

RZECZPOSPOLITA
POLSKA



Urząd Patentowy
Rzeczypospolitej Polskiej

⑫ OPIS PATENTOWY ⑰ PL ⑪ 181909

⑬ B1

⑳ Numer zgłoszenia. 320177

⑥ IntCl⁷

㉑ Data zgłoszenia: 23.05.1997

H03M 7/30
G06F 17/30

⑤④

Sposób szybkiego uaktualniania modelu danych
w bezstratnych adaptacyjnych algorytmach kompresji statystycznej

CZYTELNIWA
OGÓLNA

④③ Zgłoszenie ogłoszono:
07.12.1998 BUP 25/98

⑦③ Uprawniony z patentu:
Politechnika Śląska, Gliwice, PL
Starosolski Roman, Gliwice, PL

④⑤ O udzieleniu patentu ogłoszono:
31.10.2001 WUP 10/01

⑦② Twórcy wynalazku:
Roman Starosolski, Gliwice, PL

⑦④ Pełnomocnik:
Ziółkowska Urszula, Politechnika Śląska

⑤⑦

Sposób szybkiego uaktualniania modelu danych w bezstratnych adaptacyjnych algorytmach kompresji statystycznej, polegających na kompresji kolejnych symboli, **znamienny tym**, że uaktualnianie modelu danych przeprowadza się jedynie po kompresji wybranych symboli kompresowanego ciągu, przy czym wybór symboli, po których następuje aktualizacja modelu określa się deterministycznie.

PL 181909 B1

Sposób szybkiego uaktualniania modelu danych w bezstratnych adaptacyjnych algorytmach kompresji statystycznej

Zastrzeżenie patentowe

Sposób szybkiego uaktualniania modelu danych w bezstratnych adaptacyjnych algorytmach kompresji statystycznej, polegających na kompresji kolejnych symboli, **znamienny** tym, że uaktualnianie modelu danych przeprowadza się jedynie po kompresji wybranych symboli kompresowanego ciągu, przy czym wybór symboli, po których następuje aktualizacja modelu określa się deterministycznie.

* * *

Przedmiotem wynalazku jest sposób szybkiego uaktualniania modelu danych w bezstratnych adaptacyjnych algorytmach kompresji statystycznej znajdujący zastosowanie w systemach transmisji i składowania danych.

Kompresja jest procesem usuwania nadmiarowości z ciągów symboli, określanych tutaj także jako zbiory danych. Nadmiarowość jest naturalna dla niektórych klas zbiorów danych takich jak teksty, gdzie często powtarzają się niektóre wyrazy lub ich fragmenty, czy też obrazy w których sąsiadujące piksele mają zazwyczaj zbliżoną barwę. Dzięki kompresji można oszczędniej wykorzystywać zarówno pamięci masowe jak i szerokości pasm przenoszenia sieci komputerowych. Kosztem uzyskania tych oszczędności jest wykorzystanie pamięci operacyjnej i czasu jednostki centralnej komputera.

Bezstratne algorytmy kompresji to takie, których działanie jest odwracalne, tzn. algorytm kompresji jest bezstratny, gdy istnieje odpowiedni algorytm dekompresji, który dla każdego zbioru danych potrafi z danych skompresowanych odtworzyć dane identyczne z danymi poddanymi procesowi kompresji.

Algorytmy kompresji statystycznej można podzielić na statyczne i adaptacyjne.

Algorytmy statyczne działają w dwu etapach - najpierw przeglądają cały ciąg symboli, który należy skompresować gromadząc w tzw. modelu danych informacje o bezwarunkowych i warunkowych prawdopodobieństwach wystąpienia poszczególnych symboli. Następnie w drugim przebiegu dokonują kompresji kolejnych symboli ciągu na podstawie modelu danych uzyskanego podczas pierwszego przebiegu.

Algorytmy adaptacyjne kompresji statystycznej przeglądają ciąg symboli jednokrotnie. Każdy symbol jest kompresowany na podstawie modelu zbudowanego z wykorzystaniem symboli poprzedzających symbol aktualnie kompresowany. Po skompresowaniu danego symbolu modyfikuje się model danych tak, by uwzględniał także ten ostatnio skompresowany symbol. Algorytmy adaptacyjne nie wymagają znajomości rozmiaru zbioru danych do jego kompresowania.

Istnienie szybkich, adaptacyjnych algorytmów kompresji konieczne jest w systemach, które dysponując ograniczonymi zasobami pamięci i mocy obliczeniowej muszą przetwarzać dane napływające w czasie rzeczywistym, np. podczas akwizycji obrazu video. Szybki algorytm kompresji może także okazać się niezbędny w systemach, gdzie kompresja i dekompresja dużej ilości danych odbywa się wielokrotnie i czas jej trwania decyduje o komforcie pracy użytkownika.

Jeżeli ciąg symboli podlegający kompresji jest duży, to w przypadku niektórych klas zbiorów danych, np. obrazów, po dokonaniu kompresji odpowiedniej liczby symboli model danych staje się dokładny i dość stabilny. W znanych rozwiązaniach proponuje się zaprzestanie uaktualniania modelu po przeczytaniu arbitralnie określonej liczby symboli, co przyspiesza działanie kompresora. Takie rozwiązania równocześnie powodują, iż model nie

uwzględnia nowych symboli, a zatem gdy zmieni się charakterystyka danych model nie będzie ich prawidłowo opisywał.

W znanych rozwiązaniach celem przyspieszenia procesu kompresji kompresowany ciąg dzielony jest na dwie części. Dla pierwszej z nich każdorazowo po kompresji kolejnego symbolu ciągu następuje aktualizacja modelu danych. Druga część jest kompresowana z użyciem stałego modelu uzyskanego podczas kompresji części pierwszej.

Sposób według wynalazku polega na tym, że uaktualnianie modelu danych przeprowadza się jedynie po kompresji wybranych symboli kompresowanego ciągu, przy czym wybór symboli, po których następuje aktualizacja modelu określa się deterministycznie.

Wybór symboli, po których następuje aktualizacja modelu danych czyli każdorazowo liczba symboli, po których nie nastąpi aktualizacja modelu może być uzależniona od ilości już skompresowanych symboli. Po skompresowaniu określonej ilości symboli może to być liczba stała, lub np. w przypadku gdy ciąg może mieć charakterystykę okresową, każdorazowo po aktualizacji modelu liczbę tę można wybierać za pomocą tablicy liczb losowych lub generatora liczb pseudolosowych. Wybór tej liczby musi być określony deterministycznie, tak by algorytmy kompresora i dekompresora uaktualniały model w ten sam sposób.

Sposób według wynalazku umożliwia przyspieszenie procesu kompresji statystycznej poprzez ograniczenie czasu potrzebnego na aktualizację modelu danych bez jednoczesnego uniemożliwiania reakcji modelu na zmianę charakterystyki kompresowanych danych.

Wynalazek objaśnia przykład.

Kompresowany jest ciąg o rozmiarze 1000000 symboli zawierający opis zdigitalizowanego zdjęcia 1000 x 1000 pikseli. Pierwsze 100000 symboli kompresuje się uaktualniając model danych każdorazowo po kompresji kolejnego symbolu. Poczynając od symbolu 100001 model danych aktualizuje się tylko po kompresji co 5-tego symbolu, dzięki czemu aktualizacja modelu następuje 5-ciokrotnie rzadziej i co za tym idzie zajmuje mniej czasu.

181 909

Departament Wydawnictw UP RP. Nakład 60 egz.
Cena 2,00 zł.