

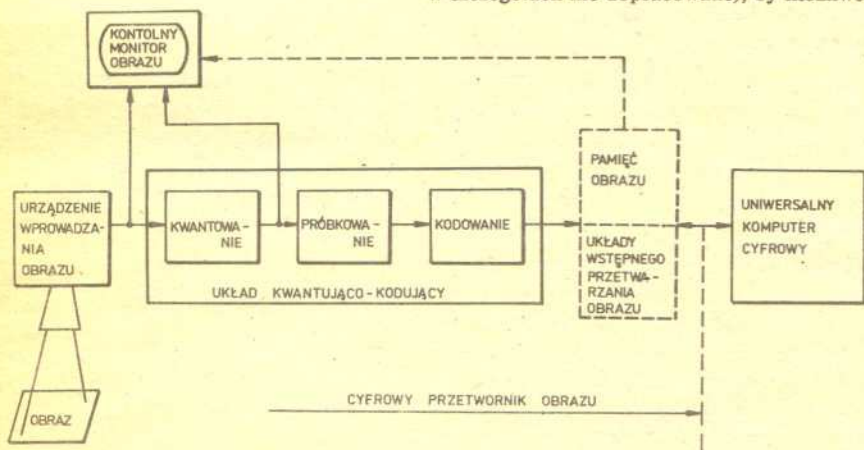
Komputerowa analiza obrazów wizualnych

Mgr inż. Zenon KULPA

W ostatnich latach zaobserwować można burzliwy rozwój jednego z nowszych zastosowań komputerów, mało znanego niespecjalistom, a mianowicie komputerowego przetwarzania i analizy obrazów wizualnych. Jak wiadomo, zmysł wzroku dostarcza człowiekowi około 1000 razy więcej informacji niż następny w kolejności słuch. Ludzie są wzrokowcami — znakomita większość użytecznej dla nich informacji, zarówno pochodzącej z otoczenia, jak i służącej do komunikacji międzyludzkiej, ma postać obrazów optycznych. Nic więc dziwnego, że wraz z postępami mechanizacji i automatyzacji różnych czynności wykonywanych dotychczas bezpośrednio przez człowieka, nadszedł również czas automatyzacji przetwarzania obrazów. Pierwsze prace w tym kierunku pojawiły się jednak dopiero na początku lat sześćdziesiątych. Co prawda, fotografia znana jest już prawie półtora stulecia, a fotokomórka od prawie wieku, lecz nie dają one jeszcze możliwości automatycznego wydzielania z obrazu użytecznej informacji zakodowanej w kształtach i rozmieszczeniu różnorodnych obiektów na obrazie. A uzyskanie takich możliwości pozwala dopiero mówić o automatyzacji analizy obrazów. Tak więc dopiero łączne pojawienie się techniki telewizyjnej (pozwalającej zamieniać obrazy na postać elektroniczną), komputerów elektronicznych (pozwalających przetwarzać informację zakodowaną w postaci elektronicznej) i techniki układów scalonych (umożliwiających budowę dostatecznie złożonych i szybkich działających układów przetwarzających i liczących) daje nadzieję uporania się z problemem automatyzacji analizy rzeczywistych obrazów. Ilość informacji, którą trzeba przetworzyć, jest bardzo duża: w średniej klasy komputerowych systemach analizy obrazów jest ona rzędu miliona bitów na 1 obraz, a wzrok przetwarza jej znacznie więcej — w siatkówce oka jest ok. 120 mln pojedynczych fotoreceptorów!

Urządzenia automatycznej analizy obrazów można z grubsza podzielić na wyspecjalizowane i uniwersalne. Te pierwsze przystosowane są do analizy pewnej określonej klasy obrazów i wydzielania z obrazów tej klasy ściśle określonych informacji w sztywny sposób. Przykładem mogą tu być tzw. czytniki pisma, pozwalające wyczytać (np. do komputera lub maszyny drukarskiej) dane w postaci tekstu drukowanego lub maszynopisu. Urządzenie rozpoznaje w tym celu litery alfabetu i inne znaki drukarskie według sztywnego algorytmu, zrealizowanego w konstrukcji urządzenia. Większość urządzeń tego typu jest przy tym przystosowana do czytania tylko jednego ściśle określonego kroju czcionki (lub w najlepszym wypadku kilku krojów). Urządzenia uniwersalne zawierają komputer uniwersalny, za pomocą którego realizowane być mogą teoretycznie dowolne algorytmy przetwarzania, analizy i rozpoznawania obrazów. Urządzenia uniwersalne stosowane są zatem głównie do celów badawczych (opracowywanie nowych metod analizy obrazów) oraz do analizy złożonych obrazów, dla których algorytmy przetwarzania są zbyt skomplikowane (i często w szczególności nie dopracowane), by możliwe było ich efektywne zrealizowanie sprzętowe.

Urządzenie uniwersalne ma strukturę ogólną podaną na rys. 1: składa się z komputera i z tzw. cyfrowego przetwornika obrazu, służącego do wprowadzania informacji obrazowej do tego komputera. Układ wprowadzania obrazu zamienia obraz optyczny na zbiór reprezentujących go wartości elektrycznych. Układy takie działają na różnych zasadach: może to być np. pojedyncza fotokomórka przesuwana mechanicznie nad obrazem (np. fotografią), układ z plamką świetlną generowaną i przesuwaną na ekranie lampy oscyloskopowej, a następnie rzutowaną na obraz, lub, najczęściej stosowana, kamera telewizyjna. Następnie układ kwantująco-kodujący rozбивa ciągły rozkład jasności (reprezentowanych elektrycznie) na skończoną liczbę oddzielnych punktów (jest to tzw. próbkowanie), jasność każdego z nich



Rys. 1. Ogólna struktura uniwersalnego systemu przetwarzania i analizy obrazów.

Kącik czytelnicy

W jednej z niedawno wydanych popularnych książek o matematyce zauważyliśmy następujące zdanie (cytujemy z pamięci): „ulamek 355/113 wynalazł w V w. p.n.e. chiński uczonec Tsu Chiu'ang Chih.” Domyślamy się, że 1/2 wynależii

Babilończycy, 22/7 Archimedes, a $\frac{m_1 \cdot m_2}{r^2}$

Newton. Kto z naszych Czytelników wie, kto wynalazł ulamek $\frac{65478958}{79654002}$? A może

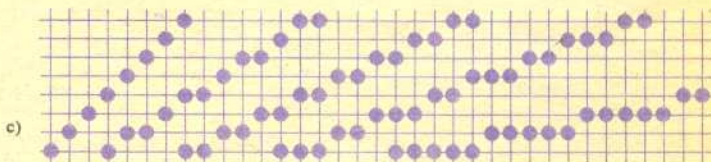
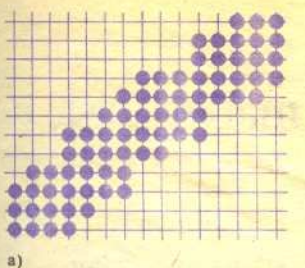
on jeszcze czeka na swojego wynalazcę?

przybliża jednym ze skończonej liczby rozróżnianych poziomów jasności (kwantowanie) i zamienia na wartość cyfrową (kodowanie). W niektórych systemach tak zakodowany obraz zapamiętywany jest, przed przesłaniem do komputera, w dodatkowej pamięci obrazu (skąd porcjami odczytuje go komputer w trakcie przetwarzania) lub też część prostszych operacji przetwarzania wykonywanych jest poza komputerem, przez specjalne układy cyfrowe (np. zliczanie liczby jasnych punktów lub spójnych składowych obrazu). Dla orientacji podajemy, że rozmiary pojedynczego obrazu wahają się w systemach uniwersalnych od 64 × 64 punkty do 1024 × 1024 punkty, a liczba rozróżnianych poziomów jasności punktu od 2 (obraz czarno-biały, bez pośrednich szarości) do 256. Odpowiada to ilości informacji na 1 obraz od 4096 do ponad 8 mln bitów. Średnie parametry: 512 × 512 punktów, 16 poziomów jasności (ponad 1 mln bitów na obraz).

Konieczność zamiany obrazu rzeczywistego (będącego ciągłym rozkładem jasności na dwuwymiarowej płaszczyźnie) na skończoną ilość liczb możliwą do przetworzenia komputerowego, pociąga za sobą konieczność próbkowania obrazu (rys. 1), tj. odczytu

wartości jasności w skończonej liczbie dyskretnych punktów (tzw. punktów rastru). Uzyskany obraz dyskretny może być dopiero przetwarzany w komputerze — podobne przekształcenie obrazu zachodzi zresztą także w siatkówce oka. W rezultacie jednak pojęcia geometryczne używane do opisu obrazów (np. pojęcie linii) tracą swój euklidesowy, ciągły sens. Rzeczywiste linie na obrazie mają skończoną grubość i dokładność rysunku — na skutek dyskretniej struktury rastru i nieuniknionych zakłóceń i błędów próbkowania — zmieniają się w szerokie obszary punktów o poszarpanych brzegach (rys. 2a) lub też ulegają poprzerywaniu na niespójne fragmenty (rys. 2b). Nawet jednak teoretycznie „idealne” dyskretne linie proste nie są wcale proste (rys. 2c). Rodzi to mnóstwo problemów przy tworzeniu algorytmów analizy obrazów. Linie rzeczywiste próbuje się sprowadzić do „idealnych” — grube „pocienia” się, usuwając z nich punkty brzegowe, a linie poszarpane łączy się uzupełniając przerwy.

Rys. 2. Przykłady rzeczywistych linii dyskretnych: „grubej” (a) i poszarpanej (b) oraz przykłady linii prostych „idealnych” o różnych nachyleniach.

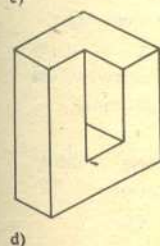
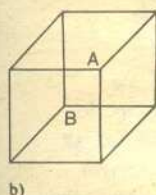
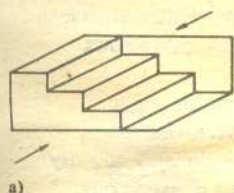
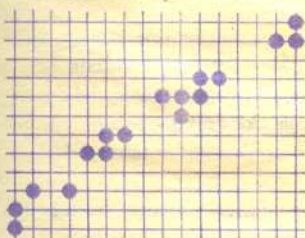


Algorytmy realizujące te operacje nie są jednak proste — muszą bowiem omijać wiele „pułapek” występujących na rzeczywistych obrazach. Gruba linia o poszarpanych brzegach powoduje przy pocienianiu powstanie „gałązek” odchodzących w przypadkowych miejscach od pocienionej linii, a ostre załamania biegu linii są przy pocienianiu zaokrąglane. Może to spowodować np. zamianę litery B na cyfrę 8 (w czytniku pisma). Linie poszarpane przebiegające blisko siebie mogą zostać z kolei przy uzupełnianiu połączone ze sobą („na krzyż”), powodując np. zinterpretowanie biegnących w jednym kierunku warstw komórek na preparacie mikroskopowym tkanki jako siatki nieregularnych linii o kierunku prostopadłym do rzeczywistego... Jeszcze większe trudności sprawia wydzielenie krawędzi obiektów na obrazie, zwłaszcza gdy są one rozmyte w różnym stopniu w różnych miejscach. Niestety, w rzeczywistych obrazach rzadko mamy do czynienia z idealnymi krawędziami.

Nawet jednak po uzyskaniu linii „cienkich”, idealnych, kłopoty nie znikają. Pojawiają się problemy interpretacji — jaki układ punktów rastru należy uznać za dyskretną wersję linii prostej, a jaki nie; jak obliczyć krzywiznę takiej „schodkowej” krzywej (rys. 2c), a jak długość jej odcinka lub obwód nieregularnej figury. Jest oczywiste, że obliczając długość punktowej linii jako sumę odległości kolejnych punktów rastru wzdłuż linii popełniamy spory błąd i to różny dla różnych nachyleń linii (rys. 2c). Badaniem i rozwiązywaniem tego typu problemów zajmuje się dział analizy obrazów zwany teorią obrazów dyskretnych lub geometrią dyskretną. Okazuje się, że ma on wiele punktów styecznych z teorią liczb — abstrakcyjnym działem matematyki, pozornie zupełnie odległym od jakichkolwiek zastosowań praktycznych. Innego rodzaju problemy wynikają w zastosowaniach automatycznej analizy obrazów do określania trójwymiarowej struktury „ogłądanej” sytuacji, np. w próbach konstrukcji „sztucznego oka” sterującego pracą robotów (przemysłowych lub tzw. „inteligentnych”). Centralnym zagadnieniem jest wtedy określanie trójwymiarowego kształtu obiektu przestrzennego na podstawie jego płaskiego wizerunku (rzutu) na płaszczyźnie obrazu. Jest to zadanie istotnie niejednoznaczne — teoretycznie istnieje nieskończenie wiele różnych obiektów przestrzennych o tym samym dwuwymiarowym rzucie. Oko człowieka (wraz z ośrodkiem wzroku w mózgu) interpretuje płaski wizerunek (obraz, rysunek, fotografię) w większości przypadków jednoznacznie, odtwarzając przestrzenną strukturę przedstawionych obiektów i nie zauważając wręcz innych (choć teoretycznie równie uprawnionych) możliwości. Dążenie do naśladowania tych możliwości oka jest przedmiotem omawianego działu analizy obrazów.

Dużą pomocą w wyjaśnieniu zasad trójwymiarowej interpretacji obrazów płaskich okazało m.in. znalezienie obrazów sprawiających także oku zasadnicze trudności w określeniu jednoznacznej interpretacji. Są to tzw. figury niejednoznaczne (rys. 3a, b) i figury niemożliwe (rys. 3c, d). Dla figur niejednoznacznych oko znajduje dwie różne, choć równouprawnione interpretacje (dla schodów Schrödera jest to „widok z góry” lub „widok z dołu”; dla sześcianu Neckera albo punkt A jest w głębi, a B bliżej, albo na odwrót). Dla figur niemożliwych oko nie jest w stanie znaleźć żadnej poprawnej interpretacji. Choć czynią one wrażenie pewnych trójwymiarowych obiektów, to ich przestrzenna struktura, sugerowana przez oko, jest niemożliwa do fizycznej realizacji. Figury te mają jednak realizowalną interpretację przestrzenną, lecz niełatwo jest ją znaleźć naszemu oku, tak doskonałemu w innych sytuacjach. Analiza tego typu figur pozwala określić wiele heurystycznych mechanizmów tworzenia w mózgu hipotez percepcyjnych przy trójwymiarowej interpretacji obrazów. Umożliwia to ich modelowanie w komputerowych systemach rozpoznawania obrazów.

Oprócz problemów zasygnalizowanych powyżej, automatyzacja analizy obrazów wymaga rozwiązania mnóstwa innych problemów różnej natury: technicznych (np. konstrukcja dokładnych i niezawodnych układów wprowadzania), informatycznych (np. zagadnienia efektywnego programowania algorytmów przetwarzania), metodologicznych, matematycznych i wielu innych.



Rys. 3. Przykłady figur niejednoznacznych: schody Schrödera (a) i sześcianu Neckera (b) oraz figur niemożliwych (c, d).