błąd dopasowania MAE był wyliczany w dziedzinie Fouriera-Mellina i rejestrowany jako $Q(i)$ (rys. 11b).



Rys. 11. Badanie zbieżności błędu MAE do obrazu oryginalnego przy przesunięciu i (perfekcyjne dopasowanie przy $i=257$)

Fig. 11. Testing of MAE error convergence to original image with shift i (perfect matching at $i=257$)

Wyraźnie zanikły lokalne minima, a zatem pomiar błędu w tej dziedzinie jest potencjalnie lepszą metodą predykcji wektorów przesunięć.

8. Wnioski

W artykule wskazano potencjalną możliwość wykorzystania liniowych przekształceń geometrycznych, takich jak obrót czy przeskalowanie, do predykcji ruchu w sekwencjach wideo kodowanych hybrydowo. Wybrano narzędzia matematyczne, umożliwiające realizację operacji poszukiwania pasujących bloków i przetestowano je pod kątem użycia w algorytmach poszukiwania wektorów przesunięć. Może się wydawać, że z uwagi na operacje

matematyczne wykonywane w każdym kroku złożoność czasowa może znacznie wzrosnąć, i co za tym idzie metody te będą trudne w zastosowaniu praktycznym. Jednak jeżeli nie traktować ich jako zmodyfikowanego pomiaru dystansu, ale użyć wprost do oceny przesunięcia, przeskalowania i obrotu w dziedzinie widmowej – tak jak zaproponowano to w [3], gdzie poszukiwanie przesunięcia odbywa się wprost w dziedzinie fourierowskiej (wzajemne widmo gęstości mocy [6]), wówczas okazuje się, że można rozważać transformaty sumacyjne jako szybkie narzędzie estymacji parametrów ruchu.

BIBLIOGRAFIA

1. Derrode S., Ghorbel F.: Robust and efficient Fourier-Mellin transform approximations for invariant grey-level image description and reconstruction. *Computer Vision and Image Understanding*, Vol. 83(1), 2001, s. 57÷78.
2. Domański M.: Zaawansowane techniki kompresji obrazów i sekwencji wizyjnych. Wyd. Politechniki Poznańskiej, Poznań 2000.
3. Essannouni F, Haj Thami R. O., Salam A., Aboutajdine D.: An efficient fast full search block matching algorithm using FFT algorithms., *Int. J. on Computer Science and Network Security (IJSCSNS)*, Vol. 26(3), 2006, s. 130÷133.
4. Girod B., Steinach E., Farber N.: Comparison of the H.263 and H.261 Video Compression Standards SPIE Proceedings Vol. CR60, Standards and Common Interfaces for Video Information Systems, Philadelphia, USA, 1995.
5. Ho H. T., Goecke R.: Optical Flow Estimation using Fourier Mellin Transform, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR*, Anchorage 2008, s. 1÷8.
6. Kuglin C., Hines D.: The phase correlation image alignment method. *IEEE International Conference on Systems, Man and Cybernetics*, September , New York 1975, s. 163÷165.
7. Ribas-corbera J., Neuhoff D. L.: On the optimal motion vector accuracy for block-based motion-compensated video coders. *Proc. SPIE Dig. Video Compr.*, 1996, s. 302÷314.
8. Robbins G. M., Huang T. S.: Inverse filtering for linear shift-variant imaging systems. *Proceedings of the IEEE*, Vol. 60(7), 1972, s. 862÷872.
9. Sullivan G. J.: Overview of International Video Coding Standards (preceding H.264-/AVC), prezentacja raportu ITU, Geneva 2005.
10. Szabatin J.: Podstawy teorii sygnałów (wyd. 4). WKŁ, Warszawa 2003.

Recenzent: Dr hab. inż. Bogdan Smółka, prof. Pol. Śląskiej

Wpłynęło do Redakcji 5 października 2009 r.

Abstract

Classical hybrid video compression methods use two schemes for frame compression. Intraframe approach in which single image is standalone entity and interframe where video frames are interpolated on basis of ‘nearby’ frames by means of motion prediction. The more precise prediction the better compression is obtained. Nowadays high precision of prediction is achieved by subpixel motion vectors estimation accuracy of block matching algorithms.

The paper presents the idea to introduce into video compression more sophisticated methods of motion prediction than simple motion vectors used in current standards like MPEG coders family. The proposal includes searching for both classical motion vectors and scaling and rotation of content as well. In the paper there are analyzed two mathematical tools to estimate these parameters: Mellin and Fourier-Mellin transforms. Both tools were discussed, implemented and tested using a few simple test images to verify their properties and to check their usefulness for block matching in pure form and in presence of noise. For block matching, Fourier-Mellin transform appeared to outperform Mellin transform due to better convergence to global minimal error without risk of local minima where search algorithms could ‘stuck’. In conclusions author mentions a possibility for fast block searching method directly in Fourier-Mellin domain using cross-spectral density.

Adres

Przemysław SKUROWSKI: Politechnika Śląska, Instytut Informatyki, ul. Akademicka 16,
44-100 Gliwice, Polska, przemyslaw.skurowski@polsl.pl.