



Jeśli chcemy zapisać na komputerze obraz wideo, powinniśmy go wcześniej skompresować. Jednak tylko inteligentne metody kompresji są w stanie uporać się z tak ogromnymi zasobami danych.

Film w pigułce

Od roku 1895, w którym bracia Lumière dokonali pierwszej projekcji swojego filmu, ruchomy obraz nic nie stracił na atrakcyjności. Jednak dopiero w ostatnich latach mogły spełnić się marzenia o stosunkowo niedrogich technikach digitalizacji obrazu wideo przeznaczonego do dalszej obróbki na komputerze. Również gotowe sekwencje filmowe, dostępne na CD-ROM-ach (Video-CD, CD-I), można obecnie bez problemu wyświetlać na domowym pececie. Takie nowe możliwości zawdzięczamy popularyzacji metod kompresji obrazu wykorzystujących standard MPEG.

W wyniku digitalizacji sekwencji wideo zapisanej w systemie PAL, z jednej sekundy obrazu powstaje około 23 MB danych, co w ciągu godziny daje niewyobrażalną wielkość 80 GB. Aby więc zdigitalizowany obraz wideo nie pozostał wyłącznie domeną superkomputerów, tak duże zbiory danych muszą zostać poddane znacznej kompresji.

Nieodwracalne (stratne) techniki kompresji (np. MPEG) są w tym przypadku znacznie bardziej efektywne niż metody bezstratne (np. GIF), jednak z oczywistych względów mają ograniczone zastosowania (kompresja zdjęć, kompresja danych wideo i audio). Wykorzystują bowiem fakt, że ludzkie organy zmysłu przekazują do mózgu tylko część odebranych informacji. Cała reszta danych wizualnych i dźwiękowych jest więc zupełnie zbędna i może zostać „odfiltrowana” przez odpowiedni algorytm kompresji. Działając zgodnie z opisaną wyżej zasadą, podczas oglądania obrazu, oko ludzkie słabiej reaguje na zmiany barw niż na zmianę

jasności. Wykorzystując tę prawidłowość można zatem – bez zauważalnych strat – zredukować ilość informacji dotyczących barw, o ile utrzymana zostanie struktura jasności obrazu.

W przypadku kompresji muzyki korzysta się natomiast z fizjologicznej właściwości ucha ludzkiego, które nie rejestruje cichych dźwięków zawartych w głośniejszych sekwencjach. Taka sama sytuacja dotyczy dwóch dźwięków o niemal jednakowych częstotliwościach: człowiek słyszy jedynie głośniejszy z nich.

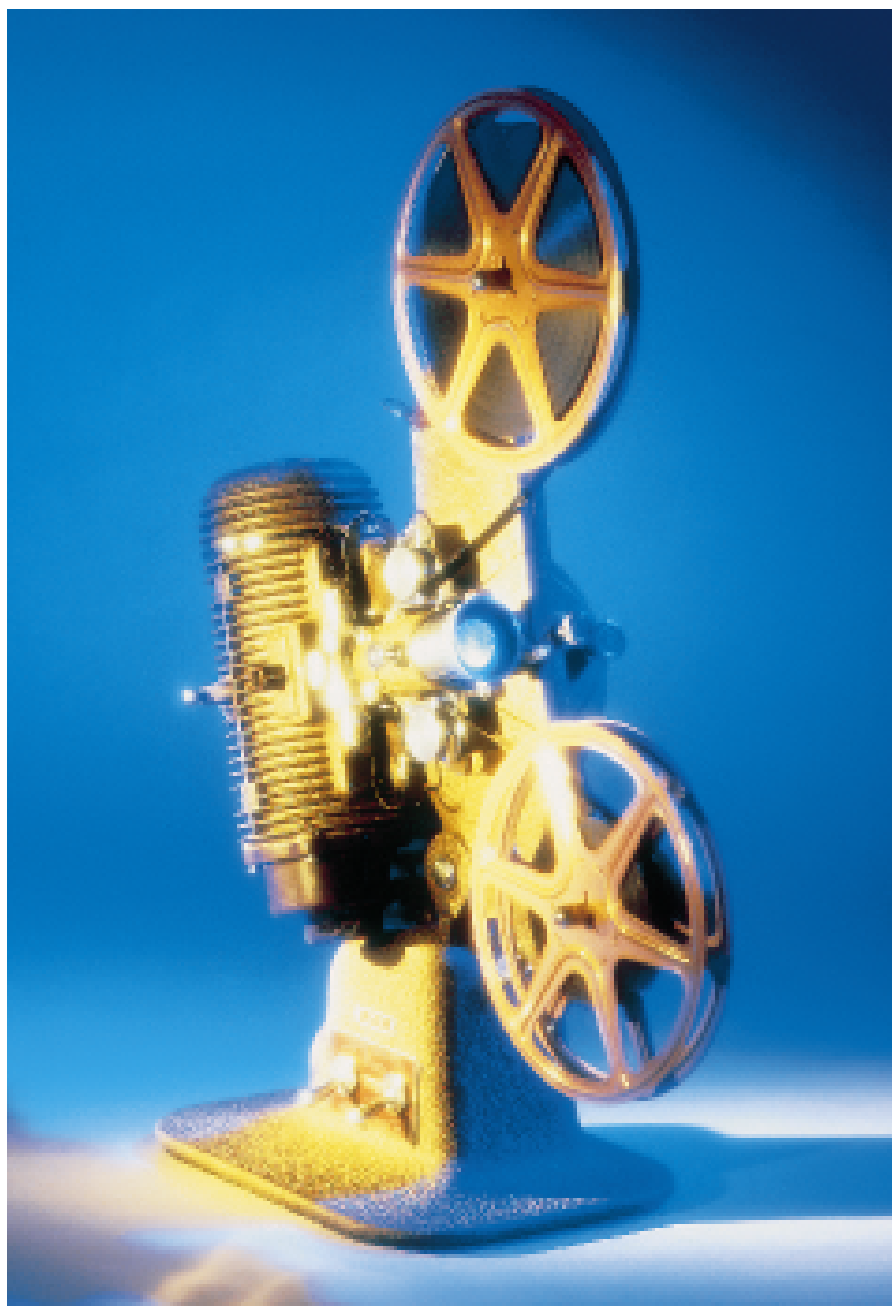
Obecnie istnieje już znaczna liczba algorytmów kodowania, opracowanych pod kątem różnych obszarów zastosowań. W ramach grupy kompresorów au-

dio i wideo technika MPEG zalicza się do metod najbardziej perspektywicznych.

Początki standardu MPEG

Pod koniec lat osiemdziesiątych grupa MPEG (Motion Picture Expert Group) rozpoczęła prace nad standardem cyfrowego zapisu ruchomych obrazów. Pierwszym owocem prac zespołu była norma MPEG-1, oparta na wcześniejszych specyfikacjach M-JPEG (Motion-Joint Photographic Expert Group) oraz H.261 komitetu CCITT (Comité Consultatif International Télégraphique et Téléphonique).

Stosunkowo prosta technika Motion-JPEG polega na kompresji pojedynczych zdjęć do formatu JPEG, które są



Na szczęście dekodowanie obrazu MPEG nie jest już tak skomplikowane. Na rynku dostępne są niedrogie karty rozszerzające do pecetów (np. Motion firmy Elsa), a niektórzy producenci kart graficznych instalują od razu w swoich produktach dekodery MPEG lub oferują je jako

Standardowo oryginalne sekwencje wideo dysponują rozdzielczością barw o proporcjach 4:2:2. W ten sposób określa się wzajemny stosunek kolejnych komponentów sygnału wideo: Y (jasność), Cr i Cb (różnice barw). Kompresję danych osiąga się poprzez redukcję

rozdzielczości, do której stosuje się tzw. filtry dziesiątkujące. Rozdzielczość pozioma jest zmniejszana o połowę w wyniku operacji ważonego uśredniania poszczególnych pikseli z jednoczesnym eliminowaniem niektórych sąsiednich punktów. W kierunku pionowym wykonuje się albo taką samą operację filtrowania, albo usuwa się po prostu co drugą linię obrazu. W ten sposób uzyskujemy obraz o rozdzielczości barw 4:2:0.

Głównym celem specyfikacji MPEG-4 jest umożliwienie transmisji dziesięciu zdjęć o dobrej jakości wraz z towarzyszącym dźwiękiem w ciągu sekundy. Przypuszczalna rozdzielczość tak przesyłanego obrazu ma wynosić 176x144 punkty obrazu. Zadanie to wymaga jednak wykorzystania kompresji o szczególnie wysokim współczynniku, na co nie pozwalają tradycyjnie stosowane techniki. Z tego względu planowane jest użycie zupełnie nowych metod, jak kompresja fraktalna czy morfologiczna.



Słowniczek

DCT (dyskretna transformacja cosinusowa) jest to metoda matematyczna umożliwiająca np. rozkład obrazu wideo. Standardy MPEG i JPEG wykorzystują dwuwymiarowe, ortonormalne transformacje DCT 8x8. W wyniku ich działania powstaje macierz współczynników transformacji, wykorzystywana następnie jako punkt wyjściowy do dalszych transformacji.

Interpolacja jest metodą matematyczną pozwalającą na dołączenie dodatkowych wartości do danej sekwencji numerycznej. Stosując rachunek różniczkowy wartości te można sprowadzić do postaci wielomianowej.

Interpolację stosuje się przy modyfikowaniu zapisanych obrazów, np. przy zmianie ich rozmiarów. W przypadku podwajania wielkości obrazu nowe piksele wyliczane są na podstawie sąsiednich, już istniejących punktów. Osoba obserwująca operację zauważy wyraźną poprawę jakości całego obrazu.

Metoda Huffmana jest to technika matematyczno-statystyczna umożliwiająca kodowanie wartości numerycznych. Jej idea polega na tym, że często występującym wartościom nadawane są krótkie kody, natomiast wartościom pojawiającym się rzadko – długie.

Takie rozwiązanie – w zależności od charakteru przetwarzanej informacji – pozwala na dość znaczną, bezstratną kompresję początkowego łańcucha danych. Przypisanie danych do odpowiednich kodów odbywa się za pomocą specjalnej tabeli.

Standard MPEG wykorzystuje pewną odmianę metody Huffmana. Z techniki tej korzystają również urządzenia faksowe należące do tzw. grupy 3.

W czasie wyświetlania skompresowanych sekwencji wideo musi zostać ponownie przywrócona pierwotna rozdzielczość. W tym celu również wykorzystuje się ważne uśrednianie (interpolację). W wyniku takiej operacji np. ciąg wartości 4-8-12 może zostać przekształcony w sekwencję 4-6-8-10-12. Poszczególne wagami są tu współczynniki filtrujące, wchodzące w skład tzw. filtra interpolacji.

Kompensacja ruchu

W sekwencjach wideo kolejne zdjęcia są do siebie bardzo podobne. Nawet w przypadku dynamicznych scen (np. wyścigu samochodowego) różnice pomiędzy poszczególnymi obrazami (ram-

kami) nie są duże. Tę prawidłowość wykorzystuje właśnie technika MPEG. Występujący między kolejnymi zdjęciami ruch danego obiektu lub osoby jest przekształcany do postaci wektora (Motion Vector).

Z uwagi na fakt, że rozpoznawanie złożonych obiektów jest w praktyce zbyt skomplikowane, analizie poddawane są małe fragmenty obrazu o wymiarach 16x16 pikseli (Macroblocks). Dla każdego bloku algorytm kompresji oblicza jego lokalne przesunięcie w stosunku do poprzedniego obrazu. Następnie ustalane są różnice zawartości w ramach takiej pary bloków. W wyniku opisanej operacji powstaje tzw. błędny obraz, który jest zapamiętywany na dysku. Złożoność obliczeniowa operacji zapisu wektora ruchu i błędnego obrazu jest tym mniejsza, im mniejsze są różnice między dwoma kolejnymi ramkami.

Opisana technika określana jest mianem blokowej kompensacji ruchu (Motion Compensation). Tę skomplikowaną procedurę matematyczną wykonuje mechanizm kodujący MPEG.

Typy obrazu MPEG

Dla potrzeb kompensacji ruchu mechanizm MPEG wykorzystuje trzy typy obrazu: I (Intra), P (Predictive-coded) i B (Bidirectional predictive-coded).

Obraz typu I zawiera skompresowaną informację dotyczącą pojedynczego, kompletnego zdjęcia (można go porównać z ramkami techniki Motion-JPEG).

Obrazy Intra są zapisywane w ciągu danych mniej więcej co pół sekundy.

Umożliwia to szybkie przeglądanie w obu kierunkach skompresowanego materiału wideo. Na podstawie zdjęć typu I tworzone są obrazy P i B.

Obrazy typu P powstają

w procesie szacowania ruchu (Motion Estimation) z wcześniej utworzonych, tymczasowych obrazów I lub P. Zarówno zdjęcia typu P, jak i I służą również ja-



Więcej szczegółów...

na temat standardu MPEG znajdziemy pod adresami: <http://www.vol.it/MPEG> i <http://www.crs4.it/HTML/LUIGI/PEG/mpegfaq.html>
Software – Xing-Player: <http://www.xingtech.com>.

ko punkty odniesienia przy ustalaniu wektorów przesunięcia.

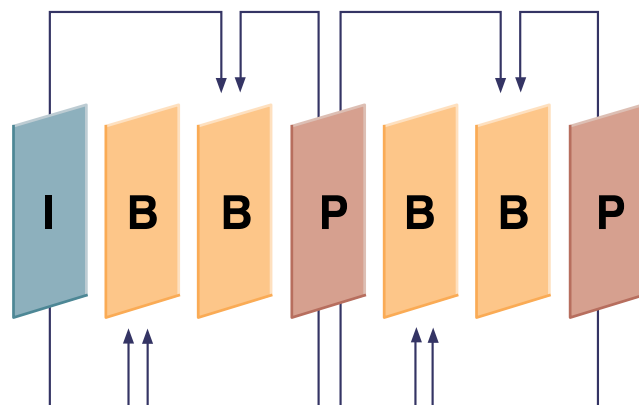
Obraz typu B generowany jest na podstawie wcześniejszego i późniejszego zdjęcia P lub I (a więc dwukierunkowo), bądź wyliczany za pomocą interpolacji sąsiednich obrazów.

W trakcie wyświetlania sekwencji MPEG następuje odtworzenie oryginalnych ramek. Jeśli np. poszczególne ramki ułożone są w sekwencji ...IBBPBBP..., to procedura odkodowania przebiega w następujący sposób: najpierw dekoduje się obraz I, co jest zadaniem prostym, gdyż zawiera on kompletną informację. Następnie za pomocą ramki typu I odkodowywany jest pierwszy obraz P. Obie ramki muszą nadal pozostawać w pamięci, gdyż kolejne pary zdjęć typu B są rekonstruowane na podstawie obrazu I i P.

Metoda kompresji bloków 16x16 do postaci typu I oraz P jest oparta na kodowanej algorytmem Huffmana transformacji DCT (patrz ramka), wykorzystującej bloki o wymiarach 8x8 pikseli (cztery bloki 8x8 tworzą jeden blok 16x16).

oprac. Jerzy Michalczyk (jk)

Generowanie ciągu danych MPEG



Obraz typu I zawiera kompletny, początkowy zestaw danych. Obrazy typu P są natomiast wyliczane na podstawie wcześniejszych, a typu B – wcześniejszych i późniejszych, obrazów I lub P