Akademia Górniczo-Hutnicza

im. Stanisława Staszica w Krakowie

WYDZIAŁ ELEKTROTECHNIKI, AUTOMATYKI, INFORMATYKI i ELEKTRONIKI

Katedra Automatyki

PRACA MAGISTERSKA

Wyszukiwanie obrazów na podstawie zawartości

Autor: Marcin Strach

Kierunek studiów: Automatyka i Robotyka

Specjalność: Robotyka

Promotor:

dr inż. Paweł Rotter

Kraków 2010

OŚWIADCZENIE AUTORA PRACY

Oświadczam, świadomy odpowiedzialności karnej za poświadczenie nieprawdy, że niniejszą pracę dyplomową wykonałem osobiście i samodzielnie, i nie korzystałem ze źródeł innych niż wymienione w pracy.

……

Podpis autora

Pragnę złożyć serdeczne podziękowania dla Pana promotora **dr inż. Pawła Rottera** za zaangażowanie i pomoc podczas pisania niniejszej pracy magisterskiej.

STRESZCZENIE PRACY

Metody wyszukiwania obrazów na podstawie zawartości są jedną z najszybciej rozwijających się dziedzin przetwarzania i analizy obrazu. Pod pojęciem metod wyszukiwania kryje się bogata różnorodność technik począwszy od prostych funkcji porównujących podobieństwa obrazów za pomocą np. dominującego koloru, kończąc na złożonych i odpornych automatycznych metodach wyszukiwania i adnotacji danych multimedialnych. Taka charakteryzacja pozwala zatem umiejscowić tę dziedzinę wiedzy pośrodku zagadnień dotyczących m. in. komputerowej analizy obrazu, zarządzania bazami danych, interakcji typu człowiek – komputer, maszyn uczących, czy wyszukiwania informacji.

Celem niniejszej pracy magisterskiej jest dokonanie przeglądu istniejących metod wyszukiwania obrazów na podstawie zawartości oraz omówienie eksperymentalnych i komercyjnych systemów, które znalazły zastosowanie w wielu dziedzinach życia społecznego. Dodatkowo praca zawiera analizę głównych problemów i przyszłych kierunków rozwoju ze szczególnym uwzględnieniem aplikacji przeznaczonych dla urządzeń mobilnych.

ABSTRACT

Content-Based Image Retrieval methods are one of the fastest developing domains of image processing and analysis. This concept holds a rich variety of techniques starting from simple functions which compare visual similarity of images and finishing with complex and robust automatic annotation and search engines. This characterization of CBIR as a field of study places it in the middle of such issues like computer vision, database management, human-computer interactions, learning machine or information retrieval.

The purpose of present master's thesis is to survey an existing CBIR methods and to give some examples of commercial systems which can be used in many domains of social life. In addition this work contains an analysis of main problems and future developments especially considering applications for mobile devices.

Spis treści

Rozdział 1. Wstęp

Gwałtowny wzrost popularności sieci World Wide Web oraz rozwój domowych bibliotek multimedialnych pociąga za sobą nieustanne powiększanie dostępnych zbiorów danych cyfrowych takich, jak zdjęcia, filmy, czy muzyka. Sytuacja ta wydaje się obecnie całkowicie naturalna i wynika z szybkiego postępu społeczeństwa informacyjnego. Jednak pociąga ona za sobą konieczność rozwoju nowych technik zarządzania bazami danych, które jednocześnie zapewniałyby wysoką efektywność oraz szybkość wyszukiwania potrzebnych informacji.

Wychodząc naprzeciw powyższym wymaganiom naukowcy zaczęli opracowywać techniki wyszukiwania informacji na podstawie zawartości (ang. Content-Based Information Retrieval), które stały się bardzo popularne na przestrzeni ostatnich dwudziestu lat (Zhang 2007). Wśród nich dużą grupę reprezentują metody koncentrujące się na sposobach indeksowania, porównywania i wyszukiwania obrazów cyfrowych. Ze względu na trudności związane z opisem wizualnym zawartości danych, początkowe metody wykorzystywane w tych systemach opierały się na tekstowym wyszukiwaniu informacji (Chang, Fu 1980). Wymagało to jednak uzupełniania obrazów tekstowymi adnotacjami, co było bardzo czasochłonne, ale konieczne do użycia znanych na ówczesny stan wiedzy technik wyszukiwania obrazów na podstawie tekstu (ang. Text-Based Image Retrieval) (Long, Zhang, Dagan Feng 2003). Jednak, jak się szybko okazało, takie podejście do problemu wyszukiwania informacji nie mogło zostać zautomatyzowane, gdyż z zasady wymagało ingerencji człowieka.

Stąd też na początku lat dziewięćdziesiątych XX w. powszechnie zaczęto rozwijać i stosować techniki Content-Based Image Retrieval (CBIR). Skupiły one uwagę naukowców zajmujących się różnymi dziedzinami informatycznymi, jak np. zagadnieniami wizji komputerowej, zarządzania bazami danych, interakcjami pomiędzy człowiekiem, a komputerem, czy w końcu aspektami cyfrowego wyszukiwania informacji. Wspólny wysiłek oraz wzajemna współpraca doprowadziły do szybkiego rozwoju metod CBIR (Zhang, Zhong 1995; Rui, Huang, Chang 1999; Del Bimbo 1999). Liczba publikacji związanych z problemami przetwarzania i analizy obrazu, efektywnego indeksowania i wyszukiwania danych gwałtownie wzrosła, co pozytywnie wpłynęło na rozwój komercyjnych i badawczo-naukowych systemów CBIR. Dziedzina ta zaczęła cieszyć się ogromnym zainteresowaniem ze względu na różnorodne możliwości zastosowania m. in. w systemach bezpieczeństwa, w medycynie diagnostycznej, czy w projektowaniu i zarządzaniu (Eakins, Graham 1999; Datta, Joshi, Li, Wang 2008).

Metody Content-Based Image Retrieval do reprezentacji i indeksowania wykorzystują wizualne cechy obrazu takie, jak np. kolor, tekstura, kształt, czy relacje przestrzenne między obiektami. W typowym systemie CBIR proces porównywania i dopasowania odbywa się za pomocą wielowymiarowych deskryptorów cech. W celu

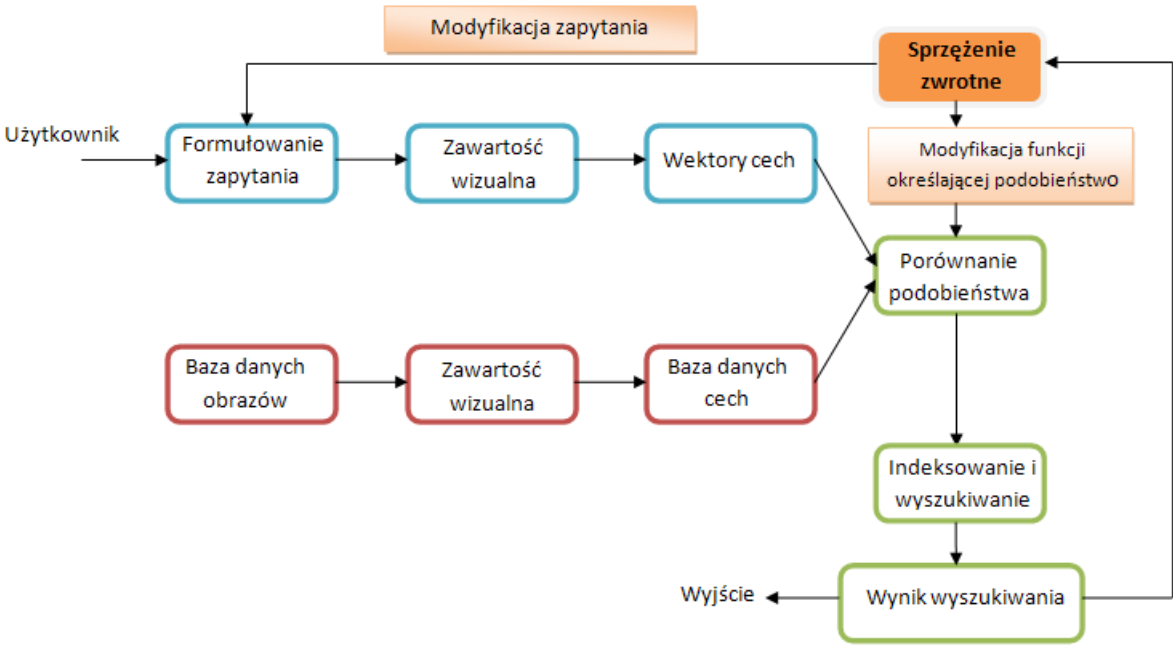
wyszukania pożądanych obrazów użytkownik wysyła do systemu zapytanie w postaci np. przykładowego zdjęcia (ang. query by example) lub szkicu (ang. query by sketch). Następnie system zamienia to zapytania zgodnie ze swoją wewnętrzną reprezentacją na wektory cech obrazu. Dzięki temu możliwe jest użycie znanych miar, bądź metryk podobieństwa, które określają odległość pomiędzy obrazem wejściowym użytkownika, a wynikiem procesu wyszukiwania.

Kolejnym krokiem do zwiększenia efektywności i skuteczności systemów CBIR jest zaadoptowanie na potrzeby procesu wyszukiwania informacji technik sprzężenia zwrotnego typu ***relevance feedback*** (Rocchio 1971; Rui, Huang, Mehrotra 1998; Minka, Picard 1997; Su, Zhang i in. 2003). Główna ideą metod sprzężenia zwrotnego jest zastosowanie subiektywizmu oraz percepcji człowieka w procesie wyszukiwania obrazów. Interakcja użytkownika polega na możliwości wpływania na wynik wyszukiwania. Obrazy wynikowe dostarczane przez system są poddawane ocenie przez użytkownika, który określa, czy dany wynik jest odpowiedni (ang. relevant), czy nieodpowiedni (ang. irrelevant). W oparciu o te dane system modyfikuje zapytanie i zwraca listę nowych obrazów. Podstawowym zadaniem w procesie sprzężenia zwrotnego jest zatem (Zhang 2003):

- użycie pozytywnych i negatywnych przykładów do tworzenia nowego zapytania,

- dopasowanie (wybór) odpowiedniej miary, bądź metryki podobieństwa.

Poniżej przedstawiamy schemat blokowy systemu Content-Based Image Retrieval z uwzględnieniem pętli sprzężenia zwrotnego:



Rys. 1.1 Przykładowy schemat systemu CBIR ze sprzężeniem zwrotnym – relevance feedback.

Za główne cele poniższej pracy magisterskiej postawiliśmy:

- Dokładne opisanie metod stosowanych w technikach wyszukiwania i rozpoznawania obrazów (ang. pattern recognition), które zostały podzielone na dwie zasadnicze części: metody bez interakcji i z interakcją użytkownika.
- Przedstawienie pierwszych i najpopularniejszych systemów CBIR – Content-Based Image Retrieval z dodatkowym uwzględnieniem obecnych aplikacji i ich możliwości.
- Przeanalizowanie głównych problemów i kierunków rozwoju technik MIR (ang. Multimedia Information Retrieval).
- Wskazanie zastosowań systemów CBIR w różnych dziedzinach życia publicznego ze szczególnych uwzględnieniem aplikacji przeznaczonych dla telefonów komórkowych i urządzeń PDA (ang. Personal Digital Assistant).

Kolejność rozdziałów pracy wynikła z chęci początkowego omówienia technik i metod stosowanych w systemach CBIR, a następnie podania przykładowych aplikacji, które stały się bardzo popularne na przestrzeni ostatnich 20 lat. W kolejnych rozdziałach omówiono główne problemy i kierunki rozwoju metod wyszukiwania obrazów na podstawie zawartości, a dopiero później przedstawiono konkretne przykłady zastosowania. Praca magisterska składa się zatem z 7 rozdziałów, które zawierają następujące treści:

W **rozdziale 1** prezentujemy główne cele pracy i omawiamy związek przyczynowo-skutkowy, który w konsekwencji doprowadził do powstania metod i systemów Content-Based Image Retrieval. Dodatkowo stwierdzamy, że systemy te są naturalną odpowiedzią na zapotrzebowanie na efektywne i szybkie narzędzia do zarządzania ogromnymi bazami danych multimedialnych.

Rozdział 2 zawiera przegląd metod stosowanych w CBIR, które zostały podzielone na dwie części w zależności od występowania elementów interakcji człowieka z systemem komputerowym. W przypadku metod bez interakcji skoncentrowaliśmy się na opisie technik wykorzystujących wizualną zawartość obrazu. Zgodnie z podziałem zastosowanym w pracy Long, Zhang, Dagan Feng (2003) wyróżniliśmy tutaj metody rozpoznawania koloru, tekstury i kształtu, które są trzema głównymi cechami służącymi do analizy zawartości obrazu. W przypadku deskryptorów koloru opisaliśmy:

- problemy właściwego doboru przestrzeni barw, jednocześnie podkreślając szerokie możliwości wyboru przestrzeni HSV (ang. Hue Saturation Value) ze względu na jej intuicyjną interpretację,
- możliwości wykorzystania teorii prawdopodobieństwa do rozkładu koloru na obrazie – użycie momentów koloru,
- najbardziej powszechną metodę bazującą na histogramie koloru, dodatkowo uwzględniając metryki stosowane przy ich porównywaniu,
- techniki wykorzystujące wektor spójności oraz korelogram koloru.

W przypadku opisu tekstury obrazu skoncentrowaliśmy się na analizie:

- cech teksturowych Tamury tj.: gruboziarnistość, kontrast, kierunkowość, liniowe podobieństwo, regularność, chropowatość,
- składników dekompozycji Wold'a tzn. harmoniczny, krótkotrwały (zanikający) i interdeterministyczny,
- filtru Gabora i transformaty falkowej.

W części poświęconej cechom kształtu wyróżniliśmy m. in.:

- metody momentowe rzędu od zerowego do czwartego oraz uwzględniliśmy niezmienniki momentowe,
- metody obracanego kąta i aktywnego konturu,
- deskryptory Fouriera.

Podrozdział poświęcony metodom bez interakcji zakończyliśmy opisem technik wykorzystywanych do wyznaczania relacji przestrzennych pomiędzy obiektami obrazu oraz podaliśmy metody indeksowania i redukcji wektora cech (np. PCA – Principal Component Analysis).

W rozdziale 2.2 przedstawiliśmy przykłady metod interakcji użytkownika z systemem komputerowym. Wśród nich wyróżniliśmy 4 sposoby formułowania zapytań: poprzez wybór kategorii, szkic, przykład oraz poprzez grupę przykładów. Najistotniejszą częścią tego rozdziału jest opis metod **sprzężenia zwrotnego** typu ***relevance feedback***. Dokonaliśmy podziału metod sprzężenia zwrotnego na dwie części:

- algorytmy klasyczne – bazujące na formule Rocchio (1971) (przemieszczanie zapytania – ang. query point movement). Podaliśmy dokładną metodykę zastosowaną w systemie MARS (Rui, Huang, Mehrotra 1997 i 19998).
- problem maszyny uczącej z pamięcią – przedstawienie sprzężenia zwrotnego jako metody probabilistycznej (reguła klasyfikacyjna Bayesa).

W celu zobrazowania postępu, jaki dokonał się w sposobach podejścia do tematu Content-Based Image Retrieval dokonaliśmy dokładnej analizy metod opartych na lokalnych deskryptorach obrazu. Wśród nich wyróżniliśmy metodę Scale Invariant Feature Transform – SIFT (Lowe 1999 i 2004) oraz metodę Speeded Up Robust Features – SURF (Bay, Tuytelaars, Van Gool 2006). Na zakończenie rozdziału podaliśmy przykład systemu wykorzystującego metody percepcyjnego grupowania człowieka do wyszukiwania i klasyfikacji obrazów zawierających duże obiekty architektoniczne takie, jak budynki, wieże, mosty itp.

W **rozdziale 3** opisaliśmy pierwsze komercyjne i badawczo – rozwojowe systemy CBIR. Skoncentrowaliśmy się na wyborze systemów, które cieszyły się dużą popularnością w latach 90-tych XX w. oraz miały znaczący wpływ na rozwój późniejszych aplikacji. Wśród nich wyróżniliśmy system QBIC firmy IBM, VIR Image Engine, Photobook, VisualSEEk i MARS. W celu zobrazowania obecnej sytuacji

panującej w świecie CBIR podaliśmy przykłady organizacji naukowych, które zrzeszają grupy badawcze zajmujące się tą dziedziną informatyki.

Rozdział 4 prezentuje główne problemy metod wyszukiwania obrazów na podstawie zawartości. Do najważniejszych z nich zaliczyliśmy występowanie (Datta, Joshi, Li, Wang 2008):

- tzw. „luki sensorycznej" czyli różnicy między definiowaniem obiektu w rzeczywistym świecie, a jego odzwierciedleniem w postaci komputerowego opisu dostarczonego za pomocą obrazu,
- tzw. „luki semantycznej" czyli różnicy zgodności pomiędzy informacją wyekstrahowaną z obrazu, a jej interpretacją, która może się zmieniać w zależności od sytuacji i celu poszukiwań.

W dalszej części rozdziału dokładnie opisaliśmy zjawisko braku wspólnych kryteriów oceny systemów CBIR. Podaliśmy także przykłady możliwych metod oceny ich skuteczności i efektywności dzieląc je na metody jednowartościowe i reprezentacje graficzne. Jednocześnie przedstawiliśmy możliwe kierunki rozwoju aplikacji CBIR w narzędziach m. in. do zarządzania multimedialnymi bazami danych, czy ochrony praw autorskich w sieci World Wide Web.

W **rozdziale 5** opisaliśmy praktyczne zastosowania metod Content-Based Image Retrieval w kilku dziedzinach życia publicznego. Szczególną uwagę zwróciliśmy na możliwości wykorzystania powyższych technik w medycynie diagnostycznej wymieniając kilka powodów, dla których systemy te powinny stać się integralną częścią narzędzi komputerowej analizy medycznej (Müller, Michoux i in. 2004). Dodatkowo przeanalizowaliśmy inne sfery życia społecznego takie, jak zapobieganie przestępczości, techniki wojskowe, czy projektowanie wnętrz i ubrań (Eakins, Graham 1999).

Rozdział 6 prezentuje praktyczne rozwiązania metod CBIR w telefonach komórkowych i urządzeniach PDA (ang. Personal Digital Assistant). Dotyczą one skonstruowania systemu, który będzie w stanie udzielić odpowiedzi na zapytanie użytkownika – turysty, który potrzebuje informacji na temat danego miejsca, punktu orientacyjnego, czy budynku. Spośród dostępnych źródeł wybraliśmy dwie aplikacje:

- *Snap2Tell* – służy do dostarczania informacji na temat punktów orientacyjnych Singapuru poprzez wykorzystanie specjalnie skonstruowanej do tego celu bazy danych STOIC – The Singapore Tourist Object Identification Collection,
- *Landmark-Based Pedestrian Navigation System* - głównym celem jest automatyczne generowanie wskazówek nawigacyjnych dla pieszych, które są bezpośrednio wyświetlane na ekranie urządzenia mobilnego.

W **rozdziale 7** dokonaliśmy całościowego podsumowania pracy magisterskiej. Powtórzyliśmy nadrzędne cele, które były podstawą do opisu technik Content-Based

Image Retrieval oraz zweryfikowaliśmy stopień ich realizacji. Na końcu umieściliśmy spis literatury wykorzystanej do napisania niniejszej pracy magisterskiej.

Rozdział 2. Metody wyszukiwania obrazów na podstawie zawartości

Technika wyszukiwania zdjęć na podstawie zawartości opiera się głównie na metodach dotyczących podstawowych cech obrazu, czyli jego zawartości wizualnej tzn. koloru, tekstury, kształtu czy relacji przestrzennych między obiektami. Metody te zostały szeroko rozwinięte i są obecnie stosowane w komercyjnych systemach CBIR. Zasadniczym problemem dotyczącym tej dziedziny nauki jest taka analiza obrazu, która wykorzystywałaby naturalne rozumienie jego zawartości w sposób, w jaki pojmuje ją człowiek. Stąd też w literaturze spotyka się próby wykorzystania percepcyjnego sposobu rozumowania człowieka w nowatorskich systemach komputerowych (Iqbal, Aggarwal 2002; Fu, Chi, Feng 2006).

W celu dokonania przeglądu metod stosowanych w przeszłości w pierwotnych, jak również i w późniejszych systemach CBIR niniejszy rozdział został podzielony na dwie części opisujące odpowiednio:

- metody bez interakcji - wykorzystujące deskryptory koloru, tekstury, kształtu, etc.
- metody z interakcją – wykorzystujące zagadnienia interakcji systemu komputerowego z użytkownikiem w postaci np. **sprzężenia zwrotnego**.

2.1 Metody bez interakcji

2.1.1 Metody rozpoznawania koloru

Kolor jest jedną z najważniejszych i najczęściej wykorzystywanych cech opisu obrazu używaną w systemach CBIR. Jest to narzędzi analizy, które zostało dokładnie zbadane i szeroko opisane przez wielu naukowców (Lew 2001; Del Bimbo 1999).

W celu omówienia deskryptorów koloru należy na początku określić system kolorów, czyli przestrzeń barw, w jakiej definiuje się konkretny wektor cech.

Każdy piksel obrazu może być reprezentowany poprzez trójwymiarową przestrzeń koloru. W literaturze spotyka się wiele przestrzeni barw, które posiadają odmienne zalety i wady. Systemy CBIR stosują różne przestrzenie kolorów wybierane zgodnie z indywidualnymi przesłankami.

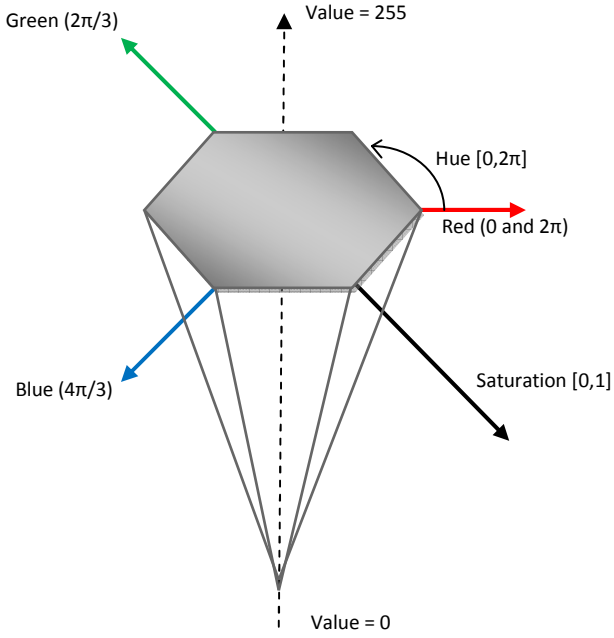
Przestrzeń RGB jest najbardziej rozpowszechnioną przestrzenią koloru. Składa się z trzech komponentów: czerwony - zielony - niebieski (ang. Red Green Blue).

Model ten nazywany jest modelem addytywnym, gdyż poszczególne barwy powstają poprzez dodawanie składowych RGB do siebie. W odróżnieniu od tej przestrzeni systemCMY (ang.Cyan Magenta Yellow) jest modelem substraktywnym, gdzie poszczególne kolory powstają przez odejmowanie składowych wyjściowego białego światła. Jest on najczęściej wykorzystywany przy drukowaniu.

Przestrzeń CIE L*a*b* lub CIE L*u*v* zaproponowana przez Huntera w 1948 r. należy do grupy przestrzeni percepcyjnie jednorodnych (ang. perceptualy uniform) tzn. różnica pomiędzy dwoma kolorami jest aproksymowana poprzez odległość między dwoma punktami przestrzeni w metryce Euklidesa (Del Bimbo 1999).

Jedną z najczęściej wykorzystywanych przestrzeni koloru w grafice komputerowej jest model HSV (HSL lub HSB). Jego nazwa to skrót od pierwszych liter wyrazów **H**ue – oznaczający barwę, **S**aturation – miarę nasycenia czyli procent koloru białego dodany do czystej barwy, **V**alue (**L**ightness lub **B**rightness) – wartość jasności czyli średnia arytmetyczna składowych RGB. Przestrzeń ta jest interpretowana jako ostrosłup, w którym (Sural 2004):

- Hue – definiuje się jako kąt w zakresie [0,2π] odpowiadający konkretnej barwie: 0 i 2π – kolor czerwony, 2π/3 – kolor zielony, 4π/3 – kolor niebieski
- Saturation – przyjmuje wartości [0,1] i zależy od odległości od środka stożka
- Value – rozumiana jako wysokość stożka (oś centralna), wartość zmienia się od 0 do 255.



Istnieje łatwe przejście pomiędzy przestrzenią HSV, a RGB zgodnie z zależnościami (Yoo, Jang, Jung i in. 2002):

$$H = \begin{cases} \theta, & dla\ G \geq B \\ 2\pi - \theta, & dla\ G < B \end{cases}$$, gdzie: $$\theta = a\cos\frac{(R-G)+(R-B)}{2\sqrt{(R-G)^2+(R-B)(G-B)}}$$

$$S = 1 - \frac{3}{R+G+B}\min(R,G,B)$$

$$V = \frac{R+G+B}{3}$$

a) momenty koloru (ang. color moments)

Z teorii prawdopodobieństwa wiadomo, że dystrybucja prawdopodobieństwa jest jednoznacznie charakteryzowana poprzez momenty centralne. Zatem jeżeli dystrybucja koloru zdjęcia zostanie zinterpretowana właśnie w taki sposób, wówczas również można posłużyć się pojęciami momentów do opisu koloru obrazu (Stricker, Orengo 1995). Można wyróżnić momenty centralne I, II i III rzędu. Matematycznie są one zdefiniowane następująco:

- moment centralny I rzędu – średnia jasność obrazu (ang. mean)

$$E_i = \frac{1}{N}\sum_{j=1}^{N} p_{ij}$$

- moment centralny II rzędu – wariancja (ang. variance)

$$\sigma_i = \left(\frac{1}{N}\sum_{j=1}^{N}\left(p_{ij} - E_i\right)^2\right)^{\frac{1}{2}}$$

- moment centralny III rzędu – współczynnik skośności, asymetria (ang. skewness)

$$s_i = \left(\frac{1}{N}\sum_{j=1}^{N}\left(p_{ij} - E_i\right)^3\right)^{\frac{1}{3}}$$

Oznaczenia:

- p_{ij} – wartość $j-tego$ piksela dla $i-tej$ barwy w danej przestrzeni,

- N - ilość pikseli obrazu.

Momenty koloru powinny być definiowane przeważnie w przestrzeni CIE L*a*b* lub CIE L*u*v* niżeli w HSV. Użycie dodatkowo trzeciego momentu centralnego zwiększa ogólną jakość wyszukiwania zdjęć w porównaniu do używania tylko pierwszych dwóch momentów. Ponieważ w przypadku stosowania trzech parametrów do opisania koloru otrzymuje się 9 liczb (3 wektory trójwymiarowe), metoda ta jest uważana za kompaktową i stosuje się ją jako pierwszą przed bardziej złożonymi algorytmami doboru deskryptora koloru (Long, Zhang, Dagan Feng 2003).

b) histogram koloru

Histogram koloru jest jednym z najczęstszych sposobów niskiego poziomu opisu deskryptora koloru. Jego najważniejszymi zaletami są: łatwość wyznaczenia, odporność na translacje i obroty obrazu, czy skalowanie. Niestety nie zawiera on żadnych informacji o strukturze przestrzennej.

Histogram może być wyznaczany dla każdej składowej przestrzeni koloru oddzielnie. Prowadzi to jednak do bardzo dużego zwiększenia wymiarowości histogramu, co z kolei czyni go bezużytecznym w procesie obliczania komputerowego. Dlatego też stosuje się metody zmniejszania wymiarowości histogramu poprzez procesy kwantyzacji.

Jedną z takich metod jest grupowanie słupków histogramu, które polega na zmniejszeniu liczby kolorów obrazu poprzez łączenie sąsiednich słupków reprezentujących podobne odcienie kolorów. W przypadku przestrzeni HSV stosowane są techniki tworzenia jednowymiarowego histogramu, który używa tylko parametrów Hue i Value opartych o parametr Saturation dla każdego piksela (Sural 2004). Innym sposobem jest metoda grupowania (ang. clustering) polegająca na pokrywaniu całego obrazu oknami o określonym rozmiarze. Następnie dla każdego okna wyznacza się średnią wartość wszystkich składowych przestrzeni barw, co daje zbiór możliwych wartości dla składowych koloru.

W procesie wykorzystania histogramu do wyszukiwania obrazów w systemach CBIR konieczne jest zdefiniowanie metryki służącej do porównania histogramów dwóch obrazów. Problem wyboru odpowiedniej metryki w przestrzeni cech został szeroko opisany przez Tadeusiewicza i Flasińskiego (1991). Przykłady obliczania odległości między histogramami są dostępne w pracy Stricker i Orengo (1995).

Metryki stosowane przy porównywaniu histogramów (Konstantinidis, Gasteratos, Andreadis 2007):

- metryka Euklidesowa

$$d_E(H_1, H_2) = \sqrt{\sum_{i=1}^{N} (H_1(i) - H_2(i))^2}$$

- metryka miejska (Manhattan)

$$d_M(H_1, H_2) = \sum_{i=1}^{N} |H_1(i) - H_2(i)|$$

- metryka cosinusowa

$$d_C(H_1, H_2) = 1 - \cos\varphi$$

$$\cos\varphi = \frac{H_1 \cdot H_2}{\|H_1\|\|H_2\|}$$

- unormowana korelacja wzajemna

$$d_X(H_1, H_2) = \frac{\sum_{i=1}^{N} H_1(i)H_2(i)}{\sum_{i=1}^{N} H_i^2}$$

- miara przekroju histogramów

$$d_I(H_1, H_2) = 1 - \frac{\sum_{i=1}^{N} \min\left(H_1(i), H_2(i)\right)}{N}$$

gdzie: H_1, H_2 – histogramy będące N wymiarowymi wektorami.

(http://www.mif.pg.gda.pl/homepages/marcin/Wyklad2.pdf).

c) wektor spójności koloru (ang. CCV - color coherence vector)

Kolejną metodą, która wykorzystuje informacje przestrzenną włączając ją do histogramu koloru, jest wektor spójności koloru (Pass, Zabith 1996). Dany słupek histogramu (składowa koloru) jest dzielony na dwa rodzaje: spójny (ang. coherent), jeżeli należy do dużego, jednorodnego pod względem koloru regionu albo niespójny (ang. incoherent), jeżeli nie należy. Oznaczając:

α_i - numer spójnego piksela w $i - tym$ słupku histogramu

β_i - numer niespójnego piksela,

wektor spójności koloru definiowany jest jako wektor $\langle(\alpha_1, \beta_1), (\alpha_2, \beta_2), \ldots, (\alpha_N, \beta_N)\rangle$, gdzie N – ilość pikseli obrazu.

Należy zwrócić uwagę, że:

$\langle \alpha_1 + \beta_1, \alpha_2 + \beta_2, \ldots, \alpha_N + \beta_N \rangle$ jest histogramem koloru obrazu.

W związku z zawieraniem informacji przestrzennej zostało udowodnione, że wektor spójności daje lepsze rezultaty wyszukiwania zdjęć w porównaniu ze zwykłym histogramem (Pass, Zabith, Miller 1997), zwłaszcza w przypadku obrazów, które posiadają głównie jednolity kolor, bądź regiony o jednolitej teksturze (Kang, Yoon, Choi i in. 2007). Zarówno w przypadku histogramu, jak i wektora spójności, przestrzeń barw HSV dostarcza lepszych rezultatów niż CIE L*a*b* lub CIE L*u*v* (Long, Zhang, Dagan Feng 2003).

d) korelogram koloru (ang. color correlogram)

Metoda budowy dekryptora cech w oparciu o korelogram koloru charakteryzuje się nie tylko różnorodnością koloru pikseli, ale także opisuje przestrzenny związek pomiędzy parami kolorów (Huang i in. 1997). Pierwszy i drugi wymiar trójwymiarowego histogramu definiuje kolory każdej pary pikseli, podczas gdy trzeci wymiar to odległość pomiędzy nimi. Korelogram koloru definiuje więc prawdopodobieństwo znalezienia piksela o j-tym kolorze w odległości k od piksela o i-tym kolorze. Matematycznie można to zapisać następująco (Long, Zhang, Dagan Feng 2003):

$$\gamma_{i,j}^{(k)} = \underset{p_1 \in I_{c(i)}, p_2 \in I}{Pr}\left[p_2 \in I_{c(j)} \middle| |p_1 - p_2| = k\right]$$

gdzie:

I – zbiór pikseli $i, j \in \{1,2,\dots,N\}$

$I_{c(i)}$- zbiór pikseli o kolorze $c(i)$ $k \in \{1,2,\dots,d\}$

$|p_1 - p_2| = k$ – odległość pomiędzy pikselami p_1i p_2.

Biorąc pod uwagę ilość wszystkich kombinacji dwóch dowolnie wybranych pikseli rozmiar korelogramu staje się bardzo duży. Stąd też używa się pojęcia autokorelogramu, który oblicza przestrzenną korelacje pomiędzy identycznymi kolorami zmniejszając efektywnie swoją wymiarowość.

W porównaniu z histogramem oraz z CCV autokorelogram daje najlepsze wyniki wyszukiwania zdjęć w systemach CBIR. Jednocześnie ze względu na swoje rozmiary jego obliczenie jest czasochłonne i znacząco obciąża komputer.

Wymienione powyżej deskryptory koloru są najbardziej popularnymi wektorami cech stosowanymi w systemach CBIR. Należy zwrócić uwagę, że popularność stosowania koloru jako parametru przy analizie obrazu wiąże się bezpośrednio z jego następującymi cechami (Lew 2001):

- łatwość implementacji algorytmów,
- liniowość koloru,
- intuicyjność,
- niezmienniczość względem translacji, obrotu, skalowania obrazu,
- niezmienniczość względem zmian geometrii obiektu,
- odporność na zmienne warunki wykonywania obrazów.

2.1.2 Metody rozpoznawania tekstury

Pojęcie tekstury jest trudne do sformułowania, dlatego też w literaturze wyróżnia się kilka odmiennych definicji (Sebe, Lew 2001). Ogólnie można przyjąć, że jest to charakterystyczny dla danego materiału wzór na powierzchni przedmiotu. (http://pl.wikipedia.org/wiki/Tekstura).

Pomimo trudności w zdefiniowaniu tekstury niewątpliwie jest ona jedną z najważniejszych cech używaną w rozpoznawaniu i przetwarzaniu obrazów. Literatura wyróżnia trzy zasadnicze podejścia przy analizie tekstury (Haralick, Shanmugam, Dinstein 1979; Sebe, Lew 2001):

- statystyczne – tekstura reprezentowana jest poprzez zbiór cech, który określa np. kontrast, entropie, korelacje. Przegląd tych metod jest dostępny w pracy Haralicka (1979),
- stochastyczne – tekstura uważana jest jako realizacja stochastycznego procesu cechującego się określonymi parametrami. Analiza polega na definicji modelu i oszacowaniu parametrów, które następnie mogą służyć jako cechy tekstury (Haindl 1991),
- strukturalne – tekstura jest uważana jako dwuwymiarowy model składający się ze zbioru podstawowych wzorów o znanym rozmieszczeniu. Na te wzory składają się mikro- i makrotekstury o znanych kształtach. Obraz tekstury jest budowany przy użyciu podstawowych wzorów z wykorzystaniem reguł określających ich lokalizację i wzajemną orientację (Haralick 1979; Haindl 1991).

a) cechy teksturowe Tamury

Metoda ta opiera się na analizie sześciu cech opisu tekstury (Tamura, Mori, Yamawaki 1978), które zostały dobrane odpowiednio do sposobu ich postrzegania przez człowieka. Są to: gruboziarnistość (ang. coarseness), kontrast (ang. contrast), kierunkowość (ang. directionality), liniowe podobieństwo (ang. linelikeness), regularność (ang. regularity), chropowatość (ang. roughness). Jednak tylko pierwsze trzy są używane w systemach CBIR (np. QBIC, Photobook):

- gruboziarnistość

Cecha ta daje informację o rozmiarze elementów tekstury. Jej sposób wyznaczenia jest następujący:

- dla każdej pary pikseli (x, y) oblicza się średnią po wszystkich sąsiadach używając do tego okien o rozmiarze $2^k \times 2^k$, dla $k = 0, 1, 2, 3, 4, 5$:

$$A_k(x,y) = \sum_{i=x-2^{k-1}}^{x+2^{k-1}-1} \sum_{j=y-2^{k-1}}^{y+2^{k-1}-1} g(i,j)/2^{2k}$$

gdzie: $g(i,j)$ - intensywność (natężenie) piksela (i,j).

- dla każdego punktu wyznacza się różnicę pomiędzy niezachodzącymi na siebie średnimi w kierunku poziomym i pionowym tzn.

$$E_{k,h}(x,y) = |A_k(x+2^{k-1},y) - A_k(x-2^{k-1},y)|$$

$$E_{k,v}(x,y) = |A_k(x,y+2^{k-1}) - A_k(x,y-2^{k-1})|$$

- numer k maksymalizujący powyższą różnicę w dowolnym kierunku jest wybierany, jako najlepszy rozmiar dla każdego piksela

$$S_{best}(x,y) = 2^k$$

- gruboziarnistość jest wyznaczana jako uśrednianie obrazu parametrem S_{best}:

$$F_{crs} = \frac{1}{m \times n} \sum_{i=1}^{m} \sum_{j=1}^{n} S_{best}(i,j)$$

gdzie: $m \times n$ - rozmiar zdjęcia podany w pikselach.

Czasami zamiast średniej S_{best} do wyznaczenia gruboziarnistości używa się histogramu, który charakteryzuje dystrybucje tego parametru. Taka reprezentacja znacznie zwiększa jakość wyszukiwania (Long, Zhang, Dagan Feng 2003).

- kontrast

W ścisłym znaczeniu kontrast to inaczej jakość obrazu. Używając bardziej szczegółowej analizy, na jego wpływ mogą mieć następujące czynniki (Tamura, Mori, Yamawaki 1978):

- ✓ skala szarości zdjęcia,
- ✓ polaryzacja dystrybucji koloru białego i czarnego na histogramie,
- ✓ ostrość krawędzi,
- ✓ częstotliwość powtarzających się wzorów.

Kontrast zdjęcia może być obliczany z zależności: $F_{con} = \frac{\sigma}{\alpha_4^z}, \ \alpha_4 = \frac{\mu_4}{\sigma_4}$

gdzie: μ_4 – moment centralny IV rzędu, σ^2 - wariancja.

Taka definicja pozwala używać kontrastu w obrębie całego obrazu, jak również lokalnie.

- kierunkowość

W celu obliczenia kierunkowości obraz poddaje się filtracji dwoma maskami 3×3 i wyznacza wektor gradientu:

$$\begin{bmatrix} -1 & 0 & 1 \\ -1 & 0 & 1 \\ -1 & 0 & 1 \end{bmatrix} i \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$$

Następnie określa się rząd wielkości i kąt wektora:

$$|\Delta G| = (|\Delta_H| + |\Delta_V|)/2$$

$$\theta = \tan^{-1}(\Delta_V/\Delta_H) + \frac{\pi}{2}$$

gdzie: Δ_V i Δ_H są pionowymi i poziomymi różnicami filtracji.

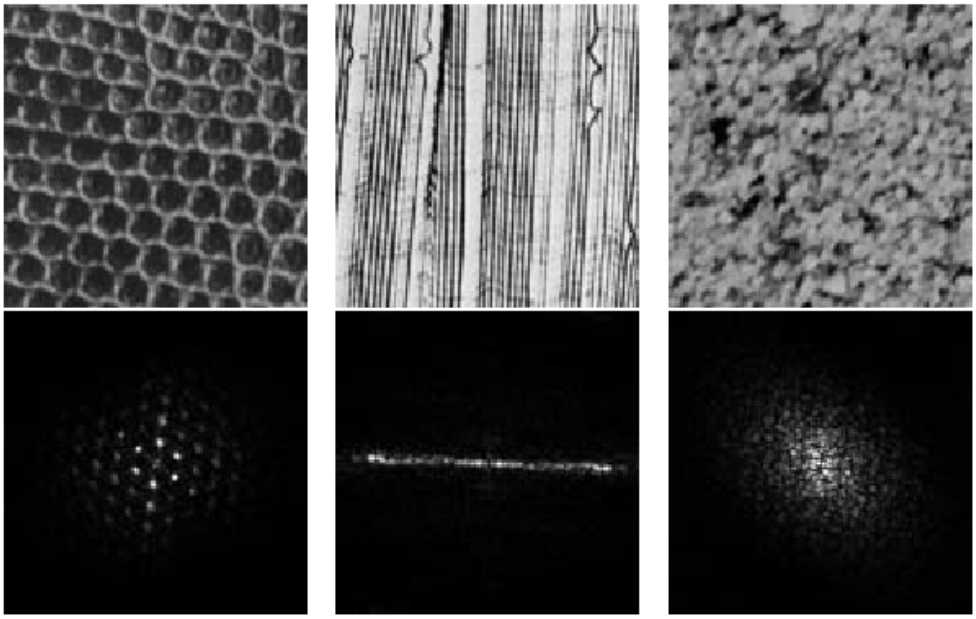
Kolejnym etapem jest wyznaczenie histogramu rozkładu kąta $\theta\,(H_D)$, który charakteryzuje się ostrymi pikami w miejscach silnej kierunkowości obrazu i niskim poziomem w pozostałej części. Wielkość kierunkowości może być wyznaczona jako miara ostrości pików histogramu:

$$F_{dir} = \sum_{p}^{n_p} \sum_{\varphi \epsilon w_p} (\varphi - \varphi_{p)}{}^2 H_D(\varphi)$$

gdzie: n_p - liczba pików histogramu, w_p - zbiór słupków tworzących dany pik, φ_p - słupek przyjmujący wartość danego piku (Long, Zhang, Dagan Feng 2003).

b) dekompozycja Wold'a

Kolejna propozycja opisu tekstury bazująca na cechach percepcyjnego sposobu rozumowania człowieka została przedstawiona w pracy Francos (1993). Opiera się ona na trzech komponentach Wold'a: harmonicznym, krótkotrwałym (zanikającym, ang. evanescent) i interdeterministycznym, które zostały przedstawione przez Liu i Picard (1996) odpowiednio jako periodyczność, kierunkowość i przypadkowość w odniesieniu do cech tekstury.



Rys. 2.1 Charakterystyka komponentów Wold'a (od lewej): składowa harmoniczna, zanikająca i interdeterministyczna. Poniżej każdej tekstury znajduje się obraz jej dyskretnej transformaty Fouriera DFT (źródło: Liu, Picard 1996).

Zgodnie z teorią przedstawioną przez powyższych naukowców dla homogenicznego i regularnego obszaru $\{y(m,n),(m,n)\epsilon Z^2\}$ dekompozycja Wold'a pozwala na zdefiniowanie obszaru jako sumy trzech ortogonalnych składników:

$$y(m,n) = u(m,n) + d(m,n) = u(m,n) + h(m,n) + e(m,n)$$

gdzie:

- $u(m,n)$ - składnik interdeterministyczny,

- $d(m,n)$ - składnik deterministyczny, który można podzielić na składnik harmoniczny $h(m,n)$ i zanikający $e(m,n)$.

W dziedzinie częstotliwości można podać analogiczne równanie:

$$F_y(\xi,\eta) = F_u(\xi,\eta) + F_d(\xi,\eta) = F_u(\xi,\eta) + F_h(\xi,\eta) + F_e(\xi,\eta)$$

przy czym $F_y(\xi,\eta), F_u(\xi,\eta), F_d(\xi,\eta), F_h(\xi,\eta), F_e(\xi,\eta)$ są widmowymi dystrybucjami (ang. spectral distribution functions SPD) odpowiednich członów $\{y(m,n)\}$, $\{u(m,n)\}$, $\{d(m,n)\}$, $\{h(m,n)\}$, $\{e(m,n)\}$.

Model oparty o składniki Wold'a może zostać zbudowany poprzez rozkład obrazu na trzy zasadnicze komponenty. W literaturze wyróżnia się dwie metody dekompozycji (Sebe, Lew 2001). Pierwsza oparta jest na estymacji maksymalnego

prawdopodobieństwa (ang. maximum likelihood estimation MLE) i dostarcza parametrycznego opisu zdjęcia (Francos, Narasimhan, Woods 1996). Druga natomiast to dekompozycja spektralna polegająca na wyborze częstotliwości Fouriera wyższych od założonego progu, jako składowych harmonicznych lub zanikających (Francos, Meiri, Porat 1993).

c) filtr Gabora

Technika analizy cech tekstury oparta na filtrach Gabora została szeroko rozwinięta i rozpowszechniona przez wielu naukowców (Turner 1986; Clark, Bovik 1987; Carreira, Orwell i in. 1998; Setchell, Campbell 1999). Wywodzi się z transformaty Fouriera i bazuje na analizie częstotliwości występujących na danym zdjęciu. Zwykła transformata Fouriera nie zachowuje żadnych informacji przestrzennych obrazu. Dlatego też początkowe podejście polega na wykorzystaniu okienkowej transformaty Fouriera (ang. windowed Fourier transform WFT). W przypadku, gdy funkcja okna jest Gaussowska otrzymuje się transformatę Gabora, która minimalizuje zasadę niepewności częstotliwości (Sebe, Lew 2001).

Transformata Gabora jest także używana do detekcji krawędzi oraz linii. Jej podstawową zaletą jest intuicyjność: im drobniejsza tekstura, tym wyższe częstotliwości widma (Rotter 2003).

Dwuwymiarowa funkcja Gabora:

$$g(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)+2\pi j f_c x\right]$$

Transformata Fouriera powyższej funkcji Gabora ma postać:

$$G(u,v) = exp\left[-\frac{1}{2}\left(\frac{(u-f_c)^2}{\sigma_u^2}+\frac{v^2}{\sigma_v^2}\right)\right]$$

gdzie:

- σ_x, σ_y - odchylenia standardowe (określają szerokość pasma przepustowego filtru),

- f_c - centralna częstotliwość filtru,

- $\sigma_u = \frac{1}{2\pi\sigma_x}, \sigma_v = \frac{1}{2\pi\sigma_y}$ - odchylenia standardowe w dziedzinie częstotliwości.

W celu użycia transformaty Gabora do określania cech tekstury konieczne jest zaprojektowanie odpowiedniego zespołu filtrów (Ma, Manjunath 1997; Sebe, Lew 2001). Otrzymuje się go poprzez skalowanie i obrót funkcji $g(x,y)$ (Rotter 2003):

$$g_{mn}(x,y) = a^{-m} g(x',y'), \;\; m = 0,1,\dots,S-1$$

$$x' = a^{-m}(x\cos\theta + y\sin\theta), \quad y' = a^{-m}(-x\sin\theta + y\cos\theta)$$

gdzie:

- $\theta = \frac{n\pi}{K}$, K – liczba orientacji, S – liczba skali

- $n = 0,1, \dots, K-1 \; i \; m = 0,1, \dots, S-1, a > 1$ - energia niezależna od parametru *m*.

d) transformata falkowa

Metoda wyznaczania wektora cech na podstawie analizy falkowej (Daubechies 1990) jest koncepcyjnie podobna do transformaty Gabora. Bazuje na dekompozycji podstawowego sygnału przy użyciu funkcji bazowych, które otrzymuje się z podstawowej funkcji falkowej (Long, Zhang, Dagan Feng 2003).

Transformata falkowa, analogicznie do transformaty Gabora, pozwala na obejście ograniczeń rozdzielczości okna Fouriera poprzez skalowanie funkcji i jej odpowiednie przesuwanie. Jej definicja jest następująca (www.falki.prv.pl):

$$\tilde{S}_{\psi}(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} s(t) \psi\left(\frac{t-b}{a}\right) dt$$

gdzie:

- a - współczynnik skali, b - współczynnik przesunięcia,

- $s(t)$ - zależny od czasu sygnał badany,

- $\tilde{S}_{\psi}(a,b)$ – współczynnik falkowy zależny od parametrów a i b,

- ψ – funkcja falkowa,

- $\psi\left(\frac{t-b}{a}\right)$ – jądro przekształcenia.

Wektor cech do analizy tekstury otrzymuje się obliczając średnią i wariancję dystrybucji energii współczynników transformaty (Smith, Chang 1996).

Metody analizy tekstury zostały pomyślnie zaadoptowane do systemów wyszukiwania obrazów na podstawie zawartości. W praktyce istnieją dwa podejścia do stosowania tekstury, jako nadrzędnej cechy dla systemów CBIR. Pierwsze polega na formułowaniu zapytania poprzez wybór tych zdjęć z określonego zbioru, które posiadają żądaną teksturę. Następnie system zwraca użytkownikowi zdjęcia, które są najbardziej zbliżone do wzoru w sensie wartości miary podobieństwa. Drugi sposób to wykorzystanie tekstury jako narzędzia do etykietowania obrazów. Następnie użytkownik jest proszony o opisanie danego fragmentu obrazu, a ta informacja jest wykorzystywana przez system do wyszukania zdjęć o podobnej teksturze (Sebe, Lew 2001).

2.1.3 Metody rozpoznawania kształtu

Analiza kształtu jest jedną z podstawowych technik reprezentacji obrazu wykorzystywanych w systemach wyszukiwania zdjęć. Dowodem tego jest mnogość pozycji naukowych omawiających podstawowe i bardziej skomplikowane metody doboru odpowiedniego wektora bazującego na cechach kształtu (Veltkamp, Hagedoorn 1999; Loncaric 1998; Wong, Shih, Liu 2007; Yadav, Nishchal i in. 2007). Zasadniczo techniki wykorzystujące deskryptory kształtu można podzielić na dwa rodzaje (Long, Zhang, Dagan Feng 2003):

- boundary-based methods – czyli metody oparte na analizie granicy (obwiedni) kształtu np. aproksymacje wielokątne, metody elementów skończonych, deskryptory kształtu Fouriera,
- region-based methods – czyli metody analizujące poszczególne regiony obrazu np. momenty geometryczne.

Niezależnie od wyboru metody, reprezentacja kształtu musi być inwariantna względem translacji, obrotu i skalowania. Dopiero, gdy warunek ten będzie spełniony deskryptor kształtu można uznać za poprawnie wyznaczony.

a) momenty

Metoda analizy kształtu oparta na momentach geometrycznych została szeroko opisana w pracy Prokopa, Nishchala i in. (1992). Momenty zostały rozpowszechnione w statystyce, gdzie służyły do opisu dystrybucji zmiennych, oraz w mechanice, gdzie wykorzystywano je do określania przestrzennego rozmieszczenia masy. Ich użycie do analizy obrazu wydaje się oczywiste w momencie przedstawienia obrazu w skali szarości, jako dwuwymiarowej funkcji gęstości. Wówczas analiza momentów może odbywać się poprzez analogię do statystyki i mechaniki.

- kartezjańska definicja momentu (Loncaric 1998):

Moment m_{pq} rzędu $p+q$ dla funkcji gęstości $f(x,y)$ definiuje się następująco:

$$m_{pq} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^p y^q f(x,y) dx dy$$

W wersji dyskretnej dla obrazu o rozmiarze $N \times M$ pikseli ($g(x,y)$):

$$m_{pq} = \sum_{y=0}^{M-1} \sum_{x=0}^{N-1} x^p y^q g(x,y)$$

Użycie momentów do analizy obrazu i reprezentacji obiektu zostało zapoczątkowane przez naukowca Hu (1962). Jego teoria jednoznaczności dowodzi, że zbiór momentów

$\{m_{pq}\}$ jest jednoznacznie zdefiniowany przez funkcję gęstości $f(x,y)$ i odwrotnie. Dodatkowo zbiór ten zawiera informacje o kształcie obiektu.

Analiza momentów zarówno niższego, jak i wyższego rzędu prowadzi do podania podstawowych, geometrycznych cech obiektu, które wykorzystuje się w technikach przetwarzania obrazu (Prokop, Reeves 1992):

- momenty zerowego rzędu – obszar

Definicja momentu zerowego rzędu dla funkcji gęstości $f(x,y)$ jest następująca:

$$m_{00} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y)dxdy$$

Moment ten reprezentuje całkowitą masę dystrybucji, a w przypadku obrazu całkowity obszar obiektu.

- momenty pierwszego rzędu – środek masy

Momenty te: $\{m_{10}, m_{01}\}$ są używane do lokalizacji środka masy obiektu. Współrzędne środka masy są obliczane ze znanych zależności:

$$\bar{x} = \frac{m_{10}}{m_{00}} \qquad \bar{y} = \frac{m_{01}}{m_{00}}$$

Jeżeli obiekt jest umiejscowiony tak, że jego środek masy pokrywa się ze środkiem układu współrzędnych tzn. $\bar{x} = 0 \; i \; \bar{y} = 0$, wówczas momenty obliczane dla tego obiektu są nazywane *momentami centralnymi* i oznaczane przez μ_{pq}.

- momenty drugiego rzędu – momenty bezwładności

Momenty te są używane do określenia następujących cech:

✓ *główne osie obiektu* – są to osie, powyżej których znajduje się maksymalny i minimalny moment drugiego rzędu. Orientację tych osi wylicza się z zależności:

$$\varphi = \frac{1}{2}\tan^{-1}\left(\frac{2\mu_{11}}{\mu_{20} - \mu_{02}}\right)$$

gdzie: $\frac{-\pi}{4} \leq \varphi \leq \frac{\pi}{4}$

✓ *elipsa obrazu* – momenty pierwszego i drugiego rzędu definiują również inercjalnie równoważną aproksymację obrazu (Teague 1980). Elipsa obrazu rozumiana jest jako eliptyczny dysk z tą samą masą i momentami drugiego rzędu, jak oryginalny obraz. Jeżeli zdefiniować elipsę przy użyciu dodatkowego układu współrzędnych: α – wzdłuż osi x i β – wzdłuż osi y:

$$\alpha = \left(\frac{2\left[\mu_{20} + \mu_{02} + \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}\right]}{\mu_{00}} \right)^{\frac{1}{2}}$$

$$\beta = \left(\frac{2\left[\mu_{20} + \mu_{02} - \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}\right]}{\mu_{00}} \right)^{\frac{1}{2}}$$

wówczas intensywność elipsy obrazu jest rozumiana, jako:

$$I = \frac{\mu_{00}}{\pi\alpha\beta}$$

✓ *promień bezwładności (ang. radii of gyration)* – określa odległość od osi punktu, w którym skupienie całkowitej masy ciała powoduje równowagę pomiędzy momentem bezwładności punktu materialnego, a całego obiektu. Promień ten można wyznaczyć ze wzorów:

$$r_x = \sqrt{\frac{m_{20}}{m_{00}}} \qquad r_y = \sqrt{\frac{m_{02}}{m_{00}}}$$

Jego podstawową zaletą jest inwariantność względem orientacji obrazu, dlatego też jest używany do reprezentacji obiektu.

- momenty trzeciego rzędu – asymetria (współczynnik skośności)

Momenty rzędów trzeciego i większych są używane bardziej do opisu cech rzutu zdjęcia na osie x, y niżeli do opisu samego zdjęcia.

Asymetrię rozkładu oblicza się ze wzorów:

$$sk_x = \frac{\mu_{30}}{\mu_{20}^{3/2}} \qquad sk_y = \frac{\mu_{03}}{\mu_{02}^{3/2}}$$

Jest ona używana do jednoznacznego określania kierunku obrotu obrazu.

- momenty czwartego rzędu – kurtoza

Jest to jedna z miar spłaszczenia rozkładu wartości cechy. Oblicza się ją ze wzorów:

$$k_x = \frac{\mu_{40}}{\mu_{20}^2} - 3 \qquad k_y = \frac{\mu_{04}}{\mu_{02}^2} - 3$$

Kurtoza rozkładu normalnego ma wartość 0.

- niezmienniki momentowe

Bezpośrednie użycie kartezjańskiej definicji momentu do budowania deskryptora kształtu nie jest stosowane, ponieważ momenty te ulegają zmianom pod wpływem przekształceń geometrycznych. Stąd też Hu (1962) zaproponował siedem niezmienników momentowych, które są inwariantne względem translacji, zmian orientacji, czy skalowania:

$$M_1 = \mu_{20} + \mu_{02}$$

$$M_2 = (\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2$$

$$M_3 = (\mu_{30} - 3\mu_{12})^2 + (3\mu_{21} - \mu_{03})^2$$

$$M_4 = (\mu_{30} + \mu_{12})^2 + (\mu_{21} + \mu_{03})^2$$

$$M_5 = (\mu_{30} - 3\mu_{12})(\mu_{30} + \mu_{12})[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] + (3\mu_{21} - \mu_{03})(\mu_{21} + \mu_{03})[3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2]$$

$$M_6 = (\mu_{20} - \mu_{02})[(\mu_{30} + \mu_{12})^2 - (\mu_{03} + \mu_{21})^2] + 4\mu_{11}(\mu_{30} + \mu_{12})(\mu_{03} + \mu_{21})$$

$$M_7 = (3\mu_{21} - \mu_{03})(\mu_{30} + \mu_{12})[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] - (\mu_{30} - 3\mu_{12})(\mu_{21} + \mu_{03})[3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2]$$

Deskryptor kształtu budowany jest jako wektor siedmioelementowy, który ze względu na swoją małą wymiarowość i niezmienniczość względem przekształceń geometrycznych bardzo dobrze nadaje się do pomiaru podobieństwa pomiędzy kształtami (Sebe, Lew 2002).

b) metoda obracanego kąta (ang. turning angle, turning function)

Metoda ta bazuje na reprezentacji kształtu obiektu jako zamkniętej sekwencji pikseli (x_S, y_S) jego obwiedni, przy czym $0 \leq s \leq N - 1$ i N - ilość wszystkich pikseli tworzących krawędź obiektu (Otterloo 1992; Veltkamp, Hagedoorn 1999). Technika obracającego się kąta $\theta(s)$ polega na pomiarze tangensa kąta przeciwnego do ruchu wskazówek zegara, jako funkcji długości łuku s zgodnie z punktem odniesienia konturu obiektu. Jej definicja jest następująca:

$$\theta(s) = \tan^{-1}\left(\frac{y_s'}{x_s'}\right)$$

$$y_s' = \frac{dy_s}{ds} \qquad x_s' = \frac{dx_s}{ds}$$

Głównym problemem przy takiej reprezentacji kształtu obiektu jest jej zależność względem rotacji obiektu i wyboru punktu odniesienia. Przesunięcie punktu odniesienia o dowolną wielkość t powoduje powstanie nowej funkcji $\theta(s + t)$. Z kolei obrót obiektu o parametr ω daje funkcję $\theta(s) + \omega$. W związku z tym w celu porównania

kształtu dwóch obiektów A i B przy użyciu metody obracanego kąta należy obliczyć minimalną odległość dla każdego możliwego przesunięcia *t* i obrotu ω (Long, Zhang, Dagan Feng 2003):

$$d_P(A,B) = \left(\min_{\omega \in R, t \in [0,1]} \int_0^1 |\theta_A(s+t) - \theta_B(s) + \omega|^P ds \right)^{\frac{1}{P}}$$

Przy wykorzystaniu powyższej zależności zakłada się , że każdy obiekt został przeskalowany tak, aby długość jego obwodu była równa jeden. Taka miara jest wówczas inwariantna względem obrotu i skalowania.

c) metody aktywnego konturu

Aktywne kontury są definiowane jako energia minimalizująca splajny (funkcje sklejane), które podlegają siłom wewnętrznym i zewnętrznym (Kass, Witkin, Terzopoulos 1988). Siły wewnętrzne (elastyczne) mają za zadanie ściśle trzymać kontur i zapobiegać jego ugięciom. Siły zewnętrzne mają natomiast prowadzić kontur w kierunku cech obrazu np. jego wysokiej intensywności. Tak zdefiniowany kontur może być rozumiany, jako równowaga pomiędzy gładkimi cechami geometrycznymi, a lokalną zgodnością z intensywnością obrazu.

Jeżeli aktywny kontur będzie zdefiniowany poprzez parametryczną reprezentację $v(s) = (x(s), y(s))$, gdzie s - znormalizowana długość łuku, wówczas jego energia całkowita ma postać (Sebe, Lew 2002):

$$E_{total} = \int_0^1 \left[E_{int}\big(v(s)\big) + E_{image}\big(v(s)\big) + E_{con}\big(v(s)\big) \right] ds$$

przy czym:

- $E_{int}\big(v(s)\big)$ - energia wewnętrzna określająca gładkie krzywe,

- $E_{image}\big(v(s)\big)$ – energia zdjęcia - lokalna zgodność z funkcją obrazu,

- $E_{con}\big(v(s)\big)$ - energia ograniczająca, pomijalna.

- energia wewnętrzna – ma minimalizować krzywiznę i sprawiać, że aktywny kontur będzie elastyczny. Jej definicja jest następująca (Kass i in. 1988):

$$E_{int}\big(v(s)\big) = \alpha(s) \left| \frac{\mathrm{dv(s)}}{\mathrm{ds}} \right|^2 + \beta(s) \left| \frac{\mathrm{d^2 v(s)}}{\mathrm{ds^2}} \right|^2$$

  - $\alpha(s)$ - współczynnik odpowiadający za elastyczność konturu,

  - $\beta(s)$ - współczynnik odpowiadający za odporność na zgięcie.

- energia zdjęcia – pojęcie to określa zdolność przyciągania aktywnego konturu do takich punktów obrazu, jak linie czy krawędzie, inaczej do miejsc o wysokim gradiencie:

$$E_{edge} = -|\nabla I(x, y)|$$

Metody aktywnego konturu w swojej pierwotnej postaci miały trzy zasadnicze wady: zależność od wyboru początkowego konturu, numeryczną niestabilność i brak zbieżności do globalnego minimum energii. Część z tych wad została usunięta m.in. przez Amini (1988), który ulepszył numeryczną niestabilność poprzez minimalizację funkcjonału energii przy użyciu dynamicznego programowania.

d) deskryptory Fouriera (ang. Fourier Descriptors)

Deskryptory Fouriera (FD) są jedną z najbardziej popularnych metod opisu wektora cech, która została pomyślnie zaadoptowana do rozwiązywania wielu problemów reprezentacji kształtu (Zahn, Roskies 1972; Zhang, Lu 2003; Yadav, Nishchal i in. 2007). Swoją popularność zawdzięcza ona m.in. łatwości obliczenia, normalizacji i odporności na zakłócenia (Yadav, Nishchal i in. 2007).

Ogólnie deskryptory Fouriera są otrzymywane poprzez poddanie transformacie Fouriera pewnych funkcji opisujących obwiednie kształtu (ang. shape signatures). Wśród tych funkcji wyróżnia się: współrzędne zespolone (ang. complex coordinates), odległość od środka ciężkości (ang. centroid distance), czy funkcje krzywizny (ang. curvature function) . Jednak deskryptory Fouriera wyznaczone przy użyciu powyższych funkcji opisujących obwiednie kształtu różnią się znacząco w kontekście wyszukiwania obrazów. Zostało udowodnione, że deskryptor FD uzyskany z odległości od środka ciężkości jest lepszy, w ogólnym znaczeniu, w porównaniu do innych funkcji obwiedni (Zhang, Lu 2001).

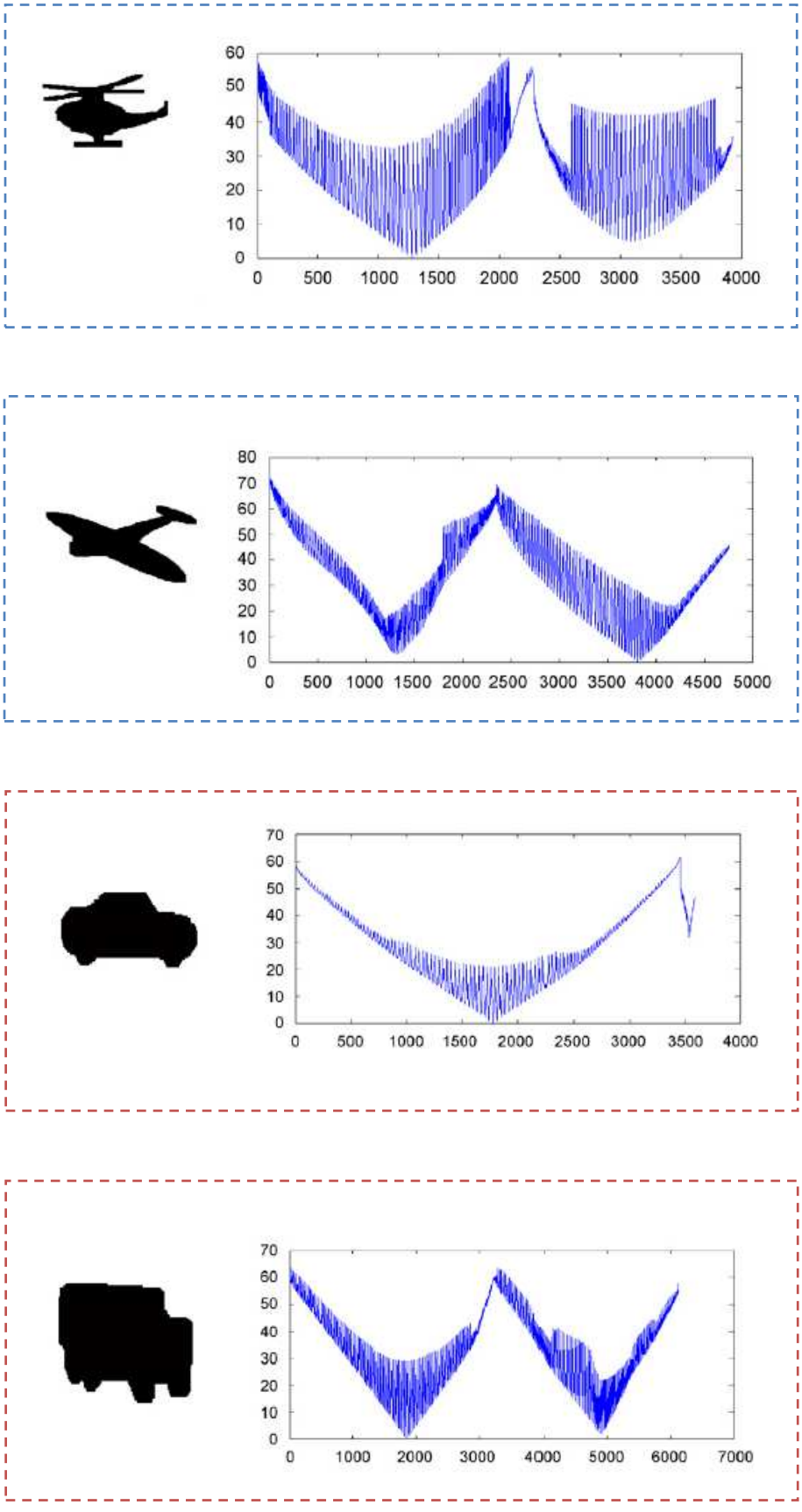
Pierwszym etapem w metodzie wyznaczenia FD jest opisanie obwiedni kształtu przy użyciu jej współrzędnych: $(x(t), y(t)), t = 0,1, \dots, N-1$, gdzie N - ilość punktów tworzących obwiednie kształtu. Funkcja określająca odległość od środka ciężkości jest wyznaczana, jako odległość punktów obwiedni od środka kształtu (x_c, y_c):

$$r(t) = ([x(t) - x_c]^2 + [y(t) - y_c]^2)^{\frac{1}{2}}, \qquad t = 0,1, \dots, N-1$$

gdzie:

$$x_c = \frac{1}{N}\sum_{t=0}^{N-1} x(t) \qquad y_c = \frac{1}{N}\sum_{t=0}^{N-1} y(t)$$

Przykłady obrazów różnych klas wraz z wykresami funkcji odległości od środka ciężkości:



Rys. 2.2 Obrazy różnych klas wraz z ich wykresami funkcji odległości od środka ciężkości (źródło: Yadav, Nishchal i in. 2007).

Dyskretna transformata Fouriera dla funkcji $r(t)$ (Gonzales, Woods 1993; Kunttu, Lepisto i in. 2006):

$$a_n = \frac{1}{N}\sum_{t=0}^{N-1} r(t) exp\left(\frac{-j2\pi nt}{N}\right), \qquad n = 0,1,\dots,N-1$$

gdzie a_n - współczynniki transformaty.

Wyznaczone współczynniki są niezmiennicze względem translacji w związku z faktem, że funkcja obwiedni jest także translacyjnie niezmiennicza. W celu opisu kształtu, współczynniki transformaty, będące faktycznym deskryptorem, muszą być znormalizowane tak, aby były niezależne od obrotu, skalowania, czy doboru punktu startowego. Postać współczynników konturu $r(t)$ niezależnych względem przekształceń geometrycznych została podana przez Granlund'a (1972):

$$a_n = exp(jn\tau) \cdot exp(j\varphi) \cdot s \cdot a_n^{(0)}$$

gdzie $a_n \; i \; a_n^{(0)}$ - współczynniki Fouriera odpowiednio po transformacji i przed, $\tau \; i \; \varphi$ - kąty wynikające odpowiednio ze zmiany punktu startowego i obrotu, s - współczynnik skali. Analiza wyrażenia (Zhang, Lu 2003):

$$b_n = \frac{a_n}{a_0} = \frac{exp(jn\tau) \cdot exp(j\varphi) \cdot s \cdot a_n^0}{exp(jn\tau) \cdot exp(j\varphi) \cdot s \cdot a_n^1} = \frac{a_n^{(0)}}{a_0^{(0)}} = b_n^{(0)} exp[j(n-1)\tau]$$

gdzie: $b_n \; i \; b_n^{(0)}$ – znormalizowane współczynniki Fouriera odpowiednio po i przed transformacją kształtu. Różnica między nimi występuje wyłącznie w wyrażeniu

exp $[j(n-1)\tau]$. Jeżeli zignorować informację o fazie i użyć jedynie informacji o wielkości współczynników, wówczas $|b_n|$ i $\left|b_n^{(0)}\right|$ są takie same. Inaczej mówiąc $|b_n|$ jest inwariantny względem translacji, obrotu, skalowania i wyboru punktu startowego. Zbiór współczynników kształtu transformaty Fouriera: $\{|b_n|, 0 < n \leq N\}$ jest zatem deskryptorem kształtu, który oznacza się $\{FD_n, 0 < n \leq N\}$. Ponieważ odległość od środka ciężkości jest liczbą rzeczywistą, zatem używa się tylko połowy współczynników FD do opisu danego kształtu. Do obliczenia podobieństwa pomiędzy przykładowym obrazem Q, a szukanym T używa się miary Euklidesa:

$$d = \left(\sum_{i=1}^{N/2} \left|FD_i^Q - FD_i^T\right|^2\right)^{1/2}$$

Zhang i Lu (2001, 2002, 2003) przeprowadzili eksperyment, w którym dowiedli, że efektywna liczba współczynników wystarczająca do opisu cech kształtu wynosi 10. Skuteczność wyszukiwania nie zmieniała się znacząco przy użyciu 15, 30, 60 lub 90 współczynników.

e) metody metryczne pomiaru podobieństwa – określanie odległości

Jednym z etapów w procesie wyszukiwania obrazów na podstawie zawartości jest wyznaczanie podobieństwa pomiędzy wzorcowym obrazem, a wynikiem wyszukiwania. W tym celu zamiast procesu dopasowania obiektowego można stosować metody metryczne pomiaru podobieństwa między obrazami. W systemach CBIR stosowanych jest wiele różnych metryk, jednak nie można wskazać najlepszej z nich. Każda ma swoje wady i zalety, a jej skuteczność w dużej mierze zależy od cech charakterystycznych obrazu.

Aby dana odległość była metryką określoną na pewnym niepustym zbiorze X, musi spełniać następujące założenia (Veltkamp, Hagedoorn 1999):

- dla dowolnych elementów $a, b, c \in X$ funkcja $d: X \times X \to R$ jest metryką, gdy:
- $d(a,b) \geq 0,\ d(a,b) = 0\ <=>\ a = b$
- $d(a,b) = d(b,a)$ - warunek symetrii
- $d(a,b) + d(b,c) \geq d(a,c)$ - nierówność trójkąta.

Wówczas parę (X, d) nazywa się przestrzenią metryczną.

Wprowadźmy oznaczenia:

- $d(I,J)$ – odległość pomiędzy wzorcowym zdjęciem I, a zdjęciem wynikowym J,

- $f_i(I)$ - liczba pikseli $i - tego$ słupka w obrazie I.

- odległość Minkowskiego

Metryka ta nadaje się do obliczania podobieństwa przy założeniu ortogonalności bazy przestrzeni cech danego deskryptora obrazu (Tadeusiewicz, Flasiński 1991). Jest to dosyć silne założenie, które jednak w większości przypadków nie jest spełnione.

Wzór jest następujący:

$$d(I,J) = \left(\sum_i |f_i(I) - f_i(J)|^p \right)^{1/p}$$

gdzie: $p = 1,2, \ldots, \infty$. Dla $p = 1$ otrzymuje się metrykę miejską (patrz rozdział 2.1.1 b), dla $p = 2$ metrykę Euklidesową, zaś dla $p \to \infty$ wzór przyjmuje postać odległości Czebyszewa. Pomimo czasochłonności obliczeniowej metryka ta znalazła szerokie zastosowanie w popularnych systemach CBIR (Long, Zhang, Dagan Feng 2003):

- ✓ system MARS używa metryki Euklidesowej do obliczania podobieństwa między deskryptorami tekstury (Rui, Mehrotra 1997),

- ✓ system Netra wykorzystuje metrykę Euklidesową do wyznaczania odległości między wektorami koloru i kształtu, oraz metrykę miejską w przypadku tekstury (Ma, Manjunath 1999),
- ✓ system Blobworld – metryka Euklidesowa dla tekstury i kształtu (Carson, Thomas, Belongie i in. 1999),
- ✓ metryka Czebyszewa została użyta przez Vorhess i Poggio (1998) do obliczania podobieństwa teksturowego.

- formy kwadratowe

Przy obliczaniu odległości Minkowskiego przyjmuje się, że wszystkie słupki histogramu są całkowicie niezależne, nie biorąc pod uwagę faktu, że niektóre z nich określają cechy percepcyjnie bardziej zbliżone do innych. W celu rozwiązania tego problemu stosuje się metrykę (formę) kwadratową postaci:

$$d(I,J) = \sqrt{\left(F_I - F_J\right)^T A \left(F_I - F_J\right)}$$

gdzie $A = \left[a_{ij}\right]$ - macierz podobieństwa między słupkami i oraz j histogramu, F_I oraz F_J - wektory o elementach odpowiednio $f_i(I)$ i $f_i(J)$.

W zależności od doboru macierzy podobieństwa A można uwzględniać różne własności przestrzeni cech (Tadeusiewicz, Flasiński 1991):

- dla $A = I$ - macierz jednostkowa, otrzymuje się metrykę Euklidesową,

- $A = T^{-1}$ prowadzi do metryki Mahalanobisa,

- przyjęcie macierzy A niezerowej tylko poza elementami diagonalnymi prowadzi do metryki Bhattacharyya.

- metryka Mahalanobisa

Jest ona używana w przypadku, gdy każdy wymiar wektora cech jest zależny od siebie. Jej definicja jest następująca:

$$d(I,J) = \sqrt{\left(F_I - F_J\right)^T T^{-1} \left(F_I - F_J\right)}$$

gdzie T jest macierzą kowariancji cech.

W przypadku niezależności wymiarów wektora do wyznaczenia odległości Mahalanobisa używa się wariancji każdego składnika cechy t_i:

$$d(I,J) = \sum_{i=1}^{N} \frac{\left(F_I - F_J\right)^2}{t_i}$$

- dywergencja Kullbacka-Leiblera (entropia względna) oraz dywergencja Jeffrey'a

Entropia względna służy do określania rozbieżności pomiędzy dwoma rozkładami cech. W rzeczywistości nie jest to metryka, gdyż nie spełnia postulatów symetryczności i nierówności trójkąta. Dla rozkładów dyskretnych dywergencja Kullbacka-Leiblera ma postać:

$$d(I,J) = \sum_i f_i(I) \log_2 \frac{f_i(I)}{f_i(J)}$$

Przyjmuje ona wartości nieujemne, a 0 wtedy, gdy dwa rozkłady są identyczne.

Z kolei dywergencja Jeffrey'a, w porównaniu do poprzedniej, jest symetryczna i stabilna numerycznie:

$$d(I,J) = \sum_i f_i(I) \log_2 \frac{f_i(I)}{\overline{f_i}} + f_i(J) \log_2 \frac{f_i(I)}{\overline{f_i}}$$

przy czym $\overline{f_i} = \frac{f_i(I)+f_i(J)}{2}$.

- odległość Hausdorffa

Metoda pomiaru podobieństwa oparta na metryce Hausdorffa stała się bardzo popularna w systemach analizy i wyszukiwania obrazów (Rucklidge 1997; Sim, Kwon, Park 1999; Huttenlocher i in. 1993). Jej podstawowe cechy to (Rotter 2003):

- ✓ odległość między dwoma zbiorami równa się 0 tylko wtedy, gdy te zbiory są identyczne,
- ✓ uwzględnianie dowolnej transformacji jednego obiektu względem drugiego,
- ✓ obliczanie prawo- i lewostronnej odległości Hausdorffa dla obiektów, które nie zostały poprawnie wyodrębnione w procesie segmentacji,
- ✓ odporność na drobne deformacje,
- ✓ zaszumienie obrazu powoduje drastyczne zwiększenie odległości Hausdorffa.

Definicja metryki:

- A, B - niepuste, domknięte i ograniczone podzbiory przestrzeni X

- odległość Hausdorffa wyraża się wzorem:

$$d_H(A,B) = max\big(d_{H+}(A,B), d_{H-}(A,B)\big)$$

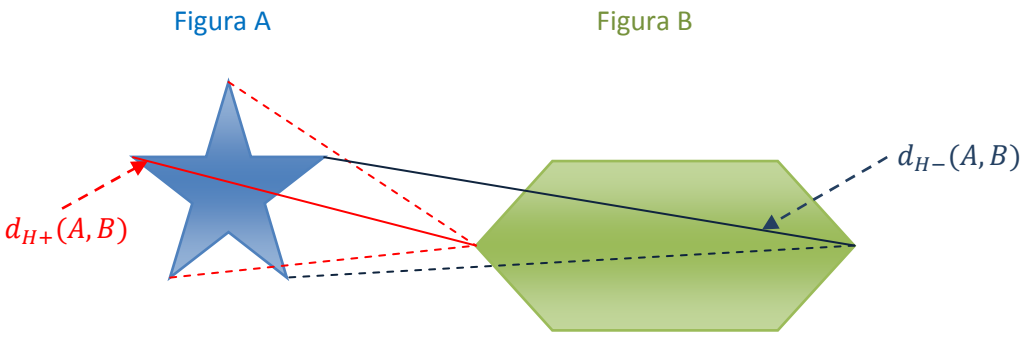
gdzie:

$$d_{H+}(A,B) = sup\{d(x,B): x \in A\}$$

$$d_{H-}(A,B) = sup\{d(y,A): y \in B\},$$

przy czym $d(v,W) = inf\{d(v,w): w \in W\}$ - odległość punktu v od zbioru W.

$d_{H+}(A,B)$, $d_{H-}(A,B)$ - odpowiednio prawo- i lewostronna odległość Hausdorfa.



Rys. 2.3 Przykład prawo- i lewostronnej odległości Hausdorffa pomiędzy figurami A i B.

Metoda wyznaczania odległości Hausdorffa odbywa się dwuetapowo:

- dla każdego piksela ze zbioru A poszukuje się najkrótszej odległości do zbioru B i spośród nich wybiera się największą - $d_{H+}(A,B)$,

- analogicznie postępuje się w przypadku odległości $d_{H-}(A,B)$,

- spośród tych dwóch odległości wybiera się większą, która jest faktyczną miarą odległości Hausdorffa.

Efektywne wykorzystanie metryki Hausdorffa do obliczania podobieństwa między obrazami wymaga użycia bardziej zaawansowanych metod. Wynika to z faktu, że wyznaczanie infimum oraz supremum dla każdego piksela obrazu wprost z powyższych wzorów jest złożone obliczeniowo. W literaturze można spotkać inne podejścia do problemu obliczania metryki Hausdorffa wykorzystujące np. aproksymację poprzez wielokąty (Alt, Behrends, Blömer 1991), triangulację Delaunay'a, czy metody przyspieszonego znajdowania minimum. Ich opis oraz przykłady zastosowania można znaleźć w pracy Rottera (2003).

Deskryptory kształtu są niewątpliwie jednym z najbardziej popularnych wektorów cech stosowanych w technikach reprezentacji i wyszukiwania obrazów. W typowych systemach CBIR można wyróżnić dwie metody wyszukiwania obrazów na podstawie kształtu:

- zapytanie poprzez przykład (ang. query by example) – użytkownik wprowadza do systemu przykładowe zdjęcie oczekując wyników podobnych do zadanego wejścia,

- zapytanie poprzez szkic (ang. query by sketch) – w tym przypadku użytkownik nie wie dokładnie czego poszukuje, dlatego wprowadza do systemu tylko szkic informacji. Jako wyjście otrzymuje on obrazy zawierające podobne kształty w kontekście zaimplementowanych technik miary podobieństwa.

2.1.4 Sposoby indeksowania i redukcji wektora cech

Kolejnym bardzo ważnym zagadnieniem systemów CBIR jest efektywne indeksowanie i wyszukiwanie obrazów oparte na wektorach cech. W związku z tym, że wektory mają dużą wymiarowość, nie nadają się do indeksowania tradycyjnymi metodami. Dlatego też stosuje się techniki zmniejszania wymiarowości wektorów, które w konsekwencji prowadzą do łatwiejszego procesu porównywania obrazów.

- analiza głównych składowych (ang. Principal Component Analysis - PCA)

Analiza głównych składowych zwana także transformatą Karhunena-Loeve jest jedną z najczęściej stosowanych i opisanych technik zmniejszania wymiarowości wektora cech (Jolliffe 1986; Heidemann 2004; Katsumata, Matsuyama 2005; Lu, He 2005). Polega ona na pozbyciu się z wektora tych cech, które wykazują pewien stopień korelacji z innymi. Dzięki temu uzyskuje się zmniejszenie liczby informacji poprzez wydobycie najważniejszych danych (Kambhatla, Leen 1997).

Algorytm PCA można podzielić na następujące etapy:

- ✓ obliczanie średniej dla każdego wiersza:

$$u[m] = \frac{1}{N}\sum_{n=1}^{N} X[m,n]$$

  gdzie $X[m,n]$ - wektor cech o danym rozmiarze,
- ✓ obliczenie macierzy odchyleń:

$$B[i,j] = X[i,j] - u[i]$$

- ✓ wyznaczenie macierzy kowariancji:

$$C = \frac{1}{N} B \cdot B^*$$

  gdzie $B^* = (\bar{B})^T$ - sprzężenie hermitowskie macierzy B,
- ✓ wyznaczenie wartości własnych macierzy kowariancji C i odpowiadających im wektorów własnych,
- ✓ z otrzymanych wartości własnych wybiera się największe – minimalizuje to straty informacji spowodowane zmniejszeniem wymiaru wektora,
- ✓ dla podzbioru wartości własnych wykonuje się rzutowanie na wektory własne:

$$y = V^T \cdot X = \begin{bmatrix} v_1^T \\ \vdots \\ v_{n-1}^T \end{bmatrix} \cdot X$$

gdzie: V - macierz wektorów własnych, X- wektor rzutowany, y – reprezentacja wektor X w nowej przestrzeni, n - ilość wektorów własnych.

(http://en.wikipedia.org/wiki/Principal_component_analysis).

Klasyczna metoda PCA jest liniową aproksymacją zbioru danych, dlatego nie stosuje się jej do nieliniowych dystrybucji. W takich przypadkach używa się techniki lokalnej PCA, która została skutecznie zaadoptowana na potrzeby interpolacji i rozpoznawania obrazów, czy też klasyfikacji twarzy (Heidemann 2004).

- metody indeksowania (Long, Zhang, Dagan Feng 2003)

Istnieje wiele sposobów podejścia do problemu indeksowania danych wielowymiarowych. Do najbardziej znanych można zaliczyć metody R-tree (R*-tree) (Beckmann 1990), linear quad-trees (Vendrig, Worring, Smeulders 1999), K-d-B tree (Robinson 1981). Ich skuteczność jest uzależniona od wielkości wektora, rośnie ekspotencjalnie wraz ze wzrostem wymiarowości. Dodatkowo powyższe sposoby zakładają wykorzystanie metryki Euklidesowej do porównywania obrazów, podczas gdy nie wszystkie systemy CBIR spełniają ten warunek. Innym podejściem do problemu indeksowania jest użycie map samoorganizujących (ang. Self-Organization Map - SOM) zaproponowane przez Zhang i Zhong (1995).

2.1.5 Modelowanie obiektów złożonych

Cechą charakterystyczną znacznej liczby obrazów jest występowanie pewnej hierarchicznej struktury ich budowy. Polega ona na tym, że dane obiekty obrazu składają się z elementarnych kształtów połączonych ze sobą w odpowiedni sposób. Przykładem takich obrazów mogą być np. auta, czy budynki składające się z dużej liczby okien symetrycznie rozmieszczonych, Stąd też w literaturze można spotkać różne podejścia do problemu tzw. *matchingu* obrazów. Wśród nich wyróżnia się np. metody porównujące wektory cech, metody bezpośredniego określania podobieństwa obiektów (Veltkamp, Hagedoorn 1999), metody ciągowe (Tadeusiewicz, Flasiński 1991), czy chociażby bardzo rozpowszechnione metody grafowe (Wong 1992; Floriani, Falcidieno 1998; Rotter 2003).

W niniejszym rozdziale skoncentrujemy się na opisie dwóch metod modelowania obiektów złożonych, które zasadniczo polegają na zidentyfikowaniu podobiektów składających się na dany obiekt i określeniu wzajemnych relacji pomiędzy nimi. Są to:

- metody reprezentacji za pomocą dwuwymiarowych ciągów (ang. 2D strings),
- metody grafowe

1. Metody ciągowe

Jedną ze znanych technik modelowania relacji przestrzennych między obiektami jest metoda reprezentacji obrazu za pomocą skojarzonego z nim ciągu (ang. 2D strings) zaproponowana przez Changa (1987). Polega ona na skonstruowania dwóch zbiorów $\{V, A\}$ będących wynikiem rzutowania obrazu odpowiednio na osie x i y układu współrzędnych. V określa zbiór obiektów obrazu, podczas gdy zbiór A opisuje rodzaj przestrzennej relacji pomiędzy obiektami. Taka reprezentacja pozwala na wyszukiwanie podobnych do siebie obrazów poprzez dopasowywanie ciągów.

Istnieje kilka odmian powyższego sposobu określania obiektów obrazów. Metoda 2-D G-string (Chang, Jungert, Li 1988) dzieli obiekty na mniejsze części i rozszerza pojęcie relacji przestrzennych dzieląc je na dwie grupy: lokalne i globalne, które wskazują czy rzuty obiektów są połączone, rozłączone, czy znajdują się w ty samym miejscu. Dodatkowo metody 2-D C-string (Lee, Hsu 1990) i 2-D B-string (Lee, Yang, Chen 1992) odpowiednio minimalizują liczbę podziału obiektów, bądź proponują jego reprezentację za pomocą dwóch zmiennych określających początek i koniec jego obwiedni.

Cechą wspólną powyższych metod ciągowych jest możliwość zastosowania ich do jednego z trzech sposobów wyszukiwania obrazów:

- pierwszy znajduje wszystkie zdjęcia zawierające obiekty $O_1, O_2, \dots, O_n$,

- drugi znajduje wszystkie zdjęcia zawierające obiekty będące w określonej relacji względem siebie bez uwzględniania odległości,

- trzeci bazuje na odległości pomiędzy obiektami.

Wyszukiwanie obrazów w oparciu o podobieństwo relacji pomiędzy obiektami pozostaje dużym problemem dla systemów CBIR. Wynika to z faktu, że wyznaczenie lokalizacji poszczególnych obiektów zdjęcia nie jest łatwe i jednoznaczne. Niektóre systemy dzielą obraz na regularne części, co prowadzi jednak do ograniczonej poprawy skuteczności.

2. Metody grafowe

Wśród metod grafowych wykorzystywanych do modelowania obiektów złożonych wyróżnia się dwa zasadnicze podejścia (Floriani, Falcidieno 1998):

- reprezentacja objętościowa (ang. volumetric represenatation)
- reprezentacja za pomocą ograniczeń (ang. boundary representation).

Pierwsza reprezentacja dotyczy technik, które opisują obiekt w postaci kombinacji podstawowych figur. Grupa ta zawiera wszelkie metody dekompozycji, które są związane z reprezentacją przestrzenną np. kodowanie poprzez drzewo ósemkowe (ang.

octree encoding), strukturalne techniki geometrii przestrzennej (ang. constructive solid geometry techniques). Reprezentacja za pomocą ograniczeń polega natomiast na dekompozycji figur na elementarne objętości (Rotter 2003). Wykorzystuje geometryczny i topologiczny opis obiektu, który jest podzielony na skończoną liczbę ograniczonych podzbiorów (ang. faces). Każdy taki podzbiór jest z kolei reprezentowany przez krawędzie i wierzchołki.

Floriani i Falcidieno w swojej pracy z 1998 r. wymyślili interesującą metodę grafową modelowania obiektów złożonych wykorzystując pojęcia wielopoziomowości i hipergrafów. Polega ona na modelowaniu obiektu trójwymiarowego za pomocą grafu HFAH (ang. Hierarchical Face Adjacency Hypergraph), który jest zdefiniowany w postaci pary $g^* = (G, T)$, gdzie G jest rodziną hipergrafów, czyli grafów zawierających krawędzie łączące więcej niż dwa wierzchołki, a T określa drzewo opisujące hierarchię obiektu. Metoda ta jest wysoce skuteczna, gdyż znacznie ułatwia definiowanie skomplikowanych figur i odzwierciedlenie hierarchii obiektów obrazu (Rotter 2003).

Znacznie prostszym sposobem modelowania obiektów złożonych jest natomiast konstrukcja grafu bez hierarchii, którego wierzchołki i krawędzie bezpośrednio odpowiadają wierzchołkom i krawędziom figury (Wong 1992).

Metody grafowe są uważane za uniwersalne narzędzie do opisu zawartości obrazu i reprezentacji jego obiektów. Jak podaje Bunke (2000) jedną z przyczyn tego zjawiska jest pewien stopień ich inwariantności względem przekształceń afinicznych. Graf narysowany na papierze w momencie translacji, rotacji lub przekształcenia do swojego lustrzanego odbicia, nadal pozostaje tym samy grafem w sensie matematycznym. Interesującą dziedziną dotyczą modelowania obiektów złożonych jest także proces tzw. *matchingu* grafów, czyli zbiór technik używanych do ich dopasowywania. W literaturze można spotkać kilka metod stosowanych w tym przypadku, wśród których warto wyróżnić metody grafu dopasowań polegające na znajdywania największego podgrafu pełnego (Schalkoff 1991).

Porównywanie obrazów jest kluczowym problemem w procesie wyszukiwania informacji w systemach CBIR, który ma decydujący wpływ na jego końcowy wynik. Skuteczność technik bazujących na wizualnym podobieństwie obiektów (regionów), czy też na obliczaniu miary tego podobieństwa zależy przede wszystkim od cech charakterystycznych obrazu, dlatego nie można wyróżnić metody, który byłaby najefektywniejszą. Problemem systemów CBIR jest zatem nie zaimplementowanie technik pomiaru podobieństwa, ale określenie odpowiedniej miary relatywnie nadającej się do automatycznego procesu wyszukiwania.

2.2 Metody z interakcją

Automatyzacja procesu wyszukiwania obrazów jest kluczowym zagadnieniem i głównym celem stawianym przed każdym systemem CBIR. Początkowe rozwiązania tego problemu zostały przedstawione w rozdziale 2.1, który opisuje techniki analizy obrazu oparte o cechy niskiego poziomu tzn. deskryptory koloru, tekstury, kształtu, etc. Niestety dokładność i skuteczność systemów wykorzystujących metody niższego rzędu jest nadal ograniczona. Wynika to z faktu, że żadna pojedyncza cecha lub ich kombinacja nie opisuje obrazu w kontekście jego zawartości. W literaturze problem ten jest określany jako tzw. „luka semantyczna" pomiędzy cechami niskiego poziomu, a interpretacją zdjęcia uwzględniającą subiektywną ocenę użytkownika (Venters, Hartley, Hewitt 2004; Muneesawang, Guan 2006; Lew, Sebe, Djeraba i in. 2006). Człowiek określa podobieństwo pomiędzy obrazami dzięki rozpoznawaniu obiektów i relacji między nimi oraz poprzez własną subiektywną ocenę. Dodatkowo to samo zdjęcie może zostać inaczej opisane przez różne osoby, bądź nawet przez tą samą w momencie, gdy zmienia się cel poszukiwań.

Podstawową koncepcją rozwiązania powyższego problemu jest zaadoptowanie w procesie wyszukiwania obrazów metod wyższego rzędu bazujących na percepcyjnym rozumowaniu człowieka oraz opartych o jego interakcje z systemem.

2.2.1 Sposoby formułowania zapytań

Ingerencja w proces wyszukiwania obrazów może być zrealizowana wyłącznie za pomocą interakcji użytkownika z systemem CBIR. Interakcja ta zwykle składa się z dwóch części: formułowania zapytania oraz prezentacji wyniku (Long, Zhang, Dagan Feng 2003).

Istnieje kilka sposobów formułowania zapytań, które w ogólności można podzielić na:

- wybór kategorii (ang. category browsing) – metoda polegająca na przeszukiwaniu zdjęć należących do określonej kategorii. W tym celu obrazy w bazie danych są sklasyfikowane do konkretnych kategorii zgodnie z ich wizualną zawartością (Vailaya, Figueiredo i in. 2001);
- zapytanie poprzez szkic (ang. query by sketch) – użytkownik zostaje poproszony o narysowanie szkicu za pomocą wbudowanego w system narzędzia do rysowania. Zapytania mogą mieć postać kilku obiektów o określonych cechach, jak np. kolor, tekstura, kształt, lokalizacja (Finlayson 1996);
- zapytanie poprzez przykład (ang. query by example) – proces wyszukiwania obrazów polega na wprowadzeniu przez użytkownika przykładowego obrazu, który zostaje następnie poddany konwersji w celu wyodrębnienia jego

podstawowych wewnętrznych cech. Zapytanie poprzez przykład może zostać rozszerzone na wyszukiwanie w oparciu o zewnętrzne i wewnętrzne obrazy bazy danych. Wewnętrzne zdjęcia bazy są sklasyfikowane, a relacje między ich charakterystycznymi obiektami (regionami) są uprzednio wyznaczone. Główną zaletą tego zapytania jest to, że użytkownik nie musi dostarczyć dokładnego opisu zdjęcia, gdyż jest ono realizowane poprzez system (Assfalg, Del Bimbo, Pala 2000);

- zapytanie poprzez grupę przykładów (ang. query by group example) – podobnie, jak wyżej z tą różnicą, że użytkownik wprowadza do systemu grupę przykładowych zdjęć. Dzięki temu cel wyszukiwania może zostać zdefiniowany bardziej precyzyjnie poprzez określenie odpowiednich relacji i odrzucenie błędnych. Wiele systemów CBIR wykorzystuje metody zapytań poprzez pozytywne i negatywne przykłady (Long, Zhang, Dagan Feng 2003).

2.2.2 Sprzężenie zwrotne (ang. relevance feedback)

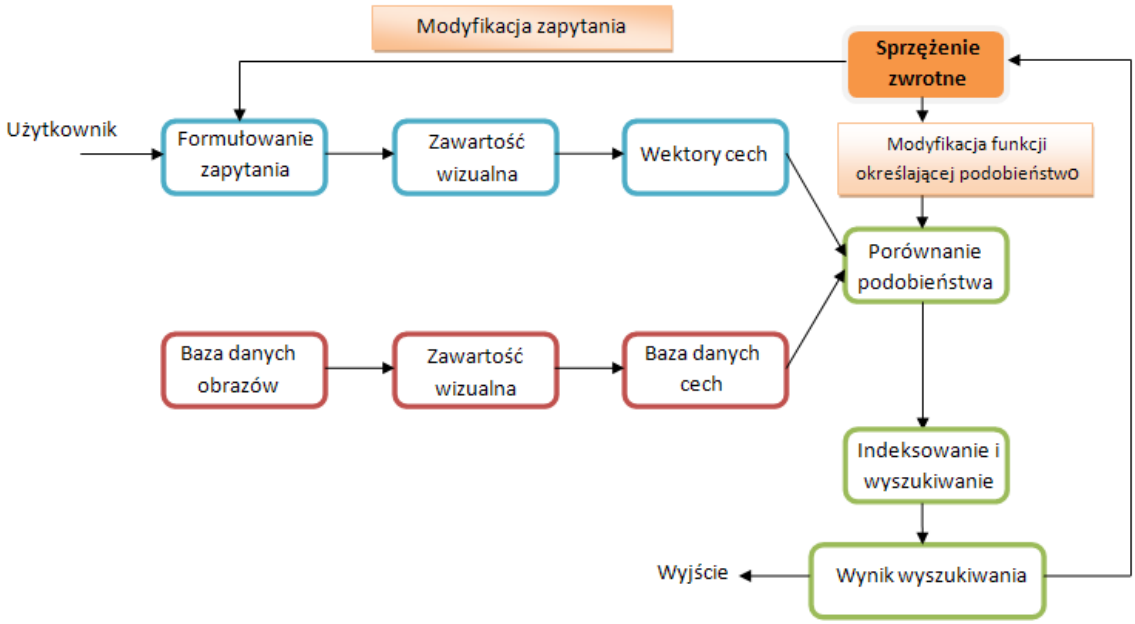
Główna ideą stosowania metod sprzężenia zwrotnego w systemach CBIR jest uwzględnienie subiektywizmu oraz percepcji człowieka w procesie wyszukiwania obrazów. Interakcja użytkownika polega więc na możliwości wpływania na wynik wyszukiwania.

Sprzężenie zwrotne, początkowo zaadoptowane na potrzeby przeszukiwania dokumentów tekstowych (MacArthur, Brodley, Shyu 2000), jest techniką aktywnego uczenia używaną do zwiększenia skuteczności systemów informatycznych. Metoda ta wykorzystuje pozytywne oraz negatywne przykłady dostarczane przez użytkownika, które następnie są porównywane z listą uprzednio uszeregowanych obrazów (zgodnie ze zdefiniowaną miarą podobieństwa). Obrazy wynikowe są poddawane ocenie przez użytkownika, który określa czy dany wynik jest odpowiedni (ang. relevant) – pozytywny przykład, lub nieodpowiedni (ang. irrelevant) – negatywny przykład. W oparciu o nowe dane system ulepsza zapytanie i zwraca listę wynikowych obrazów. Stąd też podstawowym zadaniem w procesie sprzężenia zwrotnego jest (Zhang 2003):

- użycie pozytywnych i negatywnych przykładów do tworzenia nowego zapytania,

- dopasowanie (wybór) odpowiedniej miary podobieństwa.

Zagadnienie sprzężenia zwrotnego stało się bardzo popularną dziedziną badań, co znalazło szerokie odbicie w literaturze dotyczącej systemów CBIR (Rocchio 1971; Rui, Huang, Mehrotra 1998; Minka, Picard 1997; Su, Zhang i in. 2003). Zaadoptowanie tej techniki na potrzeby procesu wyszukiwania obrazów wpłynęło znacząco na zwiększenie jego dokładności i efektywności.



Rys. 2.4 Przykładowy schemat systemu CBIR ze sprzężeniem zwrotnym – relevance feedback.

Analiza sprzężenia zwrotnego wykorzystywanego na potrzeby systemów CBIR wiąże się z takimi zagadnieniami, jak schemat uczenia się, odpowiedni wybór cech, sposób indeksowania i skalowalności. W niniejszej części pracy zostaną opisane zarówno klasyczne podejścia do problemu sprzężenia zwrotnego, jak również skoncentrujemy się na analizie tych metod w kontekście problemu maszyny uczącej z pamięcią.

a) Algorytmy klasyczne

- Początkowe podejścia do problemu sprzężenia zwrotnego w systemach CBIR zostały zapożyczone z klasycznej metody wyszukiwania dokumentów tekstowych. Dzieliły się one na dwa sposoby: „przemieszczanie zapytania” (ang. query point movement) oraz dobór właściwej miary podobieństwa (ang. re-weightening method). Oba te schematy bazowały na analizie wektora modelu (Salton, McGill 1983).

Pierwsza metoda polega na wyborze właściwego punktu zapytania poprzez przesuwanie go w stronę pozytywnych przykładów oraz z dala od negatywnych. Jedną z najbardziej znanych technik iteracyjnych zmian punktu zapytania jest formuła Rocchio (1971) postaci:

$$Q' = \alpha Q + \beta \left(\frac{1}{N_{R'}} \sum_{i \in D'_R} D_i \right) - \gamma \left(\frac{1}{N_{N'}} \sum_{i \in D'_N} D_i \right)$$

gdzie: D'_R - zbiór odpowiednich dokumentów (ang. relevant), D'_N - zbiór nieodpowiednich dokumentów (ang. irrelevant), α, β, γ - parametry określające udział oryginalnego obrazu i sprzężenia zwrotnego w zmodyfikowanym zapytaniu, N'_R i N'_N - odpowiednio liczba dokumentów w zbiorze D'_R i D'_N. Technika ta znana także jako zapytanie poprzez wektor uczący został użyta w systemie MARS (Salton, McGill 1983; Rui, Huang, Mehrotra 1997).

Druga metoda polega na zmianie wag przypisanych poszczególnym wektorom cech. Wzmacnia się wagi oraz znaczenie wymiarów cech, które pomagają odnaleźć właściwe obrazy oraz zmniejsza się wpływ wektorów cech niewłaściwych. Jest to realizowane poprzez modyfikację tych wektorów w metryce określającej stopień podobieństwa. Metryka ta może mieć postać (Zhang 2003):

$$d = \sum_{j \in [N]} \omega_j \cdot \left| X_j^{(1)} - X_j^{(2)} \right|$$

Jeżeli obraz wynikowy okaże się właściwy, składniki wektora charakteryzujące się większym podobieństwem do wzorca są uważane za bardziej istotne od tych, które nie wykazują tego podobieństwa. Stad też waga właściwego składnika - ω_i - jest modyfikowana:

$$\omega_i = \omega_i \cdot \left(1 + \bar{\delta} - \delta_i\right), \quad \delta = \left| f(Q) - f(A_j^+) \right|$$

gdzie: $\bar{\delta}$ - jest średnią z δ.

Z drugiej strony, jeżeli obraz zostanie uznany za negatywny przykład, wagi niewłaściwych składników zostaną obniżone zgodnie ze wzorem:

$$\omega_i = \omega_i \cdot \left(1 - \bar{\delta} + \delta_i\right)$$

Technika ta, znana pod nazwą *uczenie metryki* (ang. learning the metric), została zaproponowana w pracy Huang, Kumar i Metra (1997).

- Inne podejście do problemu sprzężenia zwrotnego zostało zaproponowane w pracy Minka i Picard (1997). Polega ono na zmodyfikowaniu przestrzeni zapytań poprzez odpowiedni wybór modeli cech. Zakłada się, że każdy taki model reprezentuje jeden aspekt zawartości obrazu, stąd też najlepszym sposobem na efektywne wyszukiwanie zdjęć jest użycie zbiorów modeli (ang. ”society of models”). Podejście to wykorzystuje schemat uczenia w procesie dynamicznego określania, który model lub ich kombinacja jest najlepszy do kolejnego etapu wyszukiwania.

- System MindReader zaproponowany przez Ishikawa, Subramanya i Faloustos (1998) to bardziej odporna obliczeniowo metoda, który wykorzystuje optymalizację globalnych cech obrazu. Formułuje się problem minimalizacji oraz oszacowuje

parametry procesu. W systemie MindReader funkcja odległości nie jest związana z osiami układu współrzędnych, jak ma to miejsce w tradycyjnym systemie wyszukiwania zdjęć. Bazuje ona raczej na korelacji pomiędzy atrybutami obrazu wraz ze zmianą wag przypisanych poszczególnym komponentom.

- Interesująca technika interaktywnego sprzężenia zwrotnego koncentrująca się na połączeniu metod niskiego i wyższego rzędu wyszukiwania obrazów wraz z zagadnieniem subiektywnej oceny jego zawartości przez człowieka została zaproponowana przez Rui, Huang i Mehrotra (1998).

Opis zastosowanej przez nich techniki sprzężenia zwrotnego wymaga stworzenia formalnego modelu obrazu O, który jest reprezentowany przez trójkę:

$$O = O(D, F, R)$$

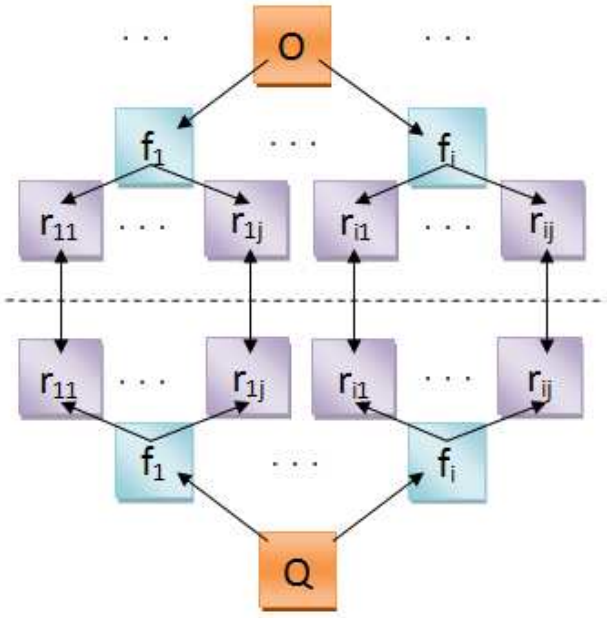
- D – określa podstawowe dane obrazu np. zdjęcie JPEG,
- $F = \{f_i\}$ - zbiór wizualnych cech obrazu niskiego poziomu: kolor, tekstura, kształt,
- $R = \{r_{ij}\}$ - zbiór metod reprezentacji danej cechy f_i np. histogram lub momenty koloru, cechy teksturowe Tamury, itp. Każda reprezentacja r_{ij} może być wektorem złożonym z wielu składników:

$$r_{ij} = [r_{ij1}, r_{ij2}, \dots, r_{ijK}]$$

gdzie: K - długość wektora.

Powyższa koncepcja modelu umożliwia używanie wielu reprezentacji opisu cech obrazu oraz stosowanie dynamicznej modyfikacji wag, w celu dokładniejszego odzwierciedlenia zawartości zdjęcia w procesie sprzężenia zwrotnego. Wyróżnia ona trzy rodzaje wag W_i, W_{ij}, W_{ijk} związane odpowiednio z cechami f_i, reprezentacjami r_{ij} oraz z ich składnikami r_{ijk}. Głównym zadaniem sprzężenia zwrotnego opracowanego przez Rui, Huang i Mehrotra jest zatem znalezienie takich wartości wag, które w najlepszy sposób modelowałyby informacje uzyskiwane od użytkownika systemu.

Model obrazu wraz ze zbiorem miar podobieństwa $M = \{m_{ij}\}$ definiuje charakterystyczną dla systemu CBIR czwórkę (D, F, R, M). Używanie konkretnej miary podobieństwa jest uzależnione od reprezentacji cechy obrazu, stąd np. metryka Euklidesowa jest używana do porównywania reprezentacji cech określanych za pomocą wektorów, podczas gdy przecięcie histogramu nadaje się najlepiej do analizy koloru. Cały proces wyszukiwania obrazów można zatem przedstawić w postaci poniższego schematu:



Rys. 2.5 Schemat procesu wyszukiwania dla metody Rui, Huang i Mehrotra (1998).

Poszczególne etapy tego procesu są następujące:

- początkowa inicjalizacja wag $W = [W_i, W_{ij}, W_{ijk}]$ do wartości $W0$ tzn. każda waga ma tą samą ważność:

$$W_i = W0_i = \frac{1}{I}$$

$$W_{ij} = W0_{ij} = \frac{1}{J_i}$$

$$W_{ijk} = W0_{ijk} = \frac{1}{K_{ij}}$$

gdzie: I - numer cechy w zbiorze F, J_i - numer reprezentacji dla cechy f_i, K_{ij} - długość wektora reprezentacji r_{ij}.

- informacja zawarta w zapytaniu użytkownika (Q) jest rozdzielana pomiędzy cechy f_i zgodnie z wagami W_i,
- dla każdej cechy f_i informacja jest znowu dzielona pomiędzy reprezentacje r_{ij} zgodnie z wagami W_{ij},
- podobieństwo obrazu do zapytania w kontekście reprezentacji r_{ij} jest obliczane według miary m_{ij} i wag W_{ijk}:

$$S(r_{ij}) = m_{ij}(r_{ij}, W_{ijk})$$

- wyznaczenie wartości podobieństwa cech za pomocą podobieństwa reprezentacji:

$$S(f_i) = \sum_j W_{ij} S(r_{ij})$$

- końcowa wartość podobieństwa to suma wszystkich podobieństw cech:

$$S = \sum_i W_i S(f_i)$$

- obrazy w bazie danych są porządkowane według ich całkowitego podobieństwa do zapytania Q. Najbardziej zbliżone rezultaty N_{RT} są zwracane do użytkownika,
- dla każdego wyniku wyszukiwania użytkownik subiektywnie określa stopień pokrewieństwa obrazu z zapytaniem poprzez wybór jednej z następujących opcji: wysoce odpowiedni (ang. highly relevant), odpowiedni, brak opinii (ang. no opinion), nieodpowiedni (ang. non-relevant), wysoce nieodpowiedni,
- system modyfikuje wagi zgodnie z informacją uzyskaną ze sprzężenia zwrotnego, dzięki czemu nowe zapytanie jest lepszą aproksymacją uzyskanych informacji,
- nowa iteracja wykorzystująca zmodyfikowane zapytanie rozpoczyna się od kroku drugiego.

Powyższy algorytm sprzężenia zwrotnego zakłada, że podobieństwa S i $S(f_i)$ są liniową kombinacją odpowiadających im niższych podobieństw. Dlatego też wprowadza się nowy wzór na końcową wartość podobieństwa postaci:

$$S = \sum_i \sum_j W_{ij} S(r_{ij}),$$

gdzie wagi W_{ij} są bezpośrednio określane za pomocą reprezentacji r_{ij}.

Kluczowymi punktami powyższej metody są dwa zagadnienia związane z *normalizacją* i *modyfikacją wag*, które zostaną opisane poniżej.

Normalizacja:

Algorytm wyszukiwania obrazów zakładał, że wartości podobieństwa reprezentacji $S(r_{ij})$ mieszczą się w tym samym zakresie np. od 0 do 1. W przeciwnym wypadku liniowa kombinacja tych wartości tworząca końcową miarę S nie miałaby sensu. Podobnie składniki r_{ijk} wektora reprezentacji powinny być znormalizowane przed obliczaniem miary m_{ij}. Stąd też autorzy wprowadzili dwa pojęcia: intra-normalizacji dla składników r_{ijk} oraz inter-normalizacji dla $S(r_{ij})$.

- intra-normalizacja

W wyniku przeprowadzenia tego procesu otrzymuje się wektor reprezentacji r_{ij}, w którym każdy składnik ma taki sam wpływ na jego kształt. Oznaczając $V = r_{ij}$ przyjmuje się, że każdy wektor reprezentacji musi zostać poddany procedurze normalizacyjnej. Przy założeniu, że w bazie danych jest M obrazów indeksowanych przez m otrzymuje się, że:

$$V = V_m = \left[V_{m,1}, V_{m,2}, \dots, V_{m,k}, \dots, V_{m,K}\right]$$

jest wektorem reprezentacji obrazu m, gdzie K - długość wektora. Następnie umieszcza się wszystkie te wektory w macierzy $\upsilon = [\upsilon_{m.k}], m = 1, \dots, M,\ k = 1, \dots, K$, gdzie $\upsilon_{m.k}$ jest k-tym składnikiem wektora V_m. Proces normalizacji odbywa się za pomocą metody Gaussa. Obliczając średnią μ_k oraz odchylenie standardowe σ_k dla sekwencji υ_k (k-ta kolumna macierzy υ) otrzymuje się znormalizowany do przedziału $[0,1]$ wektor reprezentacji postaci:

$$\upsilon_{m.k} = \frac{\upsilon_{m.k} - \mu_k}{\sigma_k}$$

- inter-normalizacja

Procedura inter-normalizacji zapewnia równy nacisk każdej składowej podobieństwa reprezentacji $S(r_{ij})$ na końcową wartość podobieństwa S. Jest ona realizowana w kilku poniższych krokach:

- dla każdej pary obrazów I_m oraz I_n oblicza się ich podobieństwo $S_{m,n}(r_{ij})$:

$$S_{m,n}(r_{ij}) = m_{ij}(r_{ij}, W_{ijk}),\ m, n = 1, \dots, M \quad m \neq n$$

- ponieważ baza danych zawiera M zdjęć, zatem występuje $C_2^M = \binom{M}{2} = \frac{M \cdot (M-1)}{2}$ możliwych wartości ich podobieństwa. Traktując je jako pewien ciąg danych można podać wartość średnią μ_{ij} oraz odchylenie standardowe σ_{ij}tej sekwencji,

- dla danego zapytania Q wyznacza się wartości podobieństwa pomiędzy Q, a obrazami w bazie danych:

$$S_{m,Q}(r_{ij}) = m_{ij}(r_{ij}, W_{ijk})$$

- następnie normalizuje się te wartości do postaci:

$$S'_{m,Q}(r_{ij}) = \frac{S_{m,Q}(r_{ij}) - \mu_{ij}}{3\sigma_{ij}},$$

co zapewnia, że 99% tych wartości będzie w zakresie $[-1,1]$ (reguła trzech sigm). Dodatkowe przekształcenie wartości do przedziału $[0,1]$ jest następujące:

$$S''_{m,Q}(r_{ij}) = \frac{S''_{m,Q}(r_{ij}) + 1}{2}$$

Modyfikacja wag:

- wagi W_{ij}:

Wagi przypisane do poszczególnych wektorów reprezentacji r_{ij} służą do bardziej precyzyjnego opisu informacji dostarczanych przez użytkownika. Wprowadźmy oznaczenia:

- $RT = [RT_1, RT_2, \dots, RT_l, \dots, RT_{N_{RT}}]$ - zbiór najbardziej zbliżonych pod względem wartości podobieństwa S wyników wyszukiwania obrazów,

- $Score_l = \begin{cases} 3 \\ 1 \\ 0 \\ -1 \\ -3 \end{cases}$ - zbiór zawierający wartości, które przypisuje użytkownik w drodze sprzężenia zwrotnego elementom RT_l.

Powyższe wartości zostały dobrane zupełnie przypadkowo. Rui, Huang i Mehrotra eksperymentalnie ustalili, że 5 liczb jest swego rodzaju kompromisem pomiędzy wygodą użytkownika, a dokładnością procesu.

- $RT^{ij} = [RT_1^{ij}, RT_2^{ij}, \dots, RT_l^{ij}, \dots, RT_{N_{RT}}^{ij}]$ - zbiór najbardziej zbliżonych pod względem wartości podobieństwa $S(r_{ij})$ wyników wyszukiwania obrazów.

W celu obliczenia wag dla elementów r_{ij} początkowo przyjmuje się $W_{ij} = 0$, a następnie:

$$W_{ij} = \begin{cases} W_{ij} + Score_l, \ gdy\ RT_l^{ij}\ należy\ do\ RT \\ W_{ij} + 0, \ gdy\ RT_l^{ij}\ nie\ należy\ do\ RT \end{cases}, l = 0, \dots, N_{RT}$$

Wszystkie obrazy znajdujące się poza zbiorem RT są oznaczane jako brak opinii, a wartość ich wag przyjmuje się równą zero. Dla $W_{ij} < 0$ przypisuje się: $W_{ij} = 0$. Oznaczając przez W_{Tij} całkowitą wartość wszystkich wag: $W_{Tij} = \sum W_{ij}$ normalizuje się wagi tak, aby ich suma była jedynką:

$$W_{ij} = \frac{W_{ij}}{W_{Tij}}$$

- wagi W_{ijk}:

Wagi W_{ijk} przypisane składnikom r_{ijk} odzwierciedlają ich różny wpływ na końcowy wektor reprezentacji, co bezpośrednio przekłada się na lepszą skuteczność wyszukiwania obrazów. Sposób modyfikacji wag odbywa się następująco:

- spośród zdjęć wybiera się te, które zostały oznaczone jako wysoce odpowiednie lub odpowiednie i umieszcza się ich wektor r_{ij} w macierzy o rozmiarze $M' \times K$, gdzie M' jest liczbą określającą ilość tych zdjęć,

- w oparciu o odwrotność odchylenia standardowego ciągu składników r_{ijk} wyznacza się zmodyfikowaną wartość poszczególnych wag:

$$W_{ijk} = \frac{1}{\sigma_{ijk}}$$

Taka modyfikacja zakłada, że użytkownik oznaczy co najmniej jedno zdjęcie, poza zapytaniem tak, aby odchylenie standardowe σ_{ijk} było różne od zera. Następnie przeprowadza się prostą normalizację wartości wag tak, jak ma to miejsce w przypadku wektorów reprezentacji:

$$W_{ijk} = \frac{W_{ijk}}{W_{Tijk}}, \; W_{Tijk} = \sum W_{ijk}$$

Podsumowując przedstawioną powyżej metodę sprzężenia zwrotnego autorstwa Rui, Huang i Mehrotra (1998) można wyróżnić trzy zasadnicze cechy użytego algorytmu:

- ✓ wielomodalność – zaproponowany model obrazu wykorzystuje kilka sposobów jego analizy opartej o charakterystyczne cechy i wektory reprezentacji. Jest to uniwersalne podejście, które umożliwia systemowi dokładniejsze modelowanie subiektywnej oceny obrazu przez człowieka.
- ✓ interaktywność – przejawia się w umiejętności zaadoptowania, na potrzeby wyszukiwania obrazów, obliczeniowych zdolności komputerowych i percepcyjnego sposobu postrzegania człowieka.
- ✓ dynamiczność – rozumiana jako możliwość modyfikowania wag w drodze sprzężenia zwrotnego. Jest to podwójna zaleta systemu, gdyż:
  - odciąża użytkownika, który nie musi definiować pierwotnego zbioru wszystkich wag, tylko określa stopień podobieństwa pomiędzy wynikiem wyszukiwania, a zapytaniem,
  - odciąża komputer, który nie imituje wysokiego poziomu rozumowania zawartości obrazu.

b) Sprzężenie zwrotne jako maszyna ucząca z pamięcią

Metoda sprzężenia zwrotnego rozumiana w kontekście problemu uczenia z pamięcią wydaje się być uzasadniona w momencie przyjęcia założenia, że system otrzymuje od użytkownika użyteczne informacje na temat odpowiedniości wyników wyszukiwania obrazów. Te informacje mogą być rozumiane jako „doświadczenie systemu" dostarczone przez użytkownika, który określa czy rezultat wyszukiwania jest pozytywny lub negatywny, i w jakim stopniu. Metoda uczenia systemu CBIR znalazła kilka odzwierciedleń w postaci sztucznych sieci neuronowych (Laaksonen, Koskela, i in. 2000), wykorzystania metod prawdopodobieństwa Bayesa (Vasconcelos, Lippman 1999), czy uczących drzew decyzyjnych (MacArthur, Brodley, Shyu 2000). W związku z niechęcią użytkowników do wprowadzania do systemu wielu przykładowych zdjęć, liczba próbek dostępna dla mechanizmu uczącego jest nie duża. Dlatego też podstawowym zadaniem metod uczących sprzężenia zwrotnego jest taka realizacja algorytmów, która z małej ilości dostępnych informacji jest w stanie utworzyć wielowymiarową przestrzeń cech (Zhang 2003). Z drugiej strony zdobyta wiedza systemu musi zostać zapamiętana, aby mogła być następnie wykorzystana do modyfikowania zapytań.

I. Sprzężenie zwrotne jako metoda probabilistyczna

Interesującym sposobem analizy sprzężenia zwrotnego stosowanego w systemach CBIR jest sformułowanie problemu w postaci zadania probabilistycznego. W tym przypadku informacja dostarczana do systemu przez użytkownika jest wykorzystywana do modyfikacji rozkładu prawdopodobieństwa wszystkich obrazów znajdujących się w bazie danych. Najczęściej stosowaną techniką jest tutaj reguła klasyfikacyjna Bayesa, która została szeroko opisana i rozpowszechniona m.in. w pracach Cox, Miller i in. (2000), Vasconcelos, Lippman (1999), Su, Zhang i in. (2003).

- Vasconcelos i Lippman (1999) uznali zadanie wyszukiwania obrazów za problem znalezienia odwzorowania:

$$g : F \rightarrow M = \{1, \dots, K\},$$

w którym: F - przestrzeń cech (reprezentacji), M – zbiór klas obrazów w bazie danych.

W przypadku, gdy celem systemu wyszukiwania jest minimalizacja prawdopodobieństwa błędnego dopasowania, można posłużyć się regułą klasyfikacji Bayesa postaci:

$$g(x) = \arg\max_i P(S_i = 1|x) = arg\max_i P(x|S_i = 1) \cdot P(S_i = 1)$$

gdzie: x – przykłady dostarczone przez użytkownika, S_i - zmienna binarna wskazująca wybór odpowiedniej klasy obrazów i. Jeżeli zamiast pojedynczego zapytania x dysponujemy sekwencją zapytań $\{x_1, \dots, x_t\}$ o interwale t, reguła Bayesa zmienia postać:

$$P(S_i = 1|x_1, \dots, x_t) = \gamma_t \cdot P(x_t|S_i = 1) \cdot P(S_i = 1|x_1, \dots, x_{t-1})$$

gdzie γ_t jest stałą normalizacyjną. Zakłada się, że znając poprawną klasę obrazów obecne zapytanie x_t jest niezależne od wcześniejszego. W konsekwencji oznacza to, że podczas każdej iteracji algorytmu w wyniku interakcji użytkownika dostarczana jest do systemu nowa informacja o przynależności obrazu do danej klasy. Powyższe równanie jest intuicyjne i zapewnia gromadzenie informacji w czasie. Dane dostarczane przez użytkownika w czasie $t - 1$ są podstawą dla iteracji t, która z kolei wpływa na kształt kolejnej. Obliczeniowo procedura ta jest bardzo wydajna, gdyż w każdej iteracji wymaga się obliczenia jedynie podobieństwa danej do zapytania.

- Su, Zhang i in. (2003) zaproponowali nowy algorytm sprzężenia zwrotnego wykorzystując do tego celu rozkład Gaussa, który umożliwia włączanie do procesu wyszukiwania obrazów informacji zawartych w pozytywnych (odpowiednich) i negatywnych przykładach dostarczanych przez użytkownika. Podstawowe założenie ich metody mówi, że podczas jednej iteracji algorytmu znalezione pozytywne przykłady obrazów należą do tej samej klasy semantycznej oraz wszystkie podlegają rozkładowi Gaussa. Dzięki temu proces wyszukiwania obrazów w bazie danych staje się zadaniem dwuetapowym:

- pierwsza część to wykorzystanie pozytywnych przykładów i reguły Bayesa do klasyfikacji obrazów. Proces wyszukiwania staje się zadaniem oszacowania prawdopodobieństwa przynależności obrazu do jednej klasy semantycznej, natomiast modyfikacja zapytania w drodze sprzężenia zwrotnego polega na zmianie parametrów funkcji Gaussa.

- druga część to wykorzystanie negatywnych przykładów oraz metody funkcji kary. Jeżeli obraz jest podobny do negatywnego przykładu, jego wartość będzie obniżona odpowiednio do stopnia podobieństwa do błędnego dopasowania.

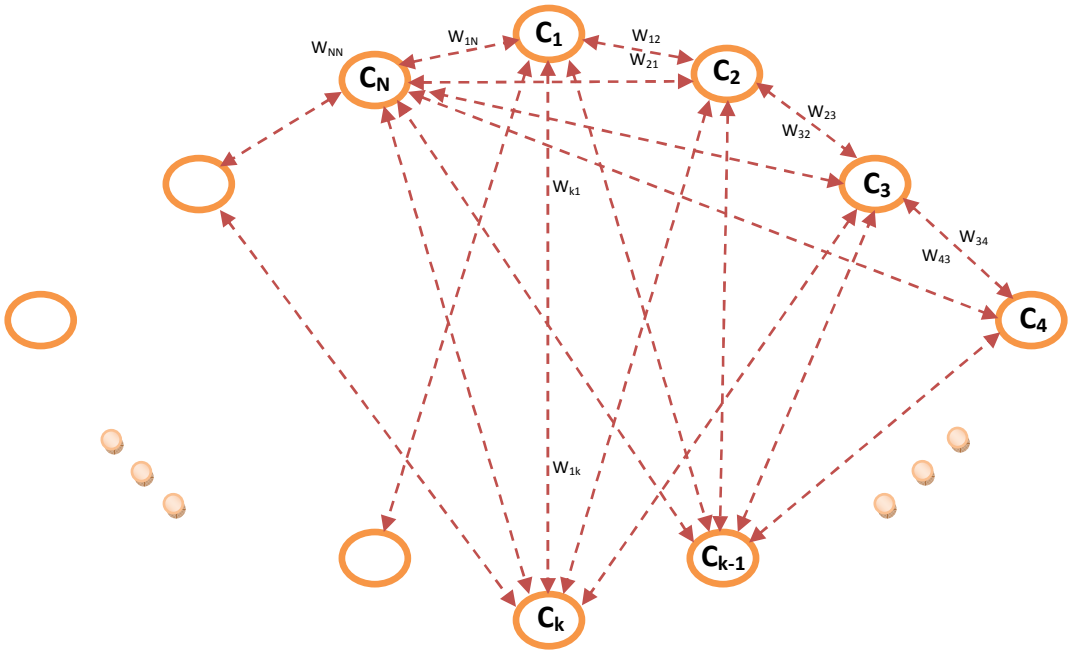
Powyższa metoda Su, Zhang i in. (2003) skutecznie zwiększa efektywność wyszukiwania obrazów w systemie CBIR. Eksperymentalnie stwierdzili oni, że w porównaniu do algorytmu sprzężenia zwrotnego autorstwa Rui, Huang i Mehrotra (1998) podejście wykorzystujące regułę klasyfikacyjną Bayesa razem z modyfikacją parametrów rozkładu Gaussa daje lepsze wyniki, co nie jest okupione zwiększonym nakładem obliczeniowym i pamięciowym.

II. Sprzężenie zwrotne jako proces pamięciowy

Podejście zaproponowane przez Lee, Ma i in. (1999) jest uważane za pierwszą próbę stworzenia metody, która zapamiętywałaby semantyczną informację dostarczaną w drodze sprzężenia zwrotnego przez użytkownika. Podstawową ideą tej techniki jest wykorzystanie sieci korelacji pomiędzy poszczególnymi grupami obrazów. Matematycznie sieć ta jest reprezentowana poprzez macierz korelacji postaci:

$$M = \begin{bmatrix} w_{11} & w_{12} & \dots & w_{1N} \\ w_{21} & w_{22} & \dots & w_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ w_{N1} & w_{N2} & \dots & w_{NN} \end{bmatrix}$$

gdzie w_{ij} reprezentują semantyczny związek pomiędzy obrazami w grupach i oraz j.



Rys. 2.6 Sieć korelacji – zapamiętywanie semantycznych powiązań pomiędzy grupami obrazów.

Wszystkie zdjęcia znajdujące się w bazie danych są dzielone na N grup w oparciu o wizualne podobieństwo cech – np. poprzez algorytm k-średnich. Początkowo współczynniki korelacji pomiędzy grupami są ustawione na zero tzn. tylko obrazy w obrębie tej samej grupy są ze sobą skorelowane: $M_0 = I_{NxN}$ – macierz jednostkowa. Następnie dla danego zapytania początkowe wyszukiwanie jest oparte o wektor cech.

Załóżmy, że po danej iteracji algorytm wyszukał $n+m$ obrazów, przy czym spośród nich n jest oznaczonych jako odpowiednie, a m - jako nieodpowiednie (mogą pochodzić z dowolnych grup). Zapamiętanie informacji zawartych w sprzężeniu zwrotnym odbywa się poprzez modyfikację macierzy korelacji zgodnie ze wzorem (Zhang 2003):

$$M_t = M_{t-1} + \sum_{i=1}^{m} F(q)F(p_i)^T - \sum_{i=1}^{n} F(q)F(n_i)^T$$

gdzie: q - wektor cech zapytania, p_i, n_i - wektory cech odpowiednio z pozytywnych i negatywnych przykładów sprzężenia zwrotnego, $F(x)$ - funkcja transformująca.

Funkcja transformująca $F(x)$ używana do łączenia sprzężenia zwrotnego i określania wielkości modyfikacji jest definiowana następująco:

$$[F(x)]_k = \varphi(\|x - c_k\|) = exp\left\{\frac{-1}{2\sigma^2}\|x - c_k\|^2\right\}$$

gdzie c_k jest środkiem ciężkości grupy k, $1 \leq k \leq N$, a x jest wektorem cech. Reprezentuje ona udział każdej próbki w grupie, który jest odwrotnie proporcjonalny do odległości pomiędzy tą próbką, a środkiem ciężkości grupy k. Inaczej mówiąc próbka znajdująca się bliżej środka grupy wnosi większy udział w procesie modyfikacji zapytania określając poprawny kierunek wyszukiwania.

Metoda realizacji sprzężenia zwrotnego w postaci maszyny uczącej z pamięcią uwzględnia nie tylko wizualne cechy obrazu (kolor, tekstura, kształt), ale także powiązania pomiędzy zawartością poszczególnych obrazów. Eksperymenty przeprowadzone przez Lee, Ma i in. (1999) dowiodły, że postępowe nauczanie systemu poprzez zapamiętywanie wyników zapytań znacząco redukuje liczbę iteracji potrzebną do osiągnięcia wysokiej dokładności wyszukiwania. Dodatkowo uwzględniono, że:

- w przypadku, gdy wewnątrz jednej grupy znajdują się dwa podzbiory o różnej semantyce, następuje jej podział na dwie inne grupy,

- jeżeli dwie grupy charakteryzują się zbliżoną zawartością obrazów tzn. mają wysoki stopień korelacji, wówczas mogą one być połączone tworząc jedną grupę. Oznacza to, że sieć korelacji dynamicznie zmienia swoją strukturę ucząc się ze sprzężenia zwrotnego od użytkownika.

2.2.3 Lokalne deskryptory obrazu

W poniższym rozdziale skoncentrujemy się na interesujących rozwiązaniach problemu wyszukiwania obrazów w systemach CBIR. Przedstawimy kilka popularnych technik wysokiego poziomu, które znalazły szerokie zainteresowanie w świecie analizy obrazu.

a) SIFT

Metoda SIFT (ang. Scale Invariant Feature Transform) została zaproponowana przez Davida Lowe'a w 1999 roku. Jest to skuteczna i wysoce uniwersalna technika bazująca na wyborze charakterystycznych punktów obrazu, które są inwariantne względem przekształceń afinicznych tj. skalowania, obrotu i translacji. Bardzo ważną zaletą jest również jej częściowa odporność na:

- zmiany oświetlenia,
- zaszumienie obrazu,
- przysłanianie obiektów,
- zmianę punktu widzenia kamery.

Charakterystyczne cechy obrazu są dokładnie zlokalizowane zarówno w dziedzinie przestrzennej, jak i częstotliwościowej redukując tym samym prawdopodobieństwa zakłócenia poprzez zaszumienie, czy przysłonięcie. Podstawą do zastosowania powyższej metody w procesie wyszukiwania zdjęć jest minimalizacja czasu wyznaczenia kluczowych deskryptorów obrazu poprzez kaskadowe podejście. Ogólnie metoda SIFT dzieli się na cztery zasadnicze części (Lowe 2004):

I. Scale-space extrema detection - detekcja potencjalnych punktów charakterystycznych przy zastosowaniu piramidy obrazów o zmieniającej się rozdzielczości (Pawlik, Mikrut 2006).
II. Keypoint localization - określenie lokalizacji oraz skali punktów poprzez pomiar ich stabilności.
III. Orientation assignment - wyznaczenie orientacji (jednej lub więcej) każdego punktu poprzez analizę lokalnych kierunków gradientu obrazu.
IV. Keypoint descriptor – stworzenie reprezentacji obrazu poprzez pomiar, dla konkretnej skali, gradientów każdego punktu, co zapewnia odporność na zmiany oświetlenia i zniekształcenie kształtów.

Poniżej przedstawiamy dokładny opis poszczególnych etapów metody SIFT, a następnie skoncentrujemy się na możliwościach adaptacji tej metody w procesie wyszukiwania obrazów.

Scale-space extrema detection

Detekcja charakterystycznych punktów obrazu, inwariantnych względem zmian skali, odbywa się za pomocą kaskadowego podejścia (ang. cascade filtering). Polega ono na

stworzeniu piramidy obrazów $L(x, y, \sigma)$ o zmiennej skali wykorzystując do tego funkcję gęstości rozkładu normalnego Gaussa:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$$

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

gdzie: $I(x, y)$ - obraz wejściowy, $*$ - splot funkcji.

Następnie tworzy się splot zdjęcia z różnicą dwóch funkcji gęstości (ang. doG – difference-of-Gaussian) różniących się współczynnikiem odchylenia standardowego σ:

$$D(x, y, \sigma) = \big(G(x, y, k\sigma) - G(x, y, \sigma)\big) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma)$$

Istnieje kilka powodów wyboru takiej postaci funkcji. Jednym z nich jest szybkość obliczania komputerowego (operacja musi zostać przeprowadzana przy zmieniającej się skali) oraz pewne podobieństwo do znormalizowanego Laplasjanu z funkcji Gaussa (ang. LoG – Laplacian of Gaussian) - $\sigma^2\nabla^2 G$. Normalizacja Laplasjanu za pomocą kwadratu odchylenia standardowego jest wymagana w celu zapewnienia niezmienniczości względem skali. Minima oraz maksima funkcji $\sigma^2\nabla^2 G$ produkują najbardziej stabilne cechy obrazu w porównaniu do funkcji gradientowych, krawędziowych, czy Harrisa.

Pomiędzy różnicą funkcji gęstości, a normalizacją Laplasjanu istnieje zależność, którą opisuje poniższe równanie:

$$\frac{\partial G}{\partial \sigma} = \sigma\nabla^2 G$$

Wynika z niego, że wyrażenie $\nabla^2 G$ może zostać wyznaczone jako iloraz różnicowy funkcji gęstości różniących się skalami $k\sigma$ oraz σ:

$$\sigma\nabla^2 G = \frac{\partial G}{\partial \sigma} \approx \frac{G(x, y, k\sigma) - G(x, y, \sigma)}{k\sigma - \sigma}$$

Stąd otrzymuje się, że:

$$G(x, y, k\sigma) - G(x, y, \sigma) \approx (k - 1)\sigma^2\nabla^2 G$$

Jak widać z powyższych zależności, jeżeli funkcje Gaussa mają skale różniące się o stały współczynnik, wówczas od razu zapewniona jest normalizacja skali σ^2 wymagana dla inwariantności Laplasjanu. Wyrażenie $(k - 1)$ jest stałe dla każdej skali, stąd nie ma ono wpływu na lokalizację ekstremów obrazu.

Metoda detekcji ekstremów rozpoczyna się od stworzenia piramidy obrazów złożonej z oktaw. Oktawa stanowi zbiór obrazów uzyskanych w wyniku splotu zdjęcia wejściowego z funkcją gęstości Gaussa dla różnych parametrów odchylenia standardowego σ. Każda para obrazów wewnątrz danej oktawy jest następnie

odejmowana tworząc różnicę difference-of-Gaussian. W momencie, gdy cała oktawa zostanie zbudowana, tworzy się kolejną poprzez zmniejszenie rozdzielczości obrazu i powtórzenie powyższego postępowania.

W celu znalezienia minimów i maksimów funkcji $D(x, y, \sigma)$ tworzy się ośmioelementowe sąsiedztwo punktu na danym obrazie wewnątrz jednej oktawy oraz odpowiednio dziewięcioelementowe na obrazach o niższej i wyższej skali. Wybór punktu jako ekstremum następuje tylko wtedy, gdy jego wartość jest mniejsza lub większa od wszystkich 26 punktów, z którymi był porównywany.

Keypoint localization

Kolejnym etapem metody SIFT jest określenie lokalizacji punktów charakterystycznych obrazu. Informacja ta pozwala odrzucić punkty wrażliwe na zaszumienie (niski kontrast), bądź też błędnie zlokalizowane wzdłuż krawędzi. Wykorzystując rozwinięcie w szereg Taylora funkcję $D(x, y, \sigma)$ można przedstawić w postaci:

$$D(x) = D + \frac{\partial D^T}{\partial x} x + \frac{1}{2} x^T \frac{\partial^2 D}{\partial x^2} x$$

przy czym funkcja D i jej pochodne są obliczane w punkcie $x = (x, y, \sigma)^T$ oznaczającym przesunięcie. Lokalizacja ekstremum $\hat{x}$ jest wyznaczana poprzez przyrównanie powyższej funkcji ze względu na x do zera:

$$\hat{x} = -\frac{\partial^2 D^{-1}}{\partial x^2} \frac{\partial D}{\partial x}$$

Jeżeli przesunięcie $\hat{x}$ jest większe niż 0.5 w każdym kierunku, oznacza to, że ekstremum leży bliżej innego punktu. W takim wypadku następuje zmiana punktu i proces interpolacji rozpoczyna się od nowa. Finalne przesunięcie $\hat{x}$ jest dodawane do lokalizacji punktu w celu oszacowania lokalizacji ekstremum. Z kolei wartość funkcji w punkcie ekstremalnym $D(\hat{x})$ wykorzystuje się do odrzucenia niestabilnych ekstremów charakteryzujących się niskim kontrastem.

Kolejnym aspektem przy określaniu położenia punktów ekstremalnych jest eliminacja wpływu krawędzi obrazu. Zastosowanie różnicy Gaussa doG charakteryzuje się znajdywaniem niestabilnych wyników wzdłuż krawędzi. Dlatego też w celu odrzucenia błędnych punktów stosuje się zasadę mówiącą, że wartość krzywizny w poprzek krawędzi jest niska, natomiast prostopadle do niej jest wysoka. Krzywizny te są obliczane za pomocą hesjanu H, czyli macierzy drugich pochodnych obliczanych jako ilorazy różnicowe:

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix}$$

Wartości własne macierzy H są proporcjonalne do krzywizny funkcji D. Jednak ich bezpośrednie obliczenie nie jest konieczne. Interesuje nas bowiem tylko rząd ich

wielkości, który z kolei można wyznaczyć za pomocą wyznacznika i śladu hesjanu. Oznaczając przez α największą wartość własną, a przez β najmniejszą, wiadoma rzeczą jest, że:

$$Tr(H) = D_{xx} + D_{yy} = \alpha + \beta$$

$$Det(H) = D_{xx}D_{yy} - \left(D_{xy}\right)^2 = \alpha\beta$$

W przypadku ujemnego wyznacznika macierzy H dany punkt obrazu nie może być ekstremalnym ze względu na znaki krzywizny. Zakładając, że istnieje prosta relacja między wartościami własnymi postaci $\alpha = r\beta$ otrzymuje się:

$$\frac{Tr(H)^2}{Det(H)} = \frac{(\alpha+\beta)^2}{\alpha\beta} = \frac{(r\beta+\beta)^2}{r\beta^2} = \frac{(r+1)^2}{r}$$

Wynika z tego, że stosunek ten zależy wyłącznie od współczynnika r, a nie od samych wartości własnych. Prowadzi to do wniosku, że sprawdzenie krzywizny funkcji D polega na spełnieniu nierówności:

$$\frac{Tr(H)^2}{Det(H)} < \frac{(r+1)^2}{r}$$

Orientation assignment

Przypisanie orientacji deskryptorom punktów charakterystycznych jest trzecią częścią metody SIFT. Wykorzystując lokalne cechy obrazu uzyskuje się kolejną zaletę metody, czyli inwariantność względem obrotu. Dla każdego obrazu $L(x, y)$, wygładzonego filtrem Gaussa przy niezmienionej skali, wyznacza się wielkość gradientu $m(x, y)$ oraz orientację $\theta(x, y)$ stosując różnicę wartości pikseli:

$$m(x,y) = \sqrt{(L(x+1,y) - L(x-1,y))^2 + (L(x,y+1) - L(x,y-1)^2}$$

$$\theta(x,y) = arctg\frac{L(x,y+1) - L(x,y-1)}{L(x+1,y) - L(x-1,y)}$$

Następnie za pomocą orientacji gradientu punktu charakterystycznego tworzy się histogram orientacji dla jego sąsiedztwa. Składa się on z 36 słupków pokrywających zakres 360°. Każdy nowy punkt dodany do histogramu jest odpowiednio ważony za pomocą wielkości swojego gradientu oraz poprzez wagowe okno Gaussa o odchyleniu standardowym σ 1.5 razy większym od zastosowanej skali.

Szczyty histogramu orientacji odpowiadają głównym kierunkom gradientu. Stworzenie nowej orientacji punktu następuje w przypadku, gdy odpowiadający mu słupek histogramu ma wysokość nie mniejszą niż 80% wysokości najwyższego słupka. Wynika stąd wniosek, że jedna lokalizacja wielu szczytów o podobnej wielkości gradientu może wyznaczać kilka punktów wysoce charakterystycznych, które będą miały tą samą lokalizację, ale różną orientację. Jak wynika z eksperymentów (Lowe

2004) tylko 15% wszystkich punktów ma przypisane kilka orientacji, co pozytywnie wpływa ma stabilność procesu dopasowania. Dodatkowo w celu zwiększenia dokładności stosuje się aproksymację paraboliczną trzech najbliższych wartości histogramu dla danego punktu.

Keypoint descriptor

Poprzednie trzy etapy metody SIFT służyły wyznaczeniu lokalizacji, skali i orientacji punktu obrazu. Określenie współrzędnych 2D tych punktów zapewniło również ich inwariantność względem przekształceń afinicznych. Ostatnim etapem metody Lowe'a jest wyznaczenie charakterystycznych deskryptorów obrazu niezmienniczych względem zmian oświetlenia i punktu widzenia kamery.

Pierwszym podejściem do powyższego problemu jest utworzenie znormalizowanej macierzy korelacji, która dopasowywałaby regiony wokół deskryptorów o jednakowej skali. Jednakże taka realizacja jest wysoce czuła na zmiany afiniczne i oświetleniowe. Stąd też wykorzystano inne podejście bazujące na mechanizmie reakcji neuronów w korze mózgowej, które wysyłają odpowiedź tylko na sygnały o określonej orientacji i częstotliwości, ale o zmiennej lokalizacji.

Metoda wyznaczenia deskryptorów obrazu składa się z kilku etapów. Początkowo określa się wielkość gradientu i orientację wokół punktu charakterystycznego używając w tym celu skali deskryptora jako miary rozmycia filtrem Gaussa. W celu osiągnięcia inwariantności orientacji współrzędne deskryptora są obracane zgodnie z orientacją punktu.

Następnie każdemu punktowi przypisuje się wagę wielkości gradientu stosując w tym celu okno Gaussa o odchyleniu standardowym σ równym połowie długości okna deskryptora. Przyczyną stosowania okna Gaussa jest eliminacja nagłych zmian postaci deskryptora wywołanych małymi zmianami pozycji próbkowanego okna oraz zmniejszenie wpływu gradientów leżących z dala od środka deskryptora (ang. misregistration errors).

Deskryptor punktu jest tworzony za pomocą histogramu orientacji z regionów o rozmiarze 4×4. Składa się on z ośmiu wektorów o różnych kierunkach, których długość wynika z wielkości histogramu. W ten sposób powstaje macierz histogramów, która zapewnia pewien stopień niewrażliwości na przesunięcia pozycji. Eksperymentalnie udowodniono (Lowe 2004), że najlepsze rezultaty są osiągane dla macierzy o rozmiarze 4×4 z ośmioma orientacjami w każdym histogramie definiując tym samym $4 \times 4 \times 8 = 128$ - elementowy wektor cech dla każdego punktu charakterystycznego.

Ostatecznie wektor cech jest modyfikowany w celu zredukowania wpływu zmian oświetlenia. Początkowo przeprowadza się normalizację wektora do jednostkowej długości, co nie wprowadza zmian w kontraście obrazu. Modyfikacja jasności powodująca, że do wartości każdego piksela dodawana jest stała wartość nie wpływa na

wartość gradientu, ponieważ jest on obliczany jako różnica pikseli, co neutralizuje stałą. Wynika stąd, że deskryptor jest inwariantny względem zmian iluminacji.

Wyszukiwanie obrazów za pomocą metody SIFT

Pierwszym etapem procesu wyszukiwania jest dopasowywanie punktów charakterystycznych do znanych punktów wyekstrahowanych z bazy zdjęć uczących tzw. training images. Wiele spośród tych dopasowań będzie niewłaściwych z powodu wieloznaczności cech wywołanej wpływem zaszumionego tła. Stąd też używa się grupy co najmniej 3 cech do porównywania obiektów, co zwiększa prawdopodobieństwo właściwego dopasowania.

- dopasowywanie punktów

W etapie tym wykorzystuje się metodę najbliższego sąsiada oraz metrykę podobieństwa opartą o odległość Euklidesa. Niestety wiele cech obrazu nie posiada poprawnego dopasowania, ponieważ albo nie zostały znalezione, albo są wynikiem zaszumionego tła. Zwiększenie poprawności dopasowania odbywa się za pomocą techniki porównującej odległość najbliższego do drugiego najbliższego sąsiada, zamiast do pierwszego. Jeżeli wynik tej odległości jest większy niż 0.8 następuje odrzucenie badanego punktu. Okazuje się, że takie podejście daje lepsze rezultaty i eliminuje około 90% błędnych dopasowań.

- efektywne wyznaczanie najbliższego sąsiada

Jako rezultat działania metody SIFT otrzymuje się wektor cech, który w powyższym przypadku ma 128 wymiarów. Tak wysoce wymiarowa przestrzeń stanowi istotny problem obliczeniowy, dla którego nie istnieją skuteczne metody zmniejszania rozmiaru. Dlatego też Lowe (2004) podał algorytm Best-Bin-First (BBF), który zwraca najbliższego sąsiada z wysokim prawdopodobieństwem.

Algorytm BBF wykorzystuje zmodyfikowaną metodę indeksowania k-d tree (patrz rozdział 2.1.4), w której histogram przestrzeni cech jest przeszukiwany w celu znalezienia najbliższej odległości do miejsca zapytania. Eksperymentalnie stwierdzono, że zakończenie procesu wyszukiwania po otrzymaniu pierwszych 200 najbliższych sąsiadów z bazy 100.000 punktów charakterystycznych powoduje stratę tylko 5% spośród poprawnych rezultatów.

- parametry przekształceń afinicznych

W celu poprawnej identyfikacji obiektów obrazu wykorzystuje się grupy zawierające co najmniej 3 charakterystyczne cechy. Każda taka grupa jest następnie wejściem dla procedury weryfikacji geometrycznej, która ma za zadanie znaleźć najlepsze parametry projekcji afinicznej wiążące zdjęcia uczące bazy danych z nowymi obrazami.

Transformata afiniczna punktu $[x\ y]^T$ do punktu obrazu $[u\ v]^T$ może być zapisana w postaci (Lowe 1999):

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} m_1 & m_2 \\ m_3 & m_4 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \end{bmatrix}$$

gdzie $\begin{bmatrix} t_x & t_y \end{bmatrix}^T$ jest wektorem translacji, a parametry rotacji i skali są zawarte wewnątrz zmiennych m_i. Kluczowym zadaniem jest znalezienie parametrów transformacji, dlatego powyższe równanie przekształca się do postaci:

$$\begin{bmatrix} x & y & 0 & 0 & 1 & 0 \\ 0 & 0 & x & y & 0 & 1 \\ & & \cdots & & & \\ & & \cdots & & & \end{bmatrix} \cdot \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \\ t_x \\ t_y \end{bmatrix} = \begin{bmatrix} u \\ v \\ \vdots \end{bmatrix}$$

Równanie to opisuje pojedyncze dopasowanie. Ponieważ do prawidłowego dopasowania potrzebne są co najmniej trzy cechy wprowadza się równanie macierzowe:

$$Ax = b$$

Rozwiązanie tego równania otrzymuje się przez obliczenie wyrażenia:

$$x = [A^T A]^{-1} A^T b,$$

które minimalizuje sumę kwadratów odległości pomiędzy lokalizacją projekcji modelu, a właściwą lokalizacją obrazu.

Finalna decyzja akceptacji, bądź odrzucenia obiektu obrazu bazuje na szczegółowym modelu probabilistycznym. Na podstawie rozmiaru modelu i ilości cech oblicza się oczekiwaną ilość błędnych dopasowań. Następnie wykorzystując metodę Bayesa określa się obecność obiektu na obrazie. Jeżeli wartość tego prawdopodobieństwa jest większa niż 0.98 przyjmuje się, że obiekt jest obecny. W przypadku obrazów tekstury liczba błędnych dopasowań może być znacznie większa, dlatego przyjmuje się, że ilość cech służąca dopasowaniu obrazów nie powinna być mniejsza niż 10.

Metoda Scale Invariant Feature Transform autorstwa Lowe znalazła szerokie zastosowanie w systemach wyszukiwania obrazów CBIR ze względu na swoją skuteczność i efektywność. Poprzez wykorzystanie wielowymiarowych deskryptorów cech inwariantnych względem obrotu, przesunięcia, zmian skali, oświetlenia i punktu widzenia kamery uzyskano metodę odporną na zakłócenia wynikające z szumu tła, czy przysłaniania obiektów. Ta nowatorska technika wyznaczania dużej liczby punktów charakterystycznych obrazu okazała się bardzo skuteczna, co nie zostało jednak okupione wysoką złożonością obliczeniową. Jak podaje autor standardowy komputer klasy PC wystarcza do znalezienia kilku tysięcy punktów obrazu niemal w czasie rzeczywistym. Podsumowując metoda SIFT nadaje się do wykorzystania m. in. w aplikacjach śledzenia ruchu i segmentacji, rekonstrukcji obrazów 3D, czy lokalizacji

robota oraz w wielu innych, wymagających przeprowadzania procesów identyfikacji i dopasowywania obrazów.

Poniżej przedstawiamy przykład skuteczności metody SIFT.

Rys. 2.7 Przykład wyszukiwania obiektów metodą SIFT. Po lewej stronie znajdują się dwa zdjęcia uczące (ang. training images). Po prawej stronie podane zostały wyniki wyszukiwania. Biały duży prostokąt przekształcony w wyniku transformacji afinicznych wyznacza granice oryginalnego zdjęcia uczącego. Mniejsze kwadraty wraz z linią określającą orientację wskazują punkty charakterystyczne użyte w procesie rozpoznawania obiektów (źródło: Lowe 2004).

b) SURF

Interesującym podejściem do problemu porównywania obrazów w systemach CBIR jest metoda zaprezentowana przez Bay, Tuytelaars i Van Gool (2006) o nazwie SURF – ang. Speeded Up Robust Features. Bazuje ona na ogólnej koncepcji SIFT autorstwa Lowe (1999), różniąc się metodologią detekcji punktów charakterystycznych obrazu i konstrukcją detektora, co znacząco wpływa na jej efektywność i skuteczność.

SURF jest interesującym pomysłem stworzenia deskryptora inwariantnego względem obrotu i zmian skali, jednocześnie zachowując wysoką charakterystyczność, powtarzalność detekcji punktów obrazu oraz odporność na jego zakłócenia. Jest to osiągane poprzez:

- wykorzystanie wszystkich pikseli obrazu na potrzeby procesu splotu zdjęć,
- użycie macierzy hesjanu do detekcji punktów wysoce charakterystycznych (ang. Fast-Hesian Detector),
- realizację deskryptorów obrazu w oparciu o funkcje Haara.

W metodzie Bay, Tuytelaars i Van Gool można wyróżnić trzy zasadnicze etapy. Pierwszy - „interest point” polega na znalezieniu charakterystycznych regionów obrazu takich, jak naroża, krawędzie, połączenia typu „T” itp. Cechą zasadniczą tego etapu jest

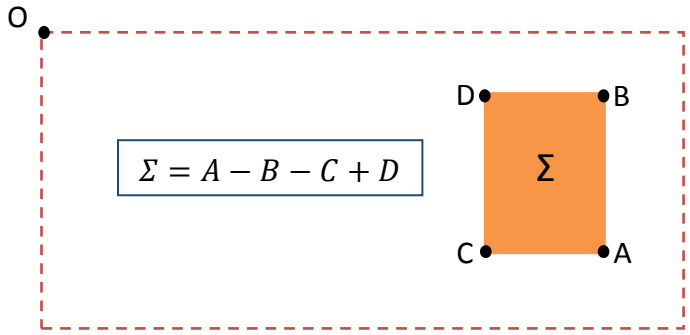
powtarzalność detekcji rozumiana jako zdolność do lokalizacji tych samych punktów przy zmieniających się warunkach otoczenia. Następną częścią jest stworzenie wektora cech reprezentującego całe sąsiedztwo badanego punktu. Taki deskryptor musi być charakterystyczny oraz odporny na zaszumienia obrazu i deformacje geometryczne. Ostatnim etapem jest proces dopasowania deskryptorów pomiędzy różnymi zdjęciami nie wykorzystując przy tym żadnych informacji o ich kolorze. Odbywa się on przy zastosowaniu metryk np. Mahalanobisa, czy Euklidesa. Kluczową rolę odgrywa tutaj wymiar wektora, który bezpośrednio wpływa na skuteczność oraz szybkość metody.

Interest Point Detection

Metoda SURF w swej pierwszej fazie wykorzystuje aproksymację macierzy hesjanu do lokalizacji charakterystycznych punktów obrazu. Do tego celu wykorzystuje się pojęcie całkowitego zdjęcia (ang. integral image), które uskutecznia etap jego splotu z danym filtrem. Całkowite zdjęcie $I_\Sigma(x)$ dla położenia $x = (x, y)^T$ reprezentuje sumę wartości wszystkich pikseli wchodzących w skład obszaru określonego przez to położenie tzn.:

$$I_\Sigma(x) = \sum_{i=0}^{i \le x} \sum_{j=0}^{j \le y} I(i,j)$$

Po wyznaczeniu $I_\Sigma(x)$ pozostają jeszcze trzy dodatkowe operacje do obliczenia sumy natężenia (intensywności) wewnątrz prostokątnego obszaru (rys. 4). Niezależność czasu obliczania od rozmiaru regionu ma szczególne znaczenie ze względu na specyfikę stosowanych filtrów.



Rys. 2.8 Obliczanie sumy natężenia.

Lokalizacja punktu charakterystycznego obrazu jest wyznaczana w oparciu o maksymalną wartość wyznacznika hesjanu. Dla punktu $x = (x, y)$ obrazu I w skali σ macierz hesjanu $H(x, \sigma)$ jest zdefiniowana jako:

$$H(x,\sigma) = \begin{bmatrix} L_{xx}(x,\sigma) & L_{xy}(x,\sigma) \\ L_{xy}(x,\sigma) & L_{yy}(x,\sigma) \end{bmatrix}$$

gdzie $L_{xx}(x,\sigma)$ jest splotem obrazu $I(x,y)$ z drugą pochodną cząstkową funkcji gęstości Gaussa $\frac{\partial^2}{\partial x^2}g(\sigma)$. Parametry $L_{xy}(x,\sigma)$ i $L_{yy}(x,\sigma)$ są obliczane analogicznie. Funkcje Gaussa są optymalne w sensie tworzenia piramidy skali, jednak w praktyce muszą być najpierw dyskretyzowane. Ma to wpływ na powtarzalność detekcji lokalizacji punktów, która zależy od stopnia obrotu obrazu. Bay, Ess i in. (2008) sprawdzili, że detektory zbudowane z macierzy hesjanu mają z natury małą powtarzalność dla obrotu zdjęć o kąt będący nieparzystą wielokrotnością $\pi/3$, natomiast wykazują najlepsze rezultaty przy kącie $\pi/2$, co jest spowodowane kwadratowym kształtem filtrów.

Aproksymacja macierzy hesjanu odbywa się za pomocą specjalnie dopasowanych filtrów kwadratowych. Dzięki zastosowaniu takiej techniki uniezależnia się czas obliczenia od wielkości filtru, co prowadzi do lepszych rezultatów niż tych uzyskiwanych za pomocą Laplasjanu z funkcji Gaussa w metodzie SIFT. Filtry o rozmiarze 9×9 dobrze aproksymują funkcję Gaussa o odchyleniu standardowym $\sigma = 1.2$, tym samym reprezentując najniższy poziom skali obrazu. Oznaczając te aproksymacje przez D_{xx}, D_{yy} i D_{xy} przyjmuje się, że końcową aproksymację hesjanu można wyznaczyć z zależności:

$$\det(H_{approx}) = D_{xx}D_{yy} - (wD_{xy})^2$$

Waga w odpowiedzi filtru jest dobierana w celu równoważenia wyznacznika, co z kolei jest konieczne do zachowania zasady energii pomiędzy funkcjami Gaussa, a ich aproksymacjami. Jest ona równa:

$$w = \frac{\left|L_{xy}(1.2)\right|_F \left|D_{yy}(9)\right|_F}{\left|L_{yy}(1.2)\right|_F \left|D_{xy}(9)\right|_F} \approx 0.9$$

gdzie $|x|_F$ oznacza normę Frobeniusa. Dodatkowo odpowiedzi filtrów są normalizowane zgodnie z ich rozmiarami, co jest zagwarantowane przez stałą wartość normy Frobeniusa.

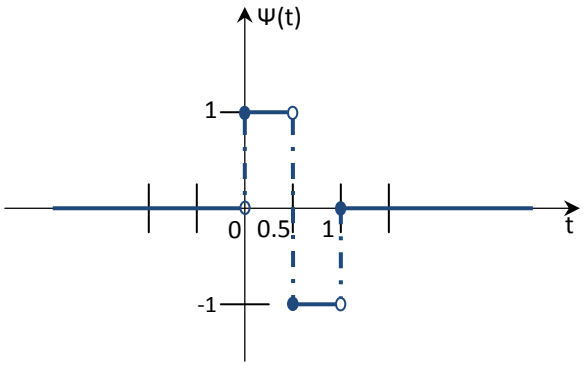
Kolejnym aspektem detekcji punktów charakterystycznych jest konieczność stosowania różnych skali obrazu, co prowadzi do powstawania tzw. piramidy. Dzięki zastosowaniu kwadratowych filtrów nie ma potrzeby rekurencyjnego używania tych samych filtrów do wyników tworzących kolejne warstwy piramidy. W celu utworzenia następnych poziomów zwiększa się rozmiar filtru, natomiast obraz pozostaje niezmieniony. Inaczej mówiąc piramida skali powstaje w wyniku zmiany rozmiaru maski, a nie poprzez iteracyjne zmniejszanie obrazu. Odpowiedź filtru o rozmiarze 9×9 pikseli jest uważana za pierwszą warstwę piramidy, która oznaczana jest poprzez skalę $s = 1.2$ (odpowiednio do pochodnych funkcji Gaussa o odchyleniu standardowym $\sigma = 1.2$). Kolejne warstwy są uzyskiwane poprzez zastosowanie filtrów 15×15, 21×21, 27×27 itd. Każda nowa oktawa, która składa się ze stałej liczby warstw, jest tworzona za pomocą filtrów, których rozmiary zwiększają się dwukrotnie tzn. zamiast

inkrementacji o 6 pikseli, jak ma to miejsce dla nowych warstw, zwiększa się rozmiary o $12, 24, 48$ itd. Wynika stąd, że dla drugiej oktawy stosuje się filtry odpowiednio o rozmiarach $15, 27, 39, 51$, a dla trzeciej oktawy $27, 51, 75, 99$ itd. Ponieważ stosunek filtrów poszczególnych warstw piramidy względem siebie pozostaje niezmieniony, aproksymacje pochodnych funkcji Gaussa są proporcjonalnie skalowane. Oznacza to, że np. filtr o rozmiarze 27×27 pikseli odpowiada skali $\sigma = s = 3 \times 1.2 = 3.6$. Ostateczny proces lokalizacji punktów charakterystycznych odbywa się dla całej piramidy skali poprzez wyznaczenie maksimów wyznacznika hesjanu dla sąsiedztwa punktu o rozmiarze $3 \times 3 \times 3$. Maksima te są następnie interpolowane, co ma istotne znaczenie dla metody, ponieważ różnice skali pomiędzy pierwszymi warstwami każdej oktawy piramidy są relatywnie duże.

Interest Point Description and Matching

Deskryptor obrazu zaadoptowany na potrzeby metody SURF opisuje dystrybucję natężenia zawartości określonej przez sąsiedztwo punktu charakterystycznego. Jest to podejście podobne do gradientów wykorzystywanych w metodzie SIFT. W tym celu używa się 64-wymiarowego wektora cech oraz wyznacza się rozkład odpowiedzi funkcji Haara pierwszego rzędu w kierunku x i y. Definicja i wykres podstawowej falki Haara są następujące:

$$\psi(t) = \begin{cases} 1, & dla \quad 0 \le t < 0.5 \\ 0, & dla \quad t < 0 \; i \; t \ge 1 \\ -1, & dla \quad 0.5 \le t < 1 \end{cases}$$



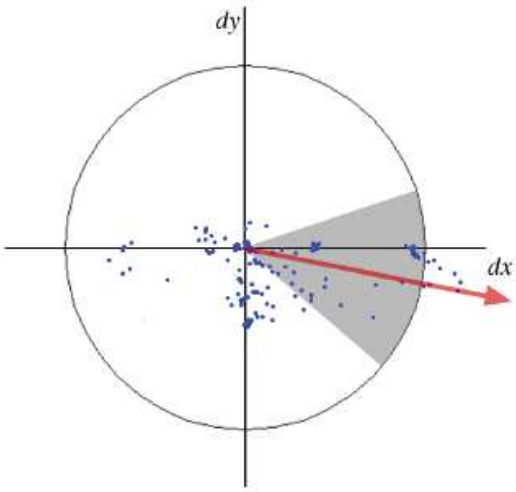
Indeksowanie realizowane jest w oparciu o znak drugich pochodnych Laplasjanu, co nie tylko zwiększa odporność deskryptora, ale także czas procesu dopasowania.

Całościowo etap konstrukcji i dopasowania deskryptora można podzielić na trzy części:

- przypisanie orientacji

W celu zachowania inwariantności deskryptora względem obrotu oblicza się odpowiedź falki Haara w kierunku x i y poprzez wyznaczenie sąsiedztwa punktu charakterystycznego w postaci okręgu o promieniu równym $6s$, gdzie s jest skalą

detekcji tego punktu. Jak wynika z badań Bay, Ess i in. (2008) potrzeba tylko sześciu operacji do wyznaczenia tych odpowiedzi dla dowolnej skali. W celu zachowania spójności obliczeń przyjmuje się, że rozmiar falki jest uzależniony od skali i wynosi $4s$. Dzięki temu można posłużyć się pojęciem całkowitego zdjęcia do szybkiego filtrowania. Po obliczeniu i modyfikacji funkcją Gaussa ($\sigma = 2s$) odpowiedzi falki Haara przedstawia się jako punkty układu współrzędnych, gdzie oś odciętych oznacza ich siłę w kierunku poziomym, a oś rzędnych w kierunku pionowym. Orientacja deskryptora jest szacowana za pomocą sumy wszystkich odpowiedzi znajdujących się w przesuwanym oknie o rozmiarze $\pi/3$. Te zsumowane odpowiedzi tworzą lokalny wektor orientacji, a najdłuższy z nich definiuje końcową orientację deskryptora. Dokładnie przedstawia to poniższy rysunek:

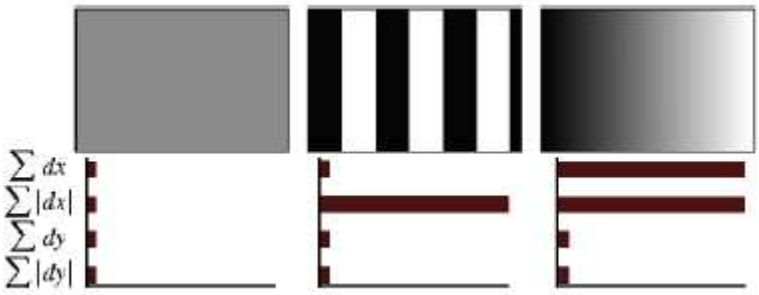


Rys. 2.9 Wyznaczanie orientacji deskryptora metody SURF za pomocą okna o rozmiarze $\pi/3$ w okrągłym sąsiedztwie punktu charakterystycznego (źródło: Bay, Ess, Tuytelaars, Van Gool 2008).

- deskryptor jako suma odpowiedzi falek Haara

Pierwszym krokiem do wyznaczenia deskryptora jest stworzenie kwadratowego okna o rozmiarze $20s$ wokół charakterystycznego punktu obrazu. Okno to dzieli się na mniejsze kwadraty o rozmiarze 2×2 (ang. sub-regions). Dla każdego takiego regionu wyznacza się odpowiedzi falki Haara w obu kierunkach oznaczając je odpowiednio d_x (kierunek poziomy) i d_y (kierunek pionowy). W celu zwiększenia odporności na zniekształcenia geometryczne i błędy lokalizacji, odpowiedzi d_x i d_y są ważone filtrem Gaussa o odchyleniu standardowym $\sigma = 3.3s$. Następnie są one sumowane w obrębie sub-regionu tworząc pierwszy zbiór wejść wektora cech. Wektor ten jest uzupełniany wartościami $|d_x|$ i $|d_y|$, co ma na celu dostarczenie informacji o polaryzacji zmian natężenia. Wynika stąd, że każdy taki region jest opisany za pomocą 4-wymiarowego

deskryptora $v = (\sum d_x, \sum d_y, \sum|d_x|, \sum|d_y|)$. Łącząc te deskryptory w obrębie całego okna otrzymuje się 64-wymiarowy wektor cech, który jest niezmienniczy względem zmian oświetlenia. Poniżej przedstawiamy przykład ilustrujący własności deskryptora dla trzech obrazów o różnym wzorze i natężeniu.



Rys. 2.10 Wartości deskryptora dla regionów o różnym wzorze. Dla jednolitego obrazu po lewej wszystkie cztery odpowiedzi Haara są równe. W przypadku występowania częstotliwości w kierunku x - obraz środkowy – tylko odpowiedź $\sum|d_x|$ jest wysoka. Wzrost intensywności obrazu po prawej w kierunku x powoduje duże wartości $\sum d_x \; i \; \sum|d_x|$ (źródło: Bay, Ess, Tuytelaars, Van Gool 2008).

- szybkie indeksowanie i dopasowanie

Proces indeksowania odbywa się poprzez wyznaczenie znaku Laplasjanu (tj. śladu macierzy hesjanu) dla danego punktu charakterystycznego. Znak ten umożliwia odróżnienie jasnych obiektów na ciemnym tle od odwrotnej sytuacji. Cecha ta nie wymaga dodatkowego nakładu obliczeniowego, ponieważ już wcześniej została wyznaczona w fazie detekcji. Etap dopasowania polega na porównywaniu deskryptorów dwóch obrazów. Do tego celu wykorzystuje się tylko cechy, które charakteryzują się jednakowym poziomem kontrastu. Wynika stąd, że już minimalna ilość informacji wystarcza do szybkiego przeprowadzenia procesu, co skutkuje niedużym zwiększeniem efektywności.

Metoda Speeded Up Robust Features autorstwa Bay, Tutyelaars i Van Gool (2006) jest ciekawym podejściem do problemu wyszukiwania i dopasowywania obrazów. Bazując na metodologii metody SIFT (Lowe 1999) koncentruje się na analizie informacji zawartych w przestrzennym rozkładzie gradientów obrazu. Jednak, jak podają Bay, Ess i in. (2008) SURF przewyższa skuteczność SIFT, co wynika z jej mniejszej wrażliwości na zaszumienie zdjęcia. Kluczowy okazuje się również kompromis pomiędzy wymiarowością wektora cech, a szybkością obliczenia algorytmu. W zaprezentowanej powyżej metodzie SURF użyto 64-wymiarowego wektora, który był zbudowany w oparciu o podział sąsiedztwa punktu

charakterystycznego na regiony o rozmiarze 4×4. Zastosowanie podziału 3×3 powodowało zmniejszenie wektora do wymiaru 36, co nieznacznie pogarszało wyniki, jednocześnie zwiększając szybkość procesu dopasowania. W celu porównania Bay, Ess i in. (2008) przetestowali również skuteczność większego wektora o rozmiarze 128. Dokonali tego poprzez podział sumd_x i $|d_x|$ na przypadki, gdy $d_y < 0$ i $d_y \geq 0$ (analogiczne postępowanie zastosowano dla d_y i $|d_y|$). Okazało się, że uzyskany deskryptor był jeszcze bardziej charakterystyczny nie powodując zbyt dużego zwolnienia prędkości obliczenia, jednak etap dopasowania okazywał się być o wiele wolniejszy, co bezpośrednio wynikało z wysokiej wymiarowości wektora cech.

2.2.4 Percepcyjne grupowanie

Analiza obrazu w kontekście jego zawartości jest kluczowym zadaniem stawianym przed każdym systemem CBIR. Oprócz podstawowych technik opisu zdjęcia wykorzystujących jego podstawowe cechy (np. deskryptory koloru, kształtu, etc.) oraz bardziej zaawansowanych matematycznie, jak np. sprzężenie zwrotne, można wyróżnić grupę metod opierających się na podstawach percepcyjnego rozumowania człowieka. To złożone i bardzo trudne podejście ma na celu zamodelowanie zdolności wizualnych człowieka do procesu wydobywania znaczących relacji pomiędzy różnymi obiektami obrazu, bez uwzględnienia żadnych informacji o jego zawartości. W początkowej fazie rozwoju technik rozumowania określono następujące koncepcje percepcyjnego grupowania (Lowe 1985):

- bliskość, sąsiedztwo (ang. proximity),

- podobieństwo (ang. similarity),

- kontynuacja (ang. continuation),

- zamknięcie (ang. closure),

- symetryczność (ang. symmetry).

W niniejszym rozdziale zostanie przedstawiona interesująca metoda autorstwa Iqbal i Aggarwal (1999), która wykorzystuje metody percepcyjnego grupowania do wyszukiwania i klasyfikacji obrazów zawierających duże obiekty architektoniczne takie, jak budynki, wieże, mosty itp. Całość ich pracy opiera się na przekonaniu, że system wizualny człowieka działa na zasadzie tworzenia struktur wysokiego poziomu opartych na percepcyjnym grupowaniu cech obrazu niskiego rzędu. Struktury te służą następnie do opisu kształtu obiektu, którego lokalizacja nie musi być dokładnie określona. Dzięki temu założeniu proces segmentacji oraz szczegółowej reprezentacji obiektu nie jest wymagany.

Detekcja obiektów architektonicznych odbywa się przy wykorzystaniu metod percepcyjnego grupowania za pomocą hierarchicznego wyznaczania następujących elementów obrazu:

- segmenty prostych linii,
- długie linie,
- zakończenia linii,
- połączenia typu „L”,
- połączenia typu „U”,
- znaczące grupy równoległe,
- graf zakończeń i wielokąty.

Powyższe cechy zostały uznane za najbardziej odpowiednie do rozpoznania obiektów architektonicznych i to one służą do zbudowania końcowego trójwymiarowego deskryptora obrazu. Przestrzeń wektora została ściśle wyznaczona poprzez podział wszystkich zdjęć na trzy klasy:

- zdjęcia zawierające struktury wykazujące charakterystyczne cechy dla obiektów architektonicznych,

- zdjęcia bez powyższych struktur,

- zdjęcia o średniej ilości struktur.

Zagadnienie percepcyjnego grupowania zostało zdefiniowane przez Iqbal i Aggarwal (2002) jako proces wykorzystania regularności obrazów wejściowych, które nie są przypadkowe (ang. non-accidental rule). Innymi słowy regularności na zdjęciu faktycznie mają swoje odpowiedniki w fizycznym świecie. Dzięki temu wyciągnięto wniosek mówiący, że poszczególna konfiguracja cech obrazu odpowiada określonej jego strukturze. Na przykład koncepcja sąsiedztwa (ang. proximity) może zostać wytłumaczona poprzez fakt, że jeżeli dwie cechy są przyległe na zdjęciu, to są one również przyległe w świecie 3D.

Wydobywanie istotnych cech obrazu

Obiekty architektoniczne cechują się ostrymi krawędziami i prostolinijnymi ograniczeniami. Posiadają zatem dużą liczbę połączeń, linii równoległych, krawędzi i wierzchołków. Te struktury są generowane poprzez obecność takich elementów, jak np. okna, drzwi, czy zewnętrzne ograniczenia budynków. Stąd też powyższe cechy zostały uznane za najbardziej odpowiednie do analizy obrazów architektonicznych.

Poniżej przedstawiamy dokładny opis detekcji charakterystycznych cech obrazu (Iqbal, Aggarwal 2002).

- Segmenty prostych linii

Wyznaczanie segmentów prostych linii na zdjęciu jest pierwszym etapem hierarchicznego procesu detekcji cech. Jest to stosunkowo łatwe zadanie, które może zostać zrealizowane za pomocą np. filtrów Sobela, Prewitta, czy Canny'ego. Iqbal i Aggarwal (2002) wykorzystali do tego celu detektor linii Burns'a, który opiera się na orientacji lokalnych gradientów sąsiedztwa danego piksela. Lokalizacja oraz cechy krawędzi są określane za pomocą struktur skojarzonych powierzchni intensywności (Burns i in. 1986). Wadą operatora Burns'a jest generowanie błędnych linii wynikających z lokalnych wariacji intensywności obrazu. Stąd też wprowadza się pojęcie tzw. „siły krawędzi" (ang. edge strength), która ma być wyższa od zadanego progu δ_e:

$$\varepsilon(L_i) > \delta_e$$

gdzie $\varepsilon(\cdot)$ oznacza stosunek siły krawędzi skojarzony z daną linią, do maksymalnej siły krawędzi dla całego obrazu. Siła ta jest obliczana jako średnia wielkość gradientu w określonym regionie. W praktyce detekcja powyższych segmentów nie działa prawidłowo, dlatego proces ten jest dzielony na etapy przedstawione poniżej.

- Długie linie

Segmenty prostych linii wyznaczone za pomocą operatora Burns'a są w praktyce podzielone na mniejsze części i muszą być łączone w celu określenia długiej linii. W tym celu poszukuje się tzw. linii reprezentatywnej (ang. representative line), która łączy w sobie zbiór podobnie zorientowanych i leżących blisko siebie linii zgodnie z koncepcją bliskości, podobieństwa i kontynuacji percepcyjnego grupowania. Region o długości $\delta_f = 2\delta_n$ zawierający segment linii L_b w swojej środkowej części jest wykorzystywany do wyznaczenia zbioru segmentów S_{f_e}, który następnie jest zastępowany linią reprezentatywną L_r w przypadku spełnienia poniższych warunków:

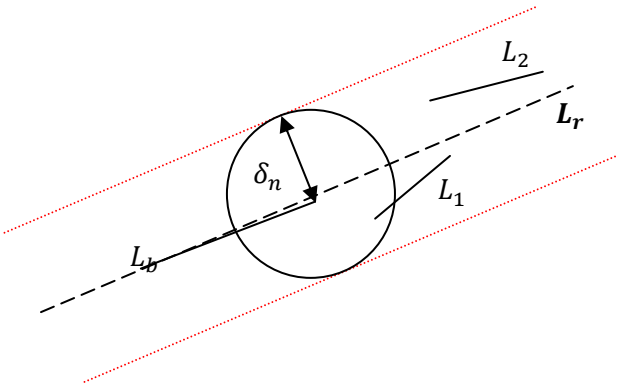
a) $A(L_b, L_i) < \delta_a$ i $max\{D_o(L_b, e_{i_1}), D_o(L_b, e_{i_2})\} < \delta_n$

lub

b) $D(e_i, e_j) < \delta_n$ lub $\vartheta\left(\varphi_{L_j}(L_i), L_j\right) > 0$

gdzie: L_b - podstawowy segment linii, L_i i L_j - dwa dowolne segmenty zbioru S_{f_e}, $A(\cdot)$ - wartość bezwzględna najmniejszego kąta między dwoma segmentami, δ_a - próg. Zmienne e_i i e_j są końcowymi punktami linii L_i i L_j, $D_o(\cdot)$ jest ortogonalną odległością pomiędzy końcowym punktem i segmentem, a $D(\cdot)$ określa odległość pomiędzy końcowymi punktami linii L_i i L_j, które są najbliżej siebie. Dodatkowo przez $\varphi_{L_j}(L_i)$ oznaczono ortogonalny rzut linii L_i na L_j, a przez $\vartheta(\cdot)$ długość zachodzenia na siebie dowolnych dwóch linii.

Schemat wyznaczania długich linii dokładnie ilustruje poniższy rysunek:



Rys. 2.11 Sposób wyznaczania linii reprezentatywnej (źródło: Iqbal, Aggarwal 2002).

W celu wyznaczenia linii reprezentatywnej potrzebna jest wiedza o jej środkowym punkcie, orientacji oraz długości. Środkowy punkt i orientacja linii L_r są określane w postaci średniej ważonej punktów środkowych i orientacji wszystkich linii tworzących zbiór S_{f_e}. Długość L_r to odległość pomiędzy dwoma najdalszymi punktami ortogonalnych rzutów wszystkich linii zbioru S_{f_e} na linię reprezentatywną. Po wyznaczeniu długich linii na obrazie mogą pozostać jeszcze mniejsze krawędzie, które eliminuje się poprzez zapewnienie odpowiedniej długości większej od zadanego progu: $\mathcal{L}(L_i) > \delta_l$.

- Zakończenia linii

Wyznaczenie punktu zakończenia jest realizowane za pomocą prostego regionu sąsiedztwa. Dla dowolnej pary dwóch linii $\{L_i, L_j\}$ muszą zachodzić poniższe warunki:

$$\delta_c \leq \theta \leq \pi - \delta_c \quad oraz \quad max\{abs(d_y(e_i, e_j)), abs(d_x(e_i, e_j)\} \leq \delta_n$$

gdzie θ wyznacza kąt pomiędzy L_i i L_j, a δ_c jest danym kątem podobieństwa.

- Połączenia typu „L”

Krawędź typu „L” powstaje przez zakończenie dwóch linii pod kątem zbliżonym do $\pi/2$. Dla każdej pary $\{L_i, L_j\}$, gdzie $i \neq j$ połączenie typu „L” musi spełniać warunki:

$$D(e_i, e_j) < \delta_n \quad oraz \quad \frac{\pi}{2} - A(L_i, L_j) < \delta_{l_a}$$

gdzie δ_{l_a} jest ustaloną wartością.

- Połączenia typu „U”

Połączenie typu „U” jest realizowane tylko za pomocą dwóch połączeń typu „L”. Tego typu krawędź może być dowodem występowania na obrazie obiektów w kształcie okien, bądź drzwi. Wykrywanie połączeń typu „U” odbywa się poprzez spełnienie następujących warunków:

a) $A(L_{11}, L_{ur}) < \delta_{ua} \;\; i \;\; A(L_{21}, L_{ur}) < \delta_{ua}$- stała wartość

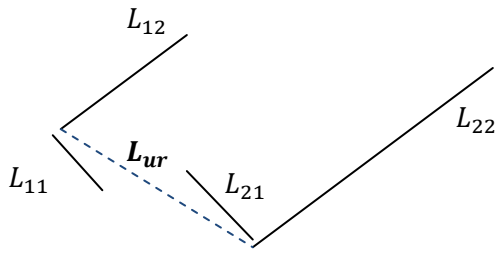
lub

b) $D(e_{11}, e_{21}) < 2\delta_n \;\; lub \;\; \vartheta\left(\varphi_{L_{21}}(L_{11}), L_{21}\right) > 0$

i

c) $L_{12}\, i\, L_{22}$ wskazują ten sam kierunek,

przy czym $L_{11}\, i\, L_{21}$ są liniami tworzącymi dwa połączenia typu „L", które również wskazują ten sam kierunek. Linie $L_{12}\, i\, L_{22}$ są pozostałymi liniami tego połączenia, a L_{ur} jest linią reprezentatywną. Dokładnie wyjaśnia to prosty schemat:



Warto zwrócić uwagę, że połączenie typu „U" powstałe z jednego połączenia typu „L" i dodatkowej linii nie jest możliwe, ponieważ taka pojedyncza linia musiałaby znajdować się blisko połączenia „L" i w ten sposób sama formowałaby takie połączenie.

- Znaczące grupy równoległe

Grupa równoległa jest zbiorem linii $S_{pg} = \{L_1, L_2, \dots, L_M\},\ M \geq 2,$ które spełniają następujące warunki:

a) linie $L_i\, i\, L_j$ mają podobne długości tzn.

$$\frac{\mathcal{L}(L_i)}{\mathcal{L}(L_j)} > \delta_{pg_1} - \; zadany\; próg$$

b) linie $L_i\, i\, L_j$ są relatywnie blisko siebie tzn.

$$\frac{D(e_{mid_i}, e_{mid_j})}{\mathcal{L}_{avg}(L_i, L_j)} < \delta_{pg_2}$$

gdzie e_{mid_i}, e_{mid_j} są środkowymi punktami odpowiadających linii $L_i\, i\, L_j$, a $\mathcal{L}_{avg}(\cdot)$ jest ich średnią długością,

c) linie $L_i\, i\, L_j$ zachodzą na siebie w następujący sposób:

$$\frac{\vartheta\left(\varphi_{y_{axis}}(L_i), \varphi_{y_{axis}}(L_j)\right)}{\mathcal{L}(\varphi_{y_{axis}}(L_i))} > \delta_{pg_3}$$

$$\frac{\vartheta\left(\varphi_{x_{axis}}(L_i), \varphi_{x_{axis}}(L_j)\right)}{\mathcal{L}(\varphi_{x_{axis}}(L_i))} > \delta_{pg_3}$$

$$\frac{\vartheta\left(\varphi_{L_j}(L_i), L_j\right)}{\mathcal{L}(\varphi_{L_j}(L_i))} > \delta_{pg_3}$$

gdzie $y_{axis}\ i\ x_{axis}$ reprezentują odpowiednie osie układu współrzędnych skojarzonego z obrazem. Powyższe warunki wykorzystują koncepcje bliskości, podobieństwa i równoległości percepcyjnego grupowania, które zapewniają formułowanie grup linii o podobnej długości i orientacji.

Niektóre spośród powyższych grup określające charakterystyczne kształty architektoniczne, jak np. okna, czy drzwi są uznawane za znaczące i dlatego muszą spełniać dodatkowe założenia postaci:

$$A(L_i, y_{axis}) < \delta_{spga} \quad lub \quad A(L_i, x_{axis}) < \delta_{spga}$$

Próg δ_{spga} przyjmuje wartość równą $\pi/4$, co zapewnia detekcję grup równoległych w dowolnej orientacji.

- Graf zakończeń i wielokąty

Wielokąty są to zamknięte figury utworzone za pomocą nierównoległych linii. Detekcja zamkniętych figur może odbywać się za pomocą śledzenia punktów zakończeń linii. Może się to zdarzyć, gdy startując z jednego punktu jesteśmy w stanie do niego powrócić idąc wzdłuż danych linii i ich zakończeń. Jednak złożoność obliczeniowa takiej metody rośnie ekspotencjalnie, dlatego też stosuje się podejście wykorzystujące teorię grafów. W celu wyznaczenia zamkniętych figur określa się graf zakończeń linii, a następnie za pomocą poniższych warunków decyduje się czy faktycznie tworzy on poprawny wielokąt.

Graf zakończeń rozumiany jest jako zbiór $G = \{V, E\}$, gdzie V jest zbiorem wierzchołków, a E jest zbiorem krawędzi. Niech $\overline{e_{ij}} \in E$ będzie krawędzią łączącą wierzchołki $\widetilde{v_i}, \widetilde{v_j} \in V$. Waga przypisana do $\overline{e_{ij}}$ jest definiowana w postaci:

$$w\left(\overline{e_{ij}}\right) = \deg(\widetilde{v_i}) + \deg(\widetilde{v_j})$$

gdzie $\deg(\cdot)$ oznacza stopień wierzchołka, czyli ilość krawędzi z nim związanych. Następnie tworzy się symetryczną macierz sąsiedztwa i określa stopnie dla wszystkich wierzchołków grafu. Wagą dowolnego drzewa rozpinającego (ang. spanning tree) jest suma wag wszystkich gałęzi tworzących to drzewo. Maksymalne drzewo rozpinające

jest używane do określenia podstawowych obwodów, które z kolei reprezentują zamkniętą figurę na obrazie.

Wielokąt P definiuje się jako podstawowy obwód, który spełnia następujące wymagania:

a) wielokąt jest prosty tzn. jego krawędzie nie przecinają się ze sobą,
b) jest on stosunkowo zwarty tzn.

$$\mathcal{L}(P) \leq \delta_{\mathcal{L}} \mathcal{L}(conv(P))$$

gdzie $conv(P)$ oznacza wypukłą powłokę (otoczkę) wielokąta P, a $\delta_{\mathcal{L}}$ jest odpowiednio dobranym progiem. $\mathcal{L}(\cdot)$ jest zdefiniowany w postaci:

$$\mathcal{L}(P) = \frac{obwód^2(P)}{pole\ powierzchni\ (P)}$$

c) nie zawiera wielu otworów:

$$n_i \leq \delta_{cv} n_{cv}$$

gdzie n_i jest liczbą wierzchołków wielokąta wewnątrz otoczki $conv(P)$, n_{cv} jest liczbą wierzchołków znajdujących się na $conv(P)$, a $\delta_{cv} \leq 1$ oznacza stałą,

d) ilość krawędzi wielokąta nie przekracza progu δ_{ne}.

W procesie klasyfikacji i identyfikacji obrazów wykorzystuje się trójwymiarowy wektor cech $X = (\widetilde{x_1}, \widetilde{x_2}, \widetilde{x_3})^T$ określający współrzędne przestrzeni obrazu, gdzie:

$$\widetilde{x_1} = \frac{liczba\ linii\ w\ połączeniach\ typu\ "L"}{całkowita\ liczba\ wyznaczonych\ linii}$$

$$\widetilde{x_2} = \frac{liczba\ linii\ w\ połączeniach\ typu\ "U"}{całkowita\ liczba\ wyznaczonych\ linii}$$

$$\widetilde{x_3} = \frac{liczba\ linii\ tworzących\ znaczące\ grupy\ równoległe\ i\ wielokąty}{całkowita\ liczba\ wyznaczonych\ linii}$$

Jak podano wcześniej wszystkie zdjęcia użyte w eksperymentach zostały podzielone na trzy klasy w zależności od stopnia występowania struktur o charakterze architektonicznym. Zostały one oznaczone odpowiednio jako $\Omega_1, \Omega_2, \Omega_3$. Każda z tych trzech klas posiada własną funkcję g_1, g_2, g_3, która służy do przypisania danego obrazu do odpowiadającej mu klasy. Odbywa się to poprzez wyznaczenie dla danego wektora X znaku nierówności:

$$g_i(X) > g_j(X)\ dla\ i \neq j, i, j \in \{1, 2, 3\}$$

Przypisanie zdjęcia do danej klasy odbywa się za pomocą algorytmu k-najbliższych sąsiadów. Następnie za pomocą miary $d(X, X_{kn})$ wyznacza się odległość pomiędzy testowym wektorem cech X, a n-tym wektorem uczącym X_{kn} należącym do k-tej klasy. Miara ta może być zdefiniowana w postaci normy:

$$d(X, X_{kn}) = \|X - X_{kn}\| = \sqrt{(X - X_{kn})^T (X - X_{kn})}$$

Iqbal i Aggarwal (2002) przeprowadzili eksperymenty skuteczności swojej metody na dwóch bazach danych o łącznej liczbie 2661 zdjęć. Otrzymany przez nich ogólny współczynnik wyszukiwania był rzędu 73,52%, Jednak efektywność percepcyjnego grupowania w kontekście pojęć precyzji (ang. precision) i powtarzalności lub kompletności (ang. recall) różniła się znacząco w obrębie trzech klas obrazów. Można to zaobserwować analizując poniższą tabelę.

Tab. 2.1 Skuteczność wyszukiwania w kontekście pojęć precyzji i kompletności. Baza danych 491 zdjęć o rozdzielczości 512×512 pikseli (źródło: Iqbal, Aggarwal 2002).

Klasa	Całkowita liczba zdjęć	Zdjęcia wyszukane	Poprawne wyniki	Kompletność [%]	Precyzja [%]
strukturalna	255	235	198	77.65	84.26
nie strukturalna	140	143	114	81.43	79.72
średnio strukturalna	96	113	49	51.04	43.36

Metody percepcyjnego grupowania oraz sposobu postrzegania obrazu przez człowieka spotkały się z szerokim zainteresowaniem w świecie analizy i przetwarzania obrazów (Jiang, Ngo, Tan 2006; Fu, Chi, Feng 2006). Metodologia Iqbal i Aggarwal (1999) bazuje na semantycznej relacji pomiędzy różnymi prymitywnymi cechami niskiego poziomu. Poprzez wykorzystanie trzech klas struktur uniezależniono się od konieczności stosowania procesów segmentacji i określania lokalizacji szukanych obiektów. Dodatkowo przeprowadzone badania pokazały, że efektywność adaptacji reguł percepcyjnych do wyszukiwania zdjęć architektonicznych może być bardzo duża.

Rozdział 3. Przegląd systemów Content-Based Image Retrieval

Głównym problemem stawianym przed systemami Content-Based Image Retrieval jest efektywne oraz szybkie wyszukiwania pożądanych informacji. Jest to jedno z kluczowych zagadnień zarządzania multimedialnymi bazami danych, które z roku na rok stają się coraz większe.

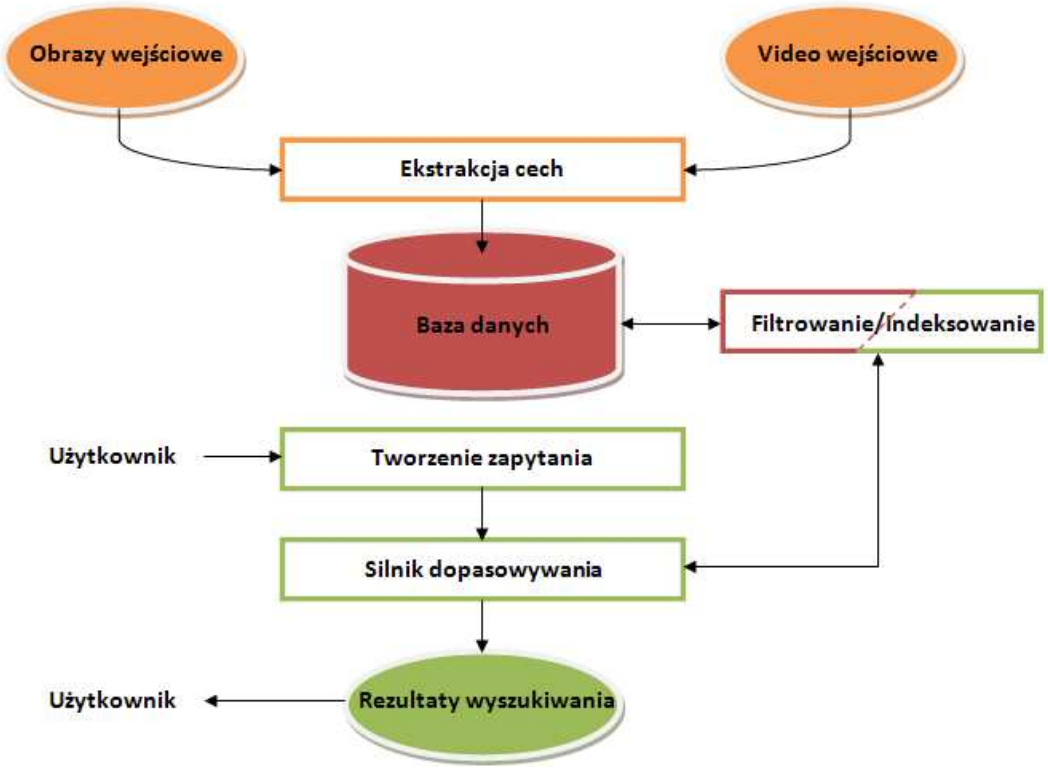
W celu rozwiązania powyższej sytuacji powstało wiele komercyjnych i badawczo-rozwojowych systemów wyszukiwania informacji MIR (ang. Multimedia Information Retrieval – Lew, Sebe i in. 2006), wśród których dużą grupę reprezentują systemy CBIR koncentrujące się na sposobach indeksowania, porównywania i wyszukiwania obrazów cyfrowych. W swojej pierwotnej wersji wyszukiwanie obrazów odbywało się poprzez ich uzupełnianie za pomocą komentarzy, a następnie wykorzystywanie algorytmów tekstowych wyszukiwania informacji (Chang, Fu 1980). Jednak takie podejście było związane z dwoma kluczowymi problemami (Rui, Huang, Chang 1999): ogromną ilością pracy w celu podpisania wszystkich zdjęć bazy danych oraz subiektywistyczną oceną ich zawartości. Stąd też w latach dziewięćdziesiątych XX w. pojawiła się nowa technika bazująca na opisie danych w kontekście ich zawartości wizualnych nazwana CBIR. Od tego czasu nastąpił olbrzymi rozwój dziedziny analizy obrazu. Powstało wiele nowatorskich technik wykorzystujących początkowo metody bez interakcji użytkownika z systemem np. porównywanie deskryptorów koloru, kształtu, tekstury, a następnie metody interakcyjne wykorzystujące modyfikacje zapytania poprzez sprzężenie zwrotne (ang. relevance feedback).

Celem niniejszego rozdziału jest prezentacja popularnych systemów komercyjnych i badawczo-rozwojowych, które miały znaczący wpływ na rozwój obecnych aplikacji CBIR. Skoncentrujemy się na ich funkcjonalności i technicznych aspektach związanych ze sposobami formułowania zapytań z wykorzystaniem sprzężenia zwrotnego, metodami indeksowania danych oraz dopasowywania podobieństwa. W drugiej części rozdziału podamy przykłady obecnych systemów, które wykorzystując rozwój narzędzi informatycznych przetwarzania i analizy obrazu znajdują zastosowania w takich dziedzinach, jak medycyna diagnostyczna, systemy wizyjne w robotyce, bezpieczeństwo czy zapewnianie jakości.

3.1 QBIC

System QBIC (ang. Query By Image Content) został stworzony w 1993 roku przez centrum badawcze IBM Almaden Research Center w USA (Niblack, Barber, Equitz i in. 1993). Jest to pierwszy komercyjny system CBIR, którego konstrukcja i

wykorzystane techniki miały olbrzymi wpływ na późniejsze systemy. Podstawowy schemat architektury systemu jest następujący:



Rys. 3.1 Architektura systemu QBIC (źródło: Niblack, Barber, Equitz i in. 1993).

QBIC wykorzystuje następujące cechy koloru (Veltkamp, Tanase 2000): wektor koloru obiektu bądź całego obrazu w przestrzeni Munsell'a, RGB i CIE L*a*b*, 256-wymiarowy histogram. W przypadku, gdy x jest n-wymiarowym histogramem i $C = [c_1, c_2, \dots, c_n]$ jest macierzą $3 \times n$ o kolumnach reprezentujących wartości RGB, średni wektor koloru ma postać $C \cdot x$. W celu opisu tekstury używane są zmodyfikowane cechy teksturowe Tamury tzn. gruboziarnistość (ang. coarness), kontrast (ang. contrast) i kierunkowość (ang. directionality). Opis kształtu składa się z pola powierzchni i kolistości obliczanych z macierzy kowariancji pikseli krawędzi obiektów oraz z algebraicznych niezmienników momentowych.

Cechą charakterystyczną systemu jest możliwość wyszukiwania obrazów na podstawie szkicu użytkownika. Zdjęcia w bazie danych są reprezentowane poprzez zredukowaną mapę binarną punktów krawędziowych. W tym celu obraz przekształcany jest do rozmiaru 64×64 piksele, a krawędzie obiektów wyznaczane są za pomocą filtru Canny'ego.

Dopasowywanie obrazów odbywa się przy wykorzystaniu koloru i wagowej metryki Euklidesa, gdzie wagi są odwrotnością odchylenia standardowego każdej składowej. W przypadku porównywania histogramów wykorzystuje się dwie miary: łatwą do obliczenia odległość między średnim kolorem oraz trudniejszą kwadratową. Pierwszą wyznacza się z zależności:

$$d^2_{avg}(x,y) = \left(x_{avg} - y_{avg}\right)^T \left(x_{avg} - y_{avg}\right),$$

a drugą przy wykorzystaniu symetrycznej macierzy podobieństwa koloru A tj.

$a_{ij} = 1 - d_{ij}/d_{max}$, gdzie d_{ij} jest odległością Euklidesa pomiędzy kolorami i oraz j w przestrzeni RGB, a $d_{max} = max_{i,j} d_{ij}$:

$$d^2_{hist} = (x-y)^T (x-y)$$

Dodatkowo podobieństwo mierzy się również dla tekstury i kształtu.

System QBIC jako pierwszy wykorzystał wielowymiarowe indeksowanie w celu zwiększenia prędkości działania. Średni kolor oraz cechy tekstury są indeksowane za pomocą algorytmu $R^* - tree$. Dodatkowo każdy wynik wyszukiwania może zostać użyty jako podstawa do kolejnego, co świadczy o występowaniu elementów sprzężenia zwrotnego w systemie (http://wwwqbic.almaden.ibm.com).

3.2 VIR Image Engine

VIR Image Engine jest to produkt grupy Virage Inc. utworzony w 1996 r. Koncepcja systemu jest podobna do QBIC tzn. wykorzystuje podstawowe cechy obrazu takie, jak średni kolor globalny i lokalny, tekstura, kształt. Aplikacja umożliwia dodatkowo tworzenie własnych nieskomplikowanych, podstawowych funkcji do wyszukiwania obrazów, przy jednoczesnym zdefiniowaniu miary podobieństwa służącej do ich porównywania (Bach, Fuller, Gupta i in. 1996). W kontekście tworzenia zapytań VIR Image Engine dostarcza użytkownikowi zbiór narzędzi GUI (ang. Graphical User Interface), który zawiera kilka udogodnień przydatnych w procesie włączania zdjęć do bazy danych, tworzenia zapytań, zmian wag, czy też wspiera kilka popularnych formatów obrazu. Dodatkowo możliwe jest modyfikowanie istniejących zdjęć i konstruowanie zapytań poprzez szkic, dzięki narzędziom do rysowania i kolorowania za pomocą dostępnej palety barw.

W przypadku porównywania dwóch obrazów za pomocą własnych funkcji wynik podobieństwa jest wyznaczany w oparciu o miarę odległości związaną z daną cechą. Pojedyncze wyniki są następnie łączone poprzez zbiór wag w strukturę wyniku, która umożliwia ponowne przeliczenie przy zmienionych wagach. System nie jest

wyposażony w sprzężenie zwrotne, ale umożliwia modyfikowanie wag dla tego samego zapytania (Veltkamp, Tanase 2000).

VIR Image Engine został zintegrowany z bazami danych Sybase, Object Design oraz dodany jako komponent do aplikacji Oracle DBMS.

Obecnie firma Autonomy Virage oferuje oprogramowanie do zarządzania multimedialnymi bazami danych o nazwie **Virage MediaBin 7**. Jest to rozwiązanie wykorzystujące tzw. technologię *Meaning Based Computing* (MBC), która definiowana jest jako zdolność rozumienia informacji i rozpoznawania relacji pomiędzy strukturalnymi i niestrukturalnymi danymi. MBC wykorzystuje np. automatyczne odsyłacze (ang. hyperlinking), które łączą użytkownika z innymi informacjami, potencjalnie odpowiadającymi szukanym. Wymaga to jednak pełnego zrozumienia zawartości odpowiednich danych (np. tekst, głos, obraz, film), co w przypadku Virage MediaBin 7 ma odbywać się w czasie rzeczywistym.

Do firm, które wykorzystują technologię Meaning Based Computing zalicza się m. in. US Department of Homeland Security, Ford Motor Company, Zurich Financial Services, Boeing, czy Shell.

Więcej informacji na temat systemu Virage MediaBin 7 można znaleźć na stronie http://www.virage.com/.

3.3 Photobook

System Photobook został stworzony przez grupę naukowców MIT Media Laboratory w 1996 r. na uniwersytecie Cambridge USA (Pentland, Picard, Sclaroff 1996). Jest to zbiór interaktywnych narzędzi do wyszukiwania i przeglądania obrazów, który dzieli się na trzy różne podejścia do problemu reprezentacji zapytania w zależności od zawartości opisywanego zdjęcia (Veltkamp, Tanase 2000). Wyróżnia się w nim:

- obrazy twarzy,
- obrazy kształtów 2D,
- obrazy teksturowe.

Deskryptory używane w każdym z trzech powyższych przypadków mogą być między sobą łączone oraz uzupełnione opisem tekstowym.

Dwie pierwsze reprezentacje są do siebie podobne w kontekście wykorzystywania wektorów własnych macierzy kowariancji, jako ortogonalnych współrzędnych systemu. W tym przypadku metoda ekstrakcji cech wymaga normalizacji pozycji, skali i orientacji zdjęcia. Dla zbioru zdjęć uczących $\Gamma_1, \Gamma_2, \ldots, \Gamma_M$, gdzie Γ_i jest macierzą intensywności o rozmiarze $n \times n$, określa się ich wariację od średniej:

$$\psi = \frac{1}{M}\sum_{i=1}^{M} \Gamma_i \quad \rightarrow \quad \varphi_i = \Gamma_i - \psi$$

Powyższy zbiór wektorów jest wejściem dla transformaty Karhunena-Loeve (analizy głównych składowych), która służy do redukcji wymiarowości wektora cech. Do opisu nowego obrazu Γ wykorzystuje się M_1 wektorów własnych zbudowanych z największych wartości własnych (patrz rozdział 2.1.4). Nowy obraz ma postać:

$$\Omega = \left(\omega_1, \omega_2, \ldots, \omega_{M_1}\right)$$

gdzie $\omega_k = u_k^T(\Gamma - \psi)$, a u_k jest wektorem własnym dla wartości λ_k macierzy kowariancji

$$C = \frac{1}{M}\sum_{i=1}^{M} \varphi_i \varphi_i^T$$

W przypadku opisu tekstury używa się dekompozycji Wold'a, która wyróżnia trzy składniki obrazu: harmoniczny, krótkotrwały (zanikający, ang. evanescent) i interdeterministyczny.

Proces dopasowywania i indeksowania w Photobook jest inny niż w przypadku wcześniejszych systemów. Porównywanie obrazów odbywa się bowiem poprzez obliczenie odległości pomiędzy reprezentacjami Ω_1 i Ω_2 postaci:

$$\epsilon_{ij}^2 = \left\|\Omega_i - \Omega_j\right\|^2$$

Z kolei porównywanie dwóch kształtów to wyznaczenie energii, która jest potrzebna do zdeformowania jednego kształtu tak, aby odpowiadał drugiemu (ang. strain energy).

Technika rozpoznawania twarzy systemu Photobook została użyta przez firmę Viisage Technology w pakiecie FaceID, który był wykorzystywany przez kilka departamentów policji oraz lotnisk w USA. Obecnie firma ta zmieniła nazwę na **Visage Technologies** i dalej zajmuje się doskonaleniem technologii opisu i wykrywania twarzy: http://www.visagetechnologies.com/index.html.

Więcej informacji na temat systemu Photobook można znaleźć na oficjalnej stronie http://vismod.media.mit.edu/vismod/demos/photobook/.

3.4 VisualSEEk

VisualSEEk jest jednym z głównych systemów CBIR, który powstał na Columbia University in NY. Został on opisany w pracy Smith i Chang (1997) jako jedna z pierwszych propozycji hybrydowych łączących wyszukiwanie obrazów w

oparciu o wektory cech oraz lokalizacje przestrzenne. Integracja ta została osiągnięta poprzez podział zdjęcia na regiony o różnym kolorze i ich reprezentację za pomocą odpowiednich zbiorów koloru. Dodatkowymi zaletami systemu są:

- automatyczne określanie lokalizacji i cech charakterystycznych regionów,
- tworzenie zapytania przy wykorzystaniu informacji zarówno o położeniu, jak i o cechach obiektów,
- wykorzystanie technik szybkiego indeksowania i wyszukiwania.

Określanie charakterystycznych regionów koloru bazuje na technice tylnej projekcji (ang. back-projection), która składa się z czterech etapów: selekcji zbioru kolorów, realizacji tylnej projekcji obrazu, progowania oraz znakowania. Pierwszy z nich polega na utworzeniu 166-wymiarowego wektora binarnego c zdefiniowanego w przestrzeni HSV. Dla danego obrazu I jego tylna projekcja B jest następująca:

$$B(x,y) = \max_j (a[k,j] \cdot c[j])$$

gdzie $k \in \{0,1,\dots,165\}$ jest indeksem koloru punktu obrazu $I(x,y)$, a $a[k,j]$ określa podobieństwo pomiędzy dwoma kolorami k i j. Oznaczając dwa kolory przestrzeni HSV odpowiednio jako $m_i(h_i,s_i,v_i)$ oraz $m_j(h_j,s_j,v_j)$ podobieństwo między nimi wyznacza się z zależności (Smith, Chang 1997):

$$a[i,j] = 1 - \frac{1}{\sqrt{5}}\left[\left(v_i - v_j\right)^2 + \left(s_i \cos h_i - s_j \cos h_j\right)^2 + \left(s_i \sin h_i - s_j \sin h_j\right)^2\right]^{1/2}$$

Oprócz zbioru kolorów w systemie VisualSEEk wyznacza się również takie cechy, jak: środki ciężkości regionów, ich obszar rozumiany jako suma wszystkich wewnętrznych pikseli oraz szerokość i wysokość minimalnego ograniczenia tego obszaru.

Proces tworzenia zapytania polega na naszkicowaniu przez użytkownika kilku obszarów o określonym kolorze, pozycji i wielkości na dostępnej siatce graficznej. Po wyznaczeniu relacji przestrzennych między obiektami system zwraca obrazy najbardziej odpowiadające szkicowi.

Etap dopasowywania dzieli się z kolei na cztery części, w których wyznacza się podobieństwo w oparciu o zbiór koloru, bezwzględną lokalizację regionu, odległość przestrzenną oraz pomiędzy obszarami. Podobieństwo koloru obliczane jest z zależności:

$$d(c_q, c_t) = \left(c_q - c_t\right)^T A \left(c_q - c_t\right)$$

gdzie A jest macierzą podobieństwa koloru. Odległość pomiędzy obszarami to wartość bezwzględna ich różnicy, natomiast odległość pomiędzy minimalnymi ograniczeniami realizowana jest za pomocą metryki L_2. Jako wynik końcowy procesu dopasowania przyjmuje się wagową sumę wszystkich powyższych odległości.

Poniżej przedstawiamy adres oficjalnej strony poświęconej systemowi VisualSEEk:

http://www.ee.columbia.edu/ln/dvmm/researchProjects/MultimediaIndexing/VisualSEEk/VisualSEEk.htm.

Smith i Chang (1997) stworzyli również system WebSEEk, który służy do wyszukiwania informacji w sieci WWW. Składa się on z trzech głównych modułów: modułu obrazu i video, modułu klasyfikacji i indeksowania oraz modułu wyszukiwania. Umożliwia tworzenie zapytań w oparciu o tekst oraz zawartość wizualną (Rui, Huang, Chang 1999).

3.5 MARS

MARS czyli Multimedia Analysis and Retrieval System jest jednym z najbardziej znanych systemów wyszukiwania informacji. Powstał na University of Illinois at Urbana-Champaign w USA jako projekt autorstwa Ortega, Rui, Chakrabarti i in. (1997). Głównym celem systemu było stworzenie infrastruktury do zarządzania multimedialnymi danymi, która miała wspierać rozwój czterech obszarów:

- Multimedia Content Representation – wyznaczanie zawartości danych multimedialnych na podstawie cech wizualnych,
- Multimedia Information Retrieval – wyszukiwanie informacji poprzez wykorzystanie technik analizy zawartości danych, w tym adaptacji modeli sprzężenia zwrotnego,
- Multimedia Feature Indexing – indeksowanie danych przy użyciu metod zapobiegających problemom wysokiej wymiarowości wektorów cech,
- Multimedia Database Management – wykorzystanie efektywnych technik zarządzania danymi, które miały uskutecznić proces wyszukiwania.

Z technicznego punktu widzenia system umożliwia tworzenie zapytań w postaci kombinacji cech niskiego poziomu (kolor, tekstura, kształt) uzupełnionych opisem tekstowym. Kolor jest reprezentowany jako histogram o dwóch wymiarach odpowiadających współrzędnym H i S przestrzeni HSV. Tekstura to z kolei dwa histogramy, z których jeden określa stopień gruboziarnistości, a drugi kierunkowości obrazu oraz dodatkowo liczba skalarna mierząca kontrast. W celu wyznaczenia cech koloru i tekstury obraz dzieli się na obszary o rozmiarze 5×5, a następnie dla każdego z nich wyznacza się histogram i wektor współczynników transformaty falkowej. Każdy obiekt zdjęcia podlega dwuetapowej segmentacji. Na początku stosuje się metodę k-średnich w przestrzeni kolor-tekstura, a następnie grupuje się odpowiednie regiony za pomocą metod przyciągania. Przyciąganie dwóch regionów i, j jest zdefiniowane jako:

$$F_{ij} = M_i M_j / d_{ij}^2$$

gdzie M_i, M_j oznaczają rozmiary tych regionów, a d_{ij} jest odległością Euklidesa pomiędzy nimi. Kształt danego obiektu jest reprezentowany przez deskryptor Fouriera (FD).

W procesie tzw. matchingu wyróżnia się trzy miary podobieństwa:

- dla koloru wyznacza się przecięcie histogramów,
- dla dwóch obrazów tekstury oblicza się wagową odległość Euklidesa pomiędzy kontrastem oraz przecięcie histogramów dla dwóch pozostałych składników,
- w przypadku porównywania kształtu wykorzystuje się wagową sumę odchylenia standardowego tzn.:

$$ratio(k) = \frac{M_2(k)}{M_1(k)} \qquad shift(k) = \theta_2(k) - \theta_1(k) - \psi$$

gdzie $k = -N_c, \dots, N_c$ (N_c – ilość współczynników FD), a $M_i(k)$ oraz $\theta_i(k)$ odpowiadają wielkości i fazie współczynników deskryptora Fouriera. ψ jest to różnica orientacji głównych osi związanych z dwoma kształtami.

Ciekawostką MARS jest wykorzystanie zagadnienia interakcji systemu z użytkownikiem do modyfikowania zapytania. W drodze sprzężenia zwrotnego możliwa jest zmian wag przypisanych poszczególnym składnikom opisu obrazu. Problem ten został szeroko opisany w rozdziale 2.2.2.

Strona poświęcona systemowi MARS:

http://www-db.ics.uci.edu/pages/research/mars/index.shtml

3.6 Porównanie systemów CBIR

Opisane powyżej systemy zostały wybrane, jako najbardziej reprezentatywne z grupy 58 systemów CBIR opisanych w pracy Veltkamp i Tanase (2000). Istnieje jeszcze wiele innych, jak np. NETRA, Chabot, Blobworld, CANDID, Surfimage, z których każdy wniósł charakterystyczny wkład w rozwój technik wyszukiwania informacji na podstawie zawartości. Niestety nie istnieje rzetelne porównanie systemów pod kątem skuteczności wyszukiwania i jakości dopasowania obrazów, które mogłoby stwierdzić najwyższą niezawodność danego systemu. Wynika to z braku ujednolicenia warunków porównywania systemów, czy chociażby obliczania dwóch najbardziej reprezentatywnych miar ich efektywności tj. precyzji (ang. precision) i kompletności (ang. recall):

$$precyzja = \frac{liczba\ istotnych\ obrazów\ odpowiadających\ zapytaniu}{liczba\ wszystkich\ obrazów\ odpowiadających\ zapytaniu}$$

$$kompletność = \frac{liczba\ istotnych\ obrazów\ odpowiadających\ zapytaniu}{całkowita\ liczba\ istotnych\ obrazów}$$

W literaturze dostępne są jedynie porównania lokalnych deskryptorów obrazu, jak np. Mikołajczyk, Schmid 2005.

Poniżej przedstawiamy tabelę, która jest zestawieniem systemów CBIR pod kątem wykorzystanych cech niskiego (kolor, tekstura, kształt) i wysokiego poziomu. Cechy niskiego poziomu zostały zgrupowane w kategorie znaczeniowe. Na przykład, różne formy wyboru najbardziej charakterystycznego koloru obrazu zostały określone jako „kolor dominujący". Obraz własny odnosi się do cechy koloru, ponieważ jest wyznaczany na podstawie globalnych wartości koloru zdjęcia. Cechy atomowe tekstury to np. kontrast, gęstość, anizotropia. Elementarny deskryptor kształtu określa z kolei takie cechy, jak środek ciężkości, pole powierzchni, orientacja, długość głównych osi, ekscentryczność (mimośród). Dodatkowo uwzględniamy również możliwość detekcji twarzy i używania tzw. słów kluczowych (ang. keywords).

Tab. 3.1 Przegląd systemów CBIR i ich właściwości (źródło: Veltkamp, Tanase 2000).

		Blobworld	CANDID	Chabot	MARS	NETRA	Photobook	QBIC	Surfimage	VIR	VisualSEEk
Słowa kluczowe				X	X			X		X	
Detekcja twarzy							X				
Lokalizacja przestrzenna		X			X	X		X			X
Kształt	Brak danych									X	
	Inny										
	Deskryptor elementarny (obwód, pole pow.)	X						X			X
	Obiekt ograniczający										X
	Model elastyczny						X		X		
	Deskryptor Fouriera				X	X					
	Dopasowywanie wzorcowe							X			
	Kierunkowy histogram krawędzi								X		
Tekstura	Brak danych									X	X
	Inny								X *		
	Losowa dekompozycja polowa						X				
	Cechy tekstury (atomowe, Tamury)	X	X		X			X	X		
	Transformata: falkowa, Gabora, Fouriera				X	X			X		
Kolor	Brak danych									X	
	Inny										
	Kolor dominujący		X								X
	Lokalny histogram	X				X					
	Momenty koloru							X			
	Globalny histogram			X	X			X	X		
	Obraz własny						X		X		

* - histogram intensywności powierzchni krzywizny.

3.7 Bieżące badania w obszarze CBIR

W następnej części rozdziału postaramy się przedstawić obecną sytuację w świecie systemów Content-Based Image Retrieval. Skoncentrujemy się na opisie kampanii badawczych, które mają na celu rozwój technik wyszukiwania informacji w bazach danych oraz podamy organizacje biorące udział w powyższych badaniach.

1. ImageCLEF

Image CLEF czyli Cross Language Image Retrieval Track jest częścią organizacji, która ma na celu promocję systemów analizy obrazu ze szczególnym uwzględnieniem zagadnień wielojęzykowości danych i ich wyszukiwania. Organizacja ta co rocznie od 2000 r. prowadzi badania mające na celu usystematyzowanie technik, porównywanie skuteczności systemów oraz rozwój metod analizy i wyszukiwania obrazów. Działalność ta nie ogranicza się tylko do problemu wielojęzykowości, ale dotyka różnych innych dziedzin m. in.:

- wyszukiwania obrazów w medycznych bazach danych,
- systemów wizyjnych w robotyce,
- wyszukiwania zdjęć w sieci WWW (zwłaszcza wikipedii),
- interakcji użytkownika z systemem komputerowym.

Udział w badaniach biorą zarówno akademickie, jak i naukowe grupy, które specjalizują się w systemach CBIR oraz metodach Cross-Language Information Retrieval (CLIR), czyli wyszukiwania informacji niezależnie od używanego języka opisu.

Dostępny jest szereg publikacji opisujący co roczne osiągnięcia oraz nowe rozwiązania poszczególnych grup badawczych (Clough, Müller, Sanderson 2005; Müller, Deselaers, i in. 2006, 2008).

Więcej informacji na temat organizacji ImageCLEF można znaleźć na oficjalnej stronie internetowej: http://www.imageclef.org/.

2. ImagEVAL

W związku z brakiem przełożenia między wysokimi wynikami skuteczności systemów w badaniach przeprowadzanych przez CLEF, a sukcesem komercyjnym, w 2005 powstała nowa organizacja ImagEVAL (współfinansowana przez francuski program „Techno Vision"), która miała za zadanie rozwiązać powyższy problem. Wyróżniono pięć kluczowych zadań stawianych przed ImagEVAL (Moëllic, Fluhr 2006):

- rozpoznawanie zmodyfikowanych obrazów,
- wyszukiwanie połączonych danych obrazu i tekstu,
- detekcja obszarów tekstowych obrazu,
- detekcja obiektów,
- wydobywanie cech semantycznych danych.

W badaniach wzięło udział 11 międzynarodowych organizacji uczelnianych oraz badawczych, które stworzyły własne aplikacje wyszukiwania informacji. Lista tych organizacji dostępna jest na oficjalnej stronie ImagEVAL: http://www.imageval.org/.

Dostępność informacji na temat poszczególnych autorskich systemów jest ograniczona, dlatego też poniżej przedstawiamy opis jednego, ale równie ciekawego systemu stworzonego przez grupę **Viper** – Visual Information Processing for Enhanced Retrieval.

Viper jest grupą badawczą działu informatyki Uniwersytetu w Genewie w Szwajcarii (Computer Science Department of the University of Geneva). Zajmuje się problemami zarządzania i przetwarzania danych multimedialnych, w szczególności zagadnieniami związanymi z:

- wyszukiwaniem informacji multimedialnych (obraz, video, audio),

- indeksowaniem danych video i obrazów na podstawie zawartości (Content-Based video and image indexing),

- automatycznym opisem danych i ich przetwarzaniem.

Interesującą aplikacją grupy Viper jest system GIFT – The **G**NU **I**mage **F**inding **T**ool. Jest to system CBIR, który wykorzystuje techniki zapytania poprzez przykład oraz sprzężenia zwrotnego w procesie modyfikowania zapytania. W całości opiera się on na analizie zawartości obrazu bez konieczności uzupełniania zdjęć dodatkowymi adnotacjami. Wykorzystuje deskryptor kolor w postaci histogramu w przestrzeni HSV oraz filtr Gabora do klasyfikacji tekstury obrazu (Müller, Squire, i in. 1999). Jest to projekt tzw. opensource umożliwiający ciągły rozwój systemu poprzez dodawanie nowych funkcjonalności.

GIFT jest pierwszym systemem CBIR, który wykorzystuje język komunikacji swojego autorstwa MRML – Multimedia Retrieval Markup Language (http://www.mrml.net/).

Jest to protokół komunikacyjny oparty na języku XML, którego celem jest oddzielenie części klienta od części serwera systemu CBIR oraz stworzenie standardowej metody komunikacji pomiędzy różnymi serwerami CBIR.

Więcej informacji na temat systemu GIFT można znaleźć na stronie: http://www.gnu.org/software/gift/.

Systemy Content-Based Image Retrieval to dynamicznie rozwijająca się dziedzina analizy obrazu. Wraz z ogromnym postępem informatycznym oraz rozwojem nowoczesnych technik przetwarzania danych, metody stosowane w tych systemach ulegają ciągłym zmianom zwiększając tym samym potencjalne możliwości ich zastosowania. Wielu uczonych zgodnie twierdzi, że przyszłością systemów CBIR są tzw. aplikacje „real world" oraz wyszukiwarki danych w sieci WWW, których sukces w dużej mierze zależy od umiejętności połączenia metod wyszukiwania na podstawie tekstu i zawartości wizualnych razem z interakcją systemu z użytkownikiem (Datta, Joshi, Li, Wang 2008).

Rozdział 4. Główne problemy i kierunki rozwoju systemów CBIR

W dzisiejszych czasach systemy CBIR rozumiane są jako technologia wspomagająca organizację danych multimedialnych poprzez określanie ich zawartości wizualnej. Analiza tej definicji pozwala stwierdzić, że pod tym pojęciem kryje się bardzo wiele różnych technik począwszy od prostych funkcji porównujących podobieństwo obrazów, kończąc na złożonych odpornych metodach wyszukiwania i adnotacji zdjęć. Taka charakteryzacja systemów pozwala zatem umiejscowić tą dziedzinę wiedzy pośrodku zagadnień dotyczących m. in. baz danych, interakcji typu człowiek – komputer, maszyn uczących, wyszukiwania informacji, czy wizji komputerowej (Wang, Boujemaa i in. 2006). Naturalną konsekwencją takiego połączenia jest zatem możliwość rozprzestrzeniania się nowoczesnych metod stosowanych na potrzeby systemów CBIR, na inne dziedziny ogólnie rozumianej informatyki.

W literaturze okres ostatnich dziesięciu lat XX w. określany jest jako początkowy etap badań i rozwoju metod wyszukiwania informacji na podstawie zawartości. Postęp ten został dokładnie opisany przez wielu naukowców np. Smeulders, Worring i in. (2000), czy Rui, Huang, Chang (1999). Jako główny problem systemów tego okresu, który do końca nie został jeszcze rozwiązany, przyjmuje się występowanie (Datta, Joshi, Li, Wang 2008):

- tzw. „luki sensorycznej" czyli różnicy między definiowaniem obiektu w rzeczywistym świecie, a jego odzwierciedleniem w postaci komputerowego opisu dostarczonego za pomocą obrazu,
- tzw. „luki semantycznej" czyli różnicy zgodności pomiędzy informacją wyekstrahowaną z obrazu, a jej interpretacją, która może się zmieniać w zależności od sytuacji i celu poszukiwań.

W niniejszym rozdziale skoncentrujemy się na opisie podstawowych problemów występujących w systemach CBIR w swojej pierwotnej formie oraz postaramy się zweryfikować, które z nich pozostają dalej nie rozwiązane. Dodatkowo przeanalizujemy możliwe kierunki rozwoju technik wyszukiwania informacji ze szczególnym uwzględnieniem możliwości ich zastosowania w nowoczesnych aplikacjach.

Opisując techniki i obiecujące kierunki rozwoju metod wyszukiwania informacji Rui, Huang i Chang (1999) podali problemy, których rozwiązanie, ich zdaniem, było kluczowe do wprowadzenia systemów CBIR do praktycznego użytku. Poniżej przedstawimy opis każdego z nich dodatkowo uwzględniając obecną sytuację, nowe trudności i możliwości rozwiązania powstających problemów.

4.1 Interakcja systemu z użytkownikiem (ang. Human in the Loop)

Podstawową różnicą pomiędzy wizyjnym systemem komputerowym, a systemem wyszukiwania obrazów jest niezaprzeczalna konieczność obecności użytkownika w procesie wyszukiwania. Oznacza to, że system musi być wyposażony w narzędzie interakcji z użytkownikiem, aby mógł skutecznie interpretować jego zapytania. Przykłady tego typu technik można znaleźć w początkowych systemach np. QBIC, WebSEEk, Photobook itp. System MARS, jako pierwszy spośród nich, do rozwiązania powyższego problemu zaproponował technikę sprzężenia zwrotnego (ang. relevance feedback), która bazowała na modyfikacji wag poszczególnych składników zapytania. Pomimo szeroko zbadanej i rozwiniętej dziedziny sprzężenia zwrotnego nie znaleziono metody, która stosowana pojedynczo, dawałaby najlepsze rezultaty. Obecnie w literaturze pojawiają się coraz to nowe podejścia do problemu interakcyjności (Datta, Joshi, Li, Wang 2008).

4.2 Adaptacja cech niskiego poziomu do opisu złożonych obrazów

Z natury człowiek do opisu zdjęcia używa trudnych i niekiedy skomplikowanych pojęć. Niestety systemy komputerowe bazują na podstawowych cechach niskiego poziomu, które w niektórych przypadkach (detekcja twarzy, czy odcisków placów) wydają się jednak wystarczające. W ogólnym przypadku zmniejszenie tzw. „semantycznej luki” wymaga użycia procesów off-line i on-line. Procesy off-line wykorzystują m. in. techniki (Liu, Zhang, Lu, Ma 2007):

- nadzorowanego uczenia (ang. supervised learning) – zakładają obecność ludzkiego nadzoru w procesie tworzenia funkcji odwzorowujących wejście systemu na jego wyjście. Analiza nowych przypadków odbywa się przy wykorzystaniu nabytej wiedzy, dzięki czemu system jest w stanie przewidzieć wyjście na podstawie zbioru danych wejściowych. Do metod stosowanych w nadzorowanym uczeniu można zaliczyć np. klasyfikatory Bayesa, maszyny wektorów nośnych (ang. suport vector machine), sieci neuronowe.

- nienadzorowanego uczenia (ang. unsupervised learning) – w tym przypadku nie mamy informacji o wyjściu systemu. Celem tej techniki jest znalezienie opisu organizacji i podziału danych wejściowych – np. algorytm k-średnich.

Do przetwarzania on-line wymagane jest inteligentne narzędzie tworzenia zapytań, które umożliwi użytkownikowi ocenę wyników wyszukiwania informacji. Może być to realizowane również w postaci algorytmów sprzężenia zwrotnego.

4.3 Wsparcie dla problemu wyszukiwania danych w sieci WWW

Sieć World Wide Web może być uważana za największą z możliwych baz danych multimedialnych, która do swojego funkcjonowania oprócz zainteresowania ze strony użytkowników potrzebuje również odpowiednich narzędzi do zarządzania. Ten problem dotyka bezpośrednio zagadnienia wyszukiwania obrazów, które nie może być realizowane tylko za pomocą technik tekstowych. Już Rui, Huang i Chang w swojej pracy z 1999 r. zauważyli, że obecne metody indeksowania i wyszukiwania danych nie są w pełni skuteczne. Z kolei Datta, Joshi, Li, Wang (2008) pokazali, że systemy wyszukiwania informacji w sieci muszą wspomagać użytkownika, gdyż jest to podstawowy sposób ich przetrwania i dalszego rozwoju.

Obecnie występuje wiele algorytmów wyszukiwania obrazów w sieci, które bazują na metodach tekstowych. Do najbardziej popularnych można zaliczyć chociażby Google, czy Yahoo! Pomimo ich bezkonkurencyjności pojawiają się także systemy oparte na wyszukiwaniu na podstawie zawartości. Wśród nich można wyróżnić system Cortina (Quack, Mönich i in. 2004), który wykorzystuje zapytanie poprzez przykład i sprzężenie zwrotne do wizualizacji wyniku. Inne podejścia wykorzystują np. programy ładujące (ang. bootstrap) do adnotacji (Feng, Shi, Chua 2004), czy nieparametryczne oszacowania gęstości dla kolekcji dzieł sztuki (Smolka, Szczepański i in. 2004).

4.4 Indeksowanie wielowymiarowych cech obrazu

Gwałtowny wzrost zasobów baz danych pociąga za sobą konieczność przyspieszania procesu wyszukiwania informacji. Stąd też kluczowym problemem stawianym przed każdym systemem CBIR jest efektywna metoda indeksowania danych wielowymiarowych. Rui, Huang, Chang (1999) pokazali, że pierwsze systemy stosunkowo dobrze radzą sobie z rozmiarami baz rzędu kilkuset lub kilku tysięcy zdjęć. W obecnej sytuacji jest to jednak zupełnie niewystarczające. Jak podają Liu, Zhang, Lu, Ma (2007) tradycyjne algorytmy indeksowania danych takie, jak K-d-B tree (Robinson 1981), quad-trees (Vendrig, Worring, Smeulders 1999), czy R-tree (R*-tree) (Beckmann 1990) nie radzą sobie z problemem wysokiej wymiarowości tzn. ich efektywność sukcesywnie maleje wraz ze wzrostem wymiarowości przestrzeni cech. Dodatkowo wnioskują, że jeżeli wymiar wektora cech jest większy od 10, skuteczność

powyższych metod indeksowania nie przekracza tej, uzyskiwanej w przypadku zwykłego sekwencyjnego przeszukiwania.

W celu rozwiązania powyższego problemu zaczęto stosować inne metody wykorzystujące np. algorytmy X-tree (Berchtold, Keim, Kriegel 1996), VA-file (Weber, Schek, Blott 1998), czy i-Distance (Yu, Ooi, Tan, Jagadish 2001). Bazują one jednak tylko na czystej technice indeksowania, nie biorąc pod uwagę specyficznych cech obrazu. Oprócz powyższych rozwiązań zaczęto również konstruować algorytmy specjalnie na potrzeby baz danych. Przykładem tego typu rozwiązań jest system FIDS – Flexible Image Database System (Berman, Shapiro 1999), który wykorzystuje nierówność trójkąta jako podstawę do indeksowania danych i redukcji ilości porównań obrazów z nowym zapytaniem. Dodatkowo dostarcza użytkownikowi łatwych narzędzi do wyszukiwania obrazów w oparciu o kombinację wcześniej zdefiniowanych miar odległości. Pomimo licznego zainteresowania dziedziną indeksowania danych pozostaje jeszcze wiele pracy do wykonania w celu znalezienia efektywnych metod, które będą spełniały wymagania stawiane im przez obecne bazy danych.

4.5 Percepcja człowieka

Problem sposobu postrzegania obrazu przez człowieka jest związany z zagadnieniami opisanymi w rozdziałach 4.1 i 4.2. Jednak ze względu na swoją istotność zostanie on dokładnie przedstawiony poniżej.

Percepcja człowieka stała się bardzo popularnym tematem, który spotkał się z wysokim zainteresowaniem wśród naukowców. Już w latach 90-tych XX w. grupy MIT (Massachusetts Institute of Technology), NEC, czy UIUC (University of Illinois at Urbana-Champaign) prowadziły badania nad stworzeniem modelu percepcji człowieka na potrzeby systemów wyszukiwania informacji. W swojej początkowej fazie modele te bazowały na technikach sprzężenia zwrotnego, gdyż była to jedyna metoda, która mogła uwzględniać złożoność zagadnień interpretacji obrazu przez użytkownika i jego subiektywnej oceny.

Obecne badania nad sposobem postrzegania obrazów przez ludzi koncentrują się jednak na aspektach psychofizycznych. Już eksperymenty przeprowadzone w 1998 r. przez Papathomas, Conway i in. (1998) kładły główny nacisk na zagadnienia semantyczności informacji, pamięci poprzednich wejść oraz na relatywnej i absolutnej ocenie podobieństwa obrazów (Rui, Huang, Chang 1999). Z kolei Rogowitz i in. (1998) przeprowadzili serię eksperymentów analizujących psychofizyczną percepcję człowieka, z których wyciągnęli wnioski o istnieniu zbieżności między cechami wizualnymi obrazu, a jego znaczeniową interpretacją zawartości.

Niestety, jak podaje Lew, Sebe, Djeraba i Jain (2006) w dalszym ciągu „luka semantyczna" nie została wypełniona. Wynika to m. in. z następujących ograniczeń systemu komputerowego:

- niski stopień rozumienia bogatego słownictwa człowieka,

- niemożność dokładnej interpretacji zapytania użytkownika w kontekście zrozumienia celu poszukiwań,

- brak wiarygodnych i reprezentatywnych zbiorów danych do oszacowywania wyników wyszukiwania,

- brak właściwych miar podobieństwa, które odpowiadałyby satysfakcji użytkownika.

4.6 Bezpieczeństwo i obrazy

Temat wiążący systemy CBIR z bezpieczeństwem informacji nie był nigdy analizowany, jako istotny problem informatyczny. Jednak olbrzymi postęp technologiczny pociąga za sobą powstawanie nowych problemów związanych m. in. z dowodami interakcji człowieka z systemem komputerowym (ang. human interactive proofs). To zagadnienie stało się ostatnio elementem łączącym metody CBIR z bezpieczeństwem (Datta, Joshi, Li, Wang 2008).

Dążenie do skonstruowania inteligentnych systemów, które imitowałyby ludzkie możliwości jest z jednej strony niezwykle interesujące, a z drugiej zarazem bardzo niebezpieczne. Ryzyko to uwidacznia się w przypadku sieci WWW i publicznych serwerów, które stają się celem ataków złośliwych programów. Takie aplikacje mogą być tworzone na przykład w celu przechwytywania i gromadzenia olbrzymich zasobów sieciowych, czy chociażby do zmieniania wyników głosowania w sieci. Jednym ze sposobów radzenia sobie z tymi problemami jest stosowanie technik weryfikujących obecność użytkownika w Internecie zwanych CAPTCHA (ang. Completely Automated Public Turing test to tell Computers and Humans Apart). CAPTCHA jest to rodzaj zabezpieczenia stosowany na stronach WWW, który ma na celu dopuszczenie do przesyłania tylko tych danych, które zostały wypełnione przez człowieka. Innymi słowy technika ta odróżnia użytkownika od automatycznego programu bazując na odpowiedziach dostarczanych na stawiane pytania. Najbardziej rozpowszechnione metody CAPTCHA wykorzystują zniekształcony tekst (Yahoo!, MSN). Ostatnio zaczęto stosować ataki technologiczne na tekstowe systemy CAPTCHA wykorzystując oprogramowanie OCR (ang. Optical Character Recognition), które służy do rozpoznawania pojedynczych znaków i całych tekstów (Mori, Malik 2003). Sytuacja ta otworzyła drogę dla systemów CAPTCHA opartych na obrazach bazując na przekonaniu, że systemy CBIR są o wiele bardziej skomplikowane niż OCR. Pierwsza normalizacja takiego systemu zabezpieczającego została podana w pracy Chew i Tygar

(2004), gdzie wykorzystano losowo wybierane obrazy i pytania związane z ich zawartością np. co znajduje się na obrazie, które spośród obrazów nie pasują do pozostałych konceptualnie. W celu uniknięcia problemu, który dotknął tekstowe zabezpieczenia CAPTCHA, systemy CBIR są używane jako technika walidacyjna do zniekształcania obrazów przed przedstawieniem ich użytkownikowi (Datta, Joshi, Li, Wang 2008).

4.7 Wybór obiektywnych kryteriów oceny skuteczności systemów CBIR

Jednym z kluczowych problemów systemów CBIR jest określenie wspólnych kryteriów oceny ich skuteczności. Zagadnienie to odnosi się nie tylko do stworzenia uniwersalnych i miarodajnych metod porównujących efektywność systemów, ale również dotyczy problemu określenia wspólnych baz danych, które mogłyby służyć jako przykładowe zbiory testowe. Obecnie wiele grup badawczych bazuje na indywidualnie wybranych kolekcjach obrazów, które służą im do uzyskania jak najlepszych wyników skuteczności działania systemów. Taka sytuacja naturalnie prowadzi do błędnych wniosków porównawczych, co spowodowane jest brakiem wspólnego schematu oceny skuteczności.

Poniżej przedstawimy możliwości zobiektywizowania kryteriów oceny systemów CBIR, koncentrując się na analizie przeprowadzonej przez Müllera i in. w 2001 r.

4.7.1 Zdefiniowanie wspólnej bazy danych

Utworzenie uniwersalnego zbioru obrazów jest trudnym zagadnieniem, które musi spełnić kilka wymagań. Przede wszystkim wspólna baza danych musi być darmowa, a jej użytkowanie nie może podlegać żadnym ograniczeniom pochodzącym z natury ochrony praw autorskich. Dodatkowo kolekcja ta musi być zróżnicowana i zawierać grupy obrazów o różnej tematyce np. obrazy medyczne, zdjęcia twarzy, przedmiotów, widoków itp.

Obecnie wśród najbardziej popularnych baz danych obrazów stosowanych w systemach CBIR wyróżnia się zbiory firmy Corel, Kodak, bazę teksturową stworzoną przez Brodatza (1966), czy chociażby własne zbiory utworzone z obrazów znajdujących się w sieci WWW. Wśród wielu naukowców powstała opinia, że baza danych firmy Corel spełnia wszystkie wymagania potrzebne do oceny systemów wyszukiwania informacji. Zdania są jednak podzielone. Liu, Zhang, Lu, Ma (2007) twierdzą, że zaletą bazy Corel jest niewątpliwie jej zróżnicowanie i wielkość, natomiast zastosowany w niej podział na kategorie nie jest prawidłowy. Stąd też w pracy Shirahatti i Barnard (2005)

zaproponowano nowy zbiór danych składający się z wyników wyszukiwania obrazów za pomocą zapytania poprzez przykład i tekst. Baza ta składa się z 16.000 obrazów, które zostały wybrane z bazy firmy Corel i uchodzi za wysoce charakterystyczną. Więcej informacji na jej temat można znaleźć na stronie internetowej: http://kobus.ca/research/data/.

4.7.2 Oszacowywanie wyników wyszukiwania

Oszacowywanie wyników wyszukiwania obrazów przez użytkownika systemu jest wysoce subiektywne i zależne od celu poszukiwań. Stąd też w technikach Content-Based Image Retrieval można wyróżnić trzy zasadnicze podejścia do powyższego problemu (Müller, Squire, Marchand-Maillet i in. 2001):

- ✓ podział bazy danych na ściśle określone podzbiory – jest to powszechna metoda, w której użytkownik nie określa poprawności wyniku wyszukiwania, gdyż jest to realizowane bezpośrednio poprzez podział obrazów na grupy. Takie rozwiązanie może jednak sztucznie zwiększać efektywność systemu, ponieważ obrazy o podobnej zawartości znajdują się blisko siebie, co tym samym ułatwia proces wyszukiwania. Powoduje to zatem błędną ocenę efektywności całego systemu.
- ✓ symulacja oceny użytkownika - niektóre badania próbują symulować opinie użytkownika bazując na przekonaniu, że określa on podobieństwo obrazów na wzór metod stosowanych w systemach CBIR z dodatkowym uwzględnieniem zakłóceń w postaci szumu (Vendrig, Worring, Smeulders 1999).

- ✓ bezpośrednie oszacowywanie wyników – praktycznie metoda ta sprowadza się do określania przez użytkownika, czy dany wynik wyszukiwania jest odpowiedni, czy też nie. Jest to wysoce czasochłonne, ale bazuje na przekonaniu, że tylko człowiek, użytkownik systemu, jest w stanie wskazać jakiej odpowiedzi oczekuje na swoje zapytanie.

4.7.3 Metody oceny skuteczności systemów CBIR

Określanie skuteczności wyszukiwania systemów CBIR jest na tyle złożonym procesem, że w literaturze traktuje się ten problem jako oddzielny temat (Huijsmans, Sebe 2005). Jest to kluczowe zagadnienie nie tylko w kontekście wyboru obecnie najlepszego systemu wyszukiwania informacji, ale w kontekście utworzenia uniwersalnego zastawu metod, który służyłby do przewidywania i wyznaczania efektywności przyszłych aplikacji.

Müller, Squire, Marchand-Maillet i in. (2001) podają następujące metody oceny efektywności systemów Content-Based Image Retrieval:

a) ocena użytkownika

Jest to interaktywna metoda, w której użytkownik ocenia trafność dostarczonych mu danych wyjściowych bezpośrednio po zapytaniu. Wśród technik wykorzystywanych w tym przypadku wyróżnia się metodę nazywaną porównanie typu przed-po (ang. Before-after comparison). Sposób ten polega na wybraniu przez użytkownika bardziej trafnego wyniku wyszukiwania spośród podanych mu dwóch przykładów. Metoda ta wymaga co najmniej dwóch oddzielnych systemów: jednego bazowego i drugiego porównawczego.

b) metody jednowartościowe

- *wybór najlepszego dopasowania*

Berman i Shapiro (1999) przeprowadzili eksperymenty, w których starali się określić, czy „najbardziej odpowiedni" obraz wynikowy znajduje się wśród pierwszych 50, czy 500 wyszukanych obrazów. Liczby te wybrano odpowiednio jako maksymalna ilość obrazów mieszcząca się na ekranie monitora oraz jako maksymalna ilość zdjęć, którą użytkownik jest w stanie przeglądnąć.

- *średnia wyszukanych obrazów (ang. average rank of relevant images)*

Metoda ta, choć użyta przez Gargi i Kasturi (1999) uchodzi za mało miarodajną, ponieważ nawet pojedynczy obraz o wysokim stopniu odpowiedniości może błędnie zwiększyć efektywność całego systemu. Zamiast niej używa się techniki pierwszego odpowiedniego obrazu, która m. in. wykorzystywana jest w ramach testów przeprowadzanych przez organizację TREC – Text REtrieval Conference i również daje poprawne wyniki w przypadku systemów CBIR.

- *precyzja i kompletność (ang. precision and recall)*

Te dwa parametry są standardami w ocenie skuteczności systemów wyszukiwania informacji. Są to najbardziej popularne oszacowania, które stosunkowo dokładnie odzwierciedlają możliwości działania systemów. Jednak muszą być zawsze podawane razem. Ani precyzja, ani kompletność oddzielnie nie nadają się do mierzenia efektywności systemów CBIR (Müller, Squire, Marchand-Maillet i in. 2001). Istnieje kilka możliwości pomiaru tych parametrów. Wśród nich wyróżnia się np. pomiar precyzji przy kompletności równej 0.5, pomiar kompletności po wyszukaniu 1000 obrazów, precyzja po 20 pierwszych wyszukanych obrazach. W literaturze można

spotkać wiele prac naukowych wykorzystujących powyższe parametry: Iqbal, Aggarwal (2002), Bay, Ess i in. (2008), Müller, Michoux i in. (2004).

- *testowanie celu (ang. target testing)*

Podejście to znacząco różni się od powyższych metod oceny skuteczności procesu wyszukiwania obrazów. Polega ono na sprawdzeniu ile obrazów musi zostać przeszukanych przez użytkownika w celu znalezienia znanego mu uprzednio obrazu wynikowego, który jest najlepszym możliwym dopasowaniem. Metoda ta została użyta m. in. w systemie PicHunter (Cox i in. 1996)

- *współczynnik błędu*

Metoda ta znalazła zastosowanie m. in. w systemach detekcji twarzy (Hwang i in. 1999). W ogólnym przypadku współczynnik błędu jest zdefiniowany następująco (Müller, Squire, Marchand-Maillet i in. 2001):

$$error\ rate = \frac{liczba\ nieodpowiednich\ obrazów}{całkowita\ liczba\ wyszukanych\ obrazów}$$

- *prawidłowa i nieprawidłowa detekcja*

Metoda ta została wykorzystana w pracy Ozer, Wolf, Akansu (1999). Polega ona na wyznaczeniu liczby prawidłowych i nieprawidłowych detekcji, które podzielone przez całkowitą liczbę wyszukanych obrazów są równe odpowiednio parametrom: współczynnikowi błędu i precyzji.

c) reprezentacje graficzne

Inne podejście do problem oszacowania efektywności wyszukiwania informacji w systemach CBIR realizowane jest za pomocą reprezentacji graficznych. Wśród nich wyróżnia się:

- *wykres precyzja – kompletność (ang. precision vs recall graphs – PR graphs)*

Metoda ta zawiera dużo informacji na temat skuteczności systemów i dlatego stała się bardzo popularna wśród społeczeństwa CBIR (Squire i in. 1999). W niektórych przypadkach można spotkać modyfikacje powyższego grafu w postaci zmiany oznaczeń odpowiednich osi. Jednak ze względu na czytelność takiego wykresu powinno się unikać takiej zamiany. Czasami naukowcy prezentują wyniki swoich osiągnięć

używając częściowych grafów precyzji i kompletności (He 1997). Taka sytuacja wydaje się przydatna w momencie, gdy chcemy szczegółowo analizować dany obszar działania. Jednak w ogólnym przypadku może to prowadzić do błędnych interpretacji, ponieważ omija się obszary o niskiej skuteczności. Stąd też graf częściowy powinien być zawsze umieszczany razem z grafem całościowym (Müller, Squire, Marchand-Maillet i in. 2001).

- *wykresy precyzji i kompletności w zależności od ilości wyszukanych obrazów*

W odróżnieniu od powyższego przypadku oddzielenie od siebie parametrów precyzji i kompletności daje mniej informacji na temat skuteczności systemów CBIR. Analizując graf kompletności należy stwierdzić, że wygląda on bardziej optymistycznie niżeli graf PR zwłaszcza w przypadku, gdy kilka odpowiednich obrazów zostanie wyszukanych pod koniec procesu (Ratan i in. 1999). Graf precyzji z kolei jest podobny do PR, ale lepiej obrazuje jaka powinna być ogólna ilość obrazów używana w procesie dopasowywania. Zarazem jest jednak bardziej czuły na zmianę ogólnej liczby odpowiednich obrazów wynikowych dla jednego zapytania.

Istnieje jeszcze kilka innych modyfikacji powyższych grafów, które zostały przedstawione w literaturze. Wśród nich można wyróżnić:

- Vasconcelos i Lippman (1999) - graf prawidłowego wyszukania w zależności od liczby wszystkich wyszukanych obrazów (ang. correctly retrieved vs all retrieved graphs),
- Belongie, Carson i in. (1998) – ułamek prawidłowych w zależności od liczby wszystkich wyszukanych obrazów (ang. fraction correct vs no. images retrieved),
- Comaniciu, Meer i in. (1999) – graf współczynnika rozpoznawania w zależności od liczby wyszukanych obrazów (ang. average recognition rate vs no. images retrieved) – jest on podobny do grafu kompletności i pokazuje średni procent odpowiednich obrazów wśród N wyszukań.

- *wykres skuteczności wyszukiwania w zależności od zakłóceń (ang. retrieval accuracy vs noise graphs)*

Jest to metoda, która obrazuje zmianę skuteczności wyszukiwania w zależności od dodawania zakłóceń w postaci szumu obrazu. Taki model pomiaru efektywności nie nadaje się jednak do stosowania w przypadku aplikacji CBIR (Müller, Squire, Marchand-Maillet i in. 2001).

Jak pokazano powyżej ilość technik stosowana do oceny skuteczności systemów Content-Based Image Retrieval jest bardzo duża. Niestety ta różnorodność nie sprzyja procesom porównawczym i stanowi duży problem. Pomimo licznych prób normalizacji

metod i stworzenia uniwersalnych kryteriów oceny efektywności (Huijsmans, Sebe 2005), naukowcy używają własnych schematów, gdyż jest to najprostszy sposób do przedstawienie swojego systemu z jak najlepszej strony. Naprzeciw powyższym problemom wychodzą organizacje informatyczne takie, jak ImageCLEF, ImagEVAL, czy Pascal (ang. Pattern Analysis, Statistical Modelling and Computational Learning), których zadaniem jest unormowanie technik oceny systemów wyszukiwania informacji poprzez określenie wspólnych dla wszystkich baz danych, celów i kryteriów. Dzięki takim przedsięwzięciom możliwa jest wzajemna współpraca pomiędzy grupami badawczymi, co niewątpliwie pozytywnie wpływa na przyszły rozwój systemów przetwarzania i analizy danych multimedialnych.

Rozdział 5. Przykłady zastosowań wyszukiwania obrazów

Techniki wyszukiwania obrazów na podstawie zawartości są dziedziną wiedzy, która cieszy się szerokim zainteresowaniem od ponad 20 lat. Początkowo systemy te głównie wykorzystywano jako aplikacje do zarządzania bazami danych, jednak jak pokazano w rozdziale IV ich możliwości są o wiele większe. Analiza obrazu w kontekście jego zawartości wizualnej wyznacza nowe kierunki rozwoju i jest przykładem na to, że nowoczesne technologie są niesamowicie owocne i przydatne w życiu codziennym.

Celem niniejszego rozdziału jest przedstawienie przykładowych zastosowań aplikacji opartych na metodach Content-Based Image Retrieval. W pracy Eakins i Graham (1999) wyróżniono m. in. następujące dziedziny wykorzystujące zalety technik CBIR:

- zapobieganie przestępczości
- techniki wojskowe,
- medycyna diagnostyczna
- projektowanie: architektura, moda, wystrój wnętrz,
- dziedzictwo kulturowe – sztuka.

Bardziej wnikliwa analiza pozwala stwierdzić, że podczas, gdy naukowcy cały czas opracowują coraz to nowe techniki CBIR, można już teraz odnaleźć kilka w pełni funkcjonujących systemów, które cieszą się dużą popularnością. Poniżej postaramy się udowodnić to stwierdzenie.

5.1 Zapobieganie przestępczości

Obecne agencje wdrażania i egzekucji prawa dysponują ogromnymi zbiorami danych wizualnych zawierających np. obrazy twarzy przestępców i podejrzanych, odciski palców, czy zdjęcia śladów opon. Materiały te są następnie wykorzystywane do porównywania z nowymi dowodami podczas popełnienia kolejnego przestępstwa. Proces ten polega zatem albo na rozpoznawaniu tożsamości sprawcy, albo na znalezieniu najbliższego podobieństwa do danych znajdujących się w archiwum. Takie metody były stosowane już w latach 80-tych XX w. m. in. przez Federalne Biuro Śledcze w Waszyngtonie, w USA (ang. Federal Bureau of Investigation - FBI) oraz przez wiele innych placówek policyjnych na świecie (Eakins, Graham 1999).

Ciekawe zastosowania technik CBIR można znaleźć w systemach porównywania odcisków palców. Stały się one bardzo popularne i są stosowane już na skalę światową. Wśród komercyjnych systemów można tutaj wymienić np. aplikację AFIS (Automated Fingerprint Identification Systems) firmy East Shore Technologies w USA (http://www.east-shore.com/index.html), czy grupę aplikacji rodziny AFIX firmy AFIX Technologies (http://www.afix.net/).

Detekcja i rozpoznawanie obrazów twarzy jest kolejnym przykładem wykorzystania technik CBIR w życiu codziennym. Zagadnienie to jest bardzo szeroko rozwijane przez naukowców (Gao, Qi 2005; Yoshino, Taniguchi i in. 2005; Abate, Nappi i in. 2004), a jego związki z CBIR sięgają lat powstawania systemu Photobook, który był wyposażony w moduł detekcji twarzy (Pentland, Picard, Sclaroff 1996). Początkowo metody te opierały się na detekcji charakterystycznych kształtów twarzy np. długość i szerokość nosa, pozycja ust, kształt brody, które stanowiły wektor cech służący do porównywania. Jednak wraz z rozwojem metod informatycznych zaczęto stosować bardziej zaawansowane techniki jak np. transformaty falkowe, Gabora, sieci neuronowe, czy klasyfikatory Bayesa. Nie sposób wymienić wszystkich technik, jednak spośród nich można wyróżnić np.:

- metody wykorzystujące kierunkowe punkty narożne (ang. directional corner points) – są odporne na zmiany skali i oświetlenia obrazu (Gao, Qi 2005),
- grafowe metody dopasowania – wykorzystują graf charakterystycznych punktów twarzy obliczając odległości pomiędzy poszczególnymi węzłami oraz ich cechy (Yoshino, Taniguchi i in. 2005),
- metody oparte na sieciach neuronowych – są odporne na zniekształcenia obrazu takie, jak translacja, rotacja, deformacja, jednak ich skuteczność maleje wraz ze wzrostem liczby klas (indywidualności) (Lawrence, Giles i in. 1997).

Systemy detekcji twarzy cieszą się dużą popularnością, co wynika bezpośrednio z możliwości ich zastosowania np.:

- do ewidencjonowania ludności,

- do poszukiwania zaginionych osób,

- w kryminalistyce – do wyszukiwania osób posługujących się fałszywymi dowodami tożsamości,

- w kontrwywiadzie, wojnie z terroryzmem,

- w mediach i rozrywce.

Przykład polskiego, komercyjnego systemu FRS (ang. Face Retrieval Systems) można znaleźć w sieci WWW na stronie internetowej: http://www.555.pl/.

5.2 Techniki wojskowe

Jak podają Eakins i Graham (1999) zastosowania systemów CBIR na potrzeby wojska są najprawdopodobniej na najwyższym poziomie technologicznym, jednak dostęp do nich jest bardzo ograniczony ze względu na brak jakichkolwiek publikacji naukowych. Można się domyślać, że wojsko wykorzystuje te techniki np.:

- w systemach rozpoznawania wrogich samolotów na obrazach z radarów,

- w procesie identyfikowania celu za pomocą zdjęć satelitarnych,

- w systemach naprowadzania rakiet samosterujących, itp.

Dodatkowo wiele spośród technik wykorzystywanych przy zapobieganiu przestępczości również może mieć zastosowanie w przypadku wojska.

5.3 Medycyna diagnostyczna

Związek pomiędzy zagadnieniami medycyny diagnostycznej, a systemami Content-Based Image Retrieval dotyczy problemu efektywnego i wydajnego zarządzania i przetwarzania zdjęć medycznych. Stąd też wykorzystanie systemów wyszukiwania obrazów na podstawie zawartości na potrzeby współczesnej medycyny stało się oddzielnym zagadnieniem analizy obrazu, które cieszy się dużym zainteresowaniem (Müller, Michoux i in. 2004; Oliveira, Cirne, Marques 2007; Lehmann, Güld, Deselaers i in. 2005).

Współczesne szpitale, w wyniku przeprowadzania licznych badań radiologicznych, kardiologicznych, czy chociażby neurologicznych, produkują kilka tysięcy obrazów i danych cyfrowych dziennie. Naturalną konsekwencją tego zjawiska są nie tylko problemy zarządzania tak ogromnymi bazami danych, ale również stworzenie efektywnych metod, które mogłyby porównywać i analizować nowe przypadki chorób pacjentów bazując na wiedzy uprzednio zdobytej.

Müller, Michoux i in. (2004) wymieniają kilka powodów, dla których systemy CBIR powinny stać się integralną częścią narzędzi komputerowej analizy medycznej. Jednym z nich jest konieczność dostarczania potrzebnej informacji we właściwym miejscu i czasie, w celu poprawy jakości i efektywności opieki medycznej. Taki cel wymaga nowoczesnych technik podejmowania decyzji bazujących np. na znajdywaniu obrazów podobnych chorób, czy tych samym części anatomicznych. O ile przykłady tego typu rozwiązań są dostępne (standard DICOM – Digital Imaging and Communications in Medicine), o tyle ich skuteczność pozostawia na razie wiele do życzenia. Innym wysoce przydatnym powodem stosowania systemów CBIR jest możliwość porównywania obrazów medycznych na podstawie ich zawartości wizualnej. Chodzi tu nie tyle o

znajdywanie podobnych przypadków chorób, ile o możliwość znajdywania przypadków pacjentów charakteryzujących się podobnymi objawami, ale mających różną diagnozę.

Próbą rozwiązania problemu zarządzania bazą danych medycznych jest środowisko PACS (ang. Picture Archiving and Communications Systems) (Lemke 2003). Jest to systemem będący kombinacją oprogramowania i sprzętu, którego głównym zadaniem jest:

- edycja i obrazowanie cyfrowych zdjęć medycznych,

- transmisja informacji o pacjentach za pomocą zabezpieczonej sieci,

- interpretacja i archiwizacja danych i raportów medycznych w standardzie DICOM.

Jest to swego rodzaju próba wprowadzenia standardu archiwizowania i opisu obrazów cyfrowych, która ma umożliwić w przyszłości szybki i bezpieczny dostęp do danych medycznych pacjentów, na potrzeby nowych przypadków choroby.

Kolejnym aspektem, który należy poruszyć przy opisie zastosowania systemów CBIR na potrzeby współczesnej medycyny, jest możliwość ich wykorzystania we wielu oddziałach szpitali. Wśród nich można wyróżnić departamenty dermatologiczne, cytologiczne, patologiczne, kardiologiczne i radiologiczne. Znane są przypadki stosowania systemów CBIR do klasyfikacji obrazów wysokiej rozdzielczości tomografii komputerowej, rezonansu magnetycznego mózgu, czy zdjęć rentgenowskich kręgosłupa (Müller, Michoux i in. 2004; Antani, Long, Thoma 2002).

Dalszy rozwój systemów CBIR w medycynie jest niewątpliwie związany z takimi pojęciami, jak nauka, badania, diagnostyka, automatyczna adnotacja i kodyfikacja (Müller, Michoux i in. 2004). W przypadku nauki studenci mogą przeszukiwać ogromne zbiory obrazów medycznych biorąc pod uwagę różne kryteria wyszukiwania np. wspólna diagnoza, czy podobieństwo wizualne. Badania również zyskają na adaptacji technik CBIR, gdyż dzięki nim będzie można rozpatrywać nowe przypadki chorób z różnych punktów widzenia, czy poprzez korelację pomiędzy cechami wizualnymi, które w konsekwencji mogą prowadzić do interesujących odkryć. W końcu sama diagnostyka także może ulec zmianie dzięki wykorzystaniu nowych narzędzi do wspierania procesu podejmowania decyzji (ang. supporting the clinical decision-making process).

5.4 Projektowanie: architektura, moda, wystrój wnętrz

Współczesne projektowanie architektoniczne polega na reprezentacji i wizualizacji obiektów za pomocą stylizowanych modeli dwu- i trójwymiarowych. Wizualizacja ta nie tylko ma zachęcić potencjalnego klienta do inwestycji, ale równocześnie musi podlegać pewnym zewnętrznym ograniczeniom, nie tylko

finansowym. Mówiąc ograniczenia mamy na myśli fakt, że architekt musi być świadom istnienia poprzednich projektów w celu uniknięcia niechcianych kopii oryginału. Stąd też jednym ze sposobów wykorzystania technik CBIR w architekturze jest możliwość wyszukiwania poprzednich projektów pod kątem określonych kryteriów (Eakins, Graham 1999). Już w latach 90-tych XX w. pojawiły się pierwsze systemy wspomagające takie wyszukiwanie, jak np. SAFARI (Eakins 1993), czy AUGURS (Yang i in. 1994). Jednak ich skuteczność, ze względu na niski poziom rozwoju technik CBIR, była niewystarczająca. Pomysł adaptacji metod wyszukiwania obrazów dla potrzeb projektowania wydaje się trafny i interesujący, jednak jak podają Eakins i Graham (1999) musi on iść w parze z rozwojem narzędzi wspomagających to projektowanie takich, jak np. CAD (ang. Computer Aided Design).

Podobieństwa do powyższego zagadnienia można upatrywać się również w procesach projektowania mody i wystroju wnętrz. Tutaj także autor ma narzucone pewne zewnętrzne ograniczenia, jak np. wybór materiałów. Zdolność do wyszukiwania kolekcji tekstylii w celu znalezienia właściwej kombinacji koloru i tekstury może być rozważana, jako problem natury Content-Based Image Retrieval (Eakins, Graham 1999).

5.5 Dziedzictwo kulturowe – sztuka

Dziedzictwo kulturowe przejawia się w ogromnej ilości rzeźb, obrazów i innych dzieł sztuki, które można oglądać m. in. w muzeach, galeriach, czy książkach. Zdolność do identyfikacji obiektów, które charakteryzują się podobnymi cechami wizualnymi może być przydatna zarówno dla naukowców badających historyczne aspekty rozwoju kultury, jak również dla miłośników i pasjonatów sztuki. Daje to szanse na rozwój systemów CBIR w kierunku wyszukiwania podobieństwa pomiędzy różnymi dziełami sztuki. Pierwsze próby stworzenia takiego systemu sięgają lat 90-tych XX w. Wśród nich można wyróżnić pracę Hirata i Kato (1992), czy chociażby system QBIC (Holt, Hartwick 1994), który był wykorzystywany m. in. jako narzędzie do zarządzania bibliotekami dzieł. Ciekawa praca w tej dziedzinie została wykonana przez Yu, Ma, Tresp i in. w 2003 r. Przedstawili oni rozwiązanie problemu wyszukiwania obrazów dzieł sztuki za pomocą profilów preferencji użytkownika. Początkowo wykorzystali maszyny wektorów nośnych (ang. Support Vector Machine – SVM) do zamodelowania każdego, indywidualnego profilu użytkownika, a następnie zmodyfikowali je bazując na metodach klasyfikacji Bayesa.

5.6 Ochrona praw autorskich

Interesującym zagadnieniem łączącym systemy CBIR z ochroną praw autorskich jest problem dotyczący detekcji fałszerstw. Obecnie sytuacje takie, jak kopiowanie zdjęć w sieci pod innym nazwiskiem lub duplikowanie znaków firmowych i log przez inne organizacje bez większych modyfikacji i z oczywistą intencją do wprowadzania konsumenta w błąd, są niestety na porządku dziennym. Stąd też powstał pomysł wykorzystania systemów CBIR do identyfikacji i weryfikacji praw autorskich. W przypadku dokładnych kopii ich detekcja wydaje się trywialna i automatyczna, jednak modyfikacja znaków firmowych wymaga już wykorzystania miar podobieństwa stosowanych w systemach wyszukiwania informacji. Mówiąc modyfikacja mamy na myśli zmiany w odcieniach kolorów, zniekształcanie obiektów obrazu, zmiany kontrastu, jasności itp. Problemy te wymagają wykorzystania odpornych metod pomiaru podobieństwa. Ich przykłady można znaleźć np. w pracach Ke i in. (2004), czy Zhang i Chang (2004).

Przedstawione powyżej przykłady zastosowań systemów CBIR w życiu codziennym są tylko częścią możliwości, jaka drzemie w tych metodach i ich aplikacjach. Oprócz nich można również wymienić chociażby takie dziedziny, jak reklama i dziennikarstwo, edukacja i nauka, czy domowa rozrywka. W następnym rozdziale postaramy się podać praktyczne przykłady zastosowania metod CBIR w telefonach komórkowych i urządzeniach PDA.

Rozdział 6. Praktyczne zastosowania metod CBIR w urządzeniach mobilnych

Jednym z obecnych trendów jest powszechna dostępność urządzeń mobilnych wyposażonych w kamery cyfrowe. Wykorzystując telefony komórkowe, czy urządzenia PDA (ang. Personal Digital Assistant) użytkownicy robią niezliczoną liczbę zdjęć, która służy im np. do uzyskania opinii innej osoby na temat danego produktu, czy chociażby do zapamiętania interesującego wydarzenia. Analizując tą sytuację z naukowego punktu widzenia zaczęto zastanawiać się, w jaki sposób można wykorzystać popularność tych urządzeń do rozwoju obecnych technik informatycznych. Interesującym zagadnieniem wydaje się możliwość wykorzystania metod i systemów Content-Based Image Retrieval do dostarczania pewnych informacji użytkownikowi. Naukowcy postawili sobie za cel skonstruowanie systemu, który będzie w stanie udzielić odpowiedzi na zapytanie użytkownika – turysty, który potrzebuje informacji na temat danego miejsca, punktu orientacyjnego, czy budynku. Scenariusz tego przypadku jest następujący: turysta za pomocą swojego telefonu komórkowego robi zdjęcie interesującego go miejsca, następnie wysyła je do serwera, który w odpowiedzi dostarcza mu informacji np. o jego położeniu, przeszłości historycznej, czy o różnych ciekawostkach z nim związanych.

W rozdziale VI postaramy się przeanalizować pracę, jaka została wykonana w dziedzinie CBIR i urządzeń mobilnych oraz opiszemy dwa systemy, które są obecnie w fazie rozwoju i ulepszeń. Są to:

- Snap2Tell – system skonstruowany przez organizację IPAL (ang. Image & Pervasive Access Lab),
- Landmark-Based Pedestrian Navigation System – system opracowywany przez Nokia Research Center w Palo Alto, we współpracy z kilkoma uniwersytetami w USA.

W 1997 r. Feiner, MacIntyre i in. skonstruowali tzw. maszynę zwiedzającą (ang. touring machine), która miała dostarczać użytkownikowi użytecznych informacji na temat konkretnych miejsc za pomocą urządzenia mobilnego. Ciekawym przykładem podobnego systemu multimedialnego jest InfoScope autorstwa Haritaoglu (2001). Wykorzystał on urządzenie PDA wyposażone w lokalizator GPS do określania położenia i dostarczania informacji. Z kolei Mai, Dodds i Tweed (2003) użyli urządzenia PDA podłączonego do laptopa i do Internetu za pomocą sieci WLAN. Podejście to zakładało, że dany punkt dostępu (ang. access point) mieści się w obszarze działania całego systemu. Zdjęcie zrobione kamerą było wysłane do serwera, który przy pomocy programu 3DMax generował obraz referencyjny. Proces dopasowania był realizowany poprzez detekcję cech liniowych. Jednak jak podają Chevallet, Lim i

Leong (2007) takie rozwiązanie oparte na konstrukcji modelu 3D jest kosztowne obliczeniowo i nie nadaje się do zastosowania dla wszystkich rodzajów scen.

Hare i Lewis w 2005 r. przedstawili system oparty na urządzeniu PDA, które poprzez bezprzewodowe połączenie umożliwiało tworzenie wizualnego zapytania do bazy danych National Gallery Image Collection. Ich metoda opierała się na algorytmie SIFT (Lowe 2004), który jest odporny na przekształcenia afiniczne obrazu. Ahmad, Abdullah, Kiranyza i Gabbouj (2005) stworzyli aplikację MUVIS, w której do procesu wyszukiwania zdjęć użyli m. in. histogramu koloru i filtru Gabora. Ich system nie wykonywał żadnego przetwarzania obrazu bezpośrednio na telefonie, a ocena skuteczności odbywała się wyłącznie na podstawie ilości odpowiednich wyszukań.

Jak widać z powyższych przykładów metody CBIR znalazły szerokie zainteresowanie w świecie urządzeń mobilnych. W ogólnym przypadku są one wykorzystywane do dostarczania użytkownikowi informacji zwrotnej na dany temat, który nie koniecznie musi dotyczyć punktów charakterystycznych, czy miejsc. Innym możliwym zastosowaniem jest chociażby dostarczanie informacji o zawartości odżywczej danego posiłku, co może być przydatne dla konsumentów dbających o własne zdrowie (Chevallet, Lim, Leong 2007).

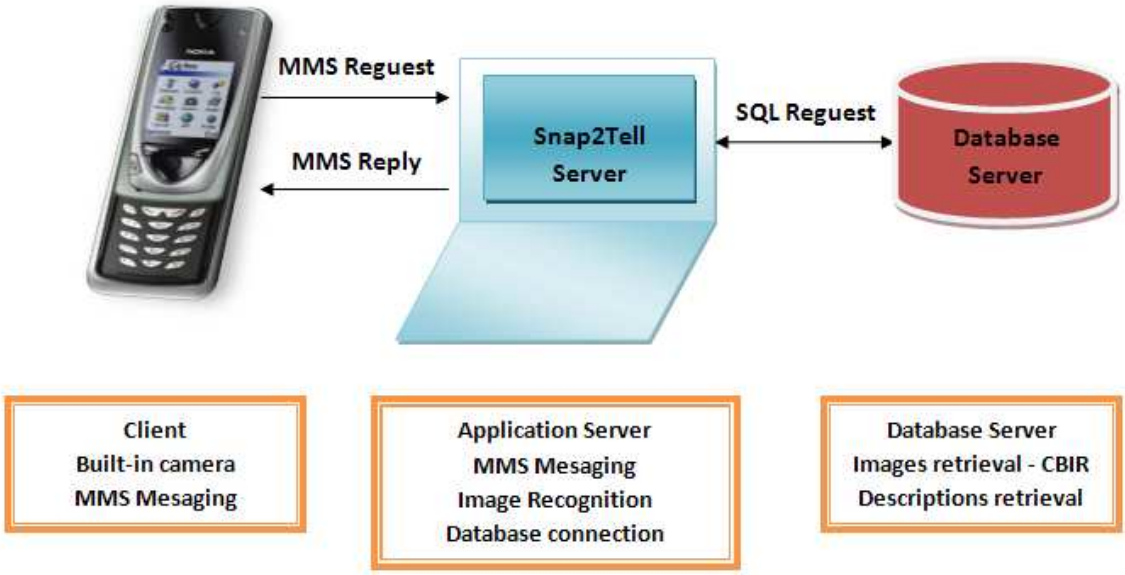
6.1 Snap2Tell

System Snap2Tell (Chevallet, Lim, Leong 2007) został skonstruowany przez organizację IPAL we współpracy z Institue for Infocomm Research (I^2R) w Singapurze. Jest to interesująca aplikacja, która dostarcza informacji na temat punktów orientacyjnych Singapuru poprzez wykorzystanie specjalnie skonstruowanej do tego celu bazy danych STOIC – The Singapore Tourist Object Identification Collection.

Baza STOIC składa się z około trzech tysięcy zdjęć obiektów, które cieszą się największym zainteresowaniem wśród turystów. Obrazy te zostały zrobione przy użyciu kilku kamer (w tym urządzeń PDA) o różnej rozdzielczości i w zmieniających się warunkach otoczenia tzn. przy zmianach oświetlenia, kąta widzenia, czy odległości. Dzięki temu uzyskano wiarygodne źródło danych na podobieństwo zdjęć wykonywanych przez zwykłych turystów przy użyciu telefonów komórkowych. Każdy obraz jest uzupełniony dodatkowymi informacjami, które opisują lokalizację miejsca, datę, autora i rodzaj sprzętu, który posłużył do wykonania zdjęcia.

System Snap2Tell jest zbudowany w oparciu o typową strukturę trójwarstwową postaci klient – serwer. Klientem w tym przypadku jest telefon komórkowy z wbudowanym aparatem fotograficznym, który obsługuje standardy MMS i GPRS. Chevallet, Lim,

Leong (2007) wykorzystali do tego celu model Nokia 7650. Poniżej przedstawiamy strukturę systemu.



Rys. 6.1 Architektura systemu Snap2Tell typu klient – serwer (źródło: Chevallet, Lim, Leong 2007).

Działanie systemu jest następujące. Użytkownik robi zdjęcie za pomocą telefonu komórkowego i przy użyciu aplikacji Snap2Tell wysyła zapytanie do serwera. Serwer uzyskuje informację o lokalizacji obiektu na obrazie dzięki pomocy operatora sieci. Wówczas możliwe jest wysłanie zapytania w języku SQL do bazy danych STOIC, w celu wyszukania tzw. meta-danych związanych z określoną lokalizacją. Kolejnym etapem jest proces porównania obrazu znajdującego się w serwerze z danymi uzyskanymi z bazy STOIC. Jeżeli w procesie dopasowania uzyska się wynik powyżej pewnego założonego progu, wówczas następuje wysłanie pozytywnej informacji zwrotnej do użytkownika, która może zawierać zarówno tekst, jak i opis w postaci pliku audio.

W celu przetestowania skuteczności działania aplikacji Snap2Tell Chevallet, Lim i Leong przeprowadzili szereg eksperymentów, w których brali pod uwagę zmiany kilku czynników:

- wpływ ilości słupków histogramu obrazu na wynik wyszukiwania podobieństwa

W tym przypadku chodzi o znalezienie równowagi pomiędzy ilością słupków histogramu obrazu wysyłanego przez telefon komórkowy do serwera, a precyzją

rozpoznawania określonej sceny. Analiza wyników pozwoliła stwierdzić, że wraz ze wzrostem ilości słupków jakość wyszukiwania również wzrasta aż do pewnego momentu. Dla określonego punktu dalsze zwiększanie ilości słupków powoduje błędy w odzwierciedleniu, gdyż nieznaczna zmiana w rozkładzie koloru obrazu powoduje przesuwanie się pikseli pomiędzy przyległymi słupkami. Stąd też autorzy wysunęli wnioski, że najlepszą precyzję uzyskuje się dla liczby 11 słupków histogramu zdjęcia.

- wpływ układu, kompozycji zdjęcia (ang. influence of image composition)

Użytkownik systemu wysyłając zapytanie do serwera oczekuje, że informacja zwrotna będzie w najlepszym stopniu odpowiadała jego oczekiwaniom tzn. że obiekt zostanie prawidłowo rozpoznany. W tym celu wysunięto hipotezę, że środek wysłanego obrazu jest bardziej istotny niżeli jego krawędzie, gdyż to on powinien zawierać najwięcej użytecznych informacji. Stąd też autorzy postanowili podzielić obraz na bloki i dobierać ich wagi w zależności od położenia. Jednak takie podejście nie przyniosło znacznego zwiększenia skuteczności.

Chevallet, Lim i Leong posunęli się więc krok dalej implementując w swoim algorytmie narzędzia do kadrowania obrazu. Miało to pozwolić użytkownikowi wybranie najważniejszego obszaru zdjęcia, który posłużyłby do lepszego odzwierciedlenia szukanego obiektu. Dzięki tej technice skutecznie odizolowano szukany budynek od otoczenia i tła. Jak można się było spodziewać metoda ta dała lepsze rezultaty, ponieważ kadrowanie jednoznacznie przedstawia to, czego tak naprawdę poszukuje użytkownik.

- wpływ wielkości bazy danych STOIC

Wybór nowego zestawu kamer umożliwił dokładniejsze odzwierciedlenie zróżnicowania jakości rozumianej w sensie wielkości pikseli, ostrości i konsystencji koloru. Nowy zbiór danych zawierał 5 zdjęć każdej sceny, co umożliwiło dokładniejsze odzwierciedlenie miejsca, budynku, czy ogólnie punktu charakterystycznego.

- wpływ jakości urządzeń mobilnych

Do eksperymentów użyto 8 kamer o różnej wielkości matrycy CCD. Niestety nie podano dokładnych nazw urządzeń. Poniżej przedstawiamy zbiorczą tabelę, która pozwoli wyciągnąć konstruktywne wnioski.

Tab. 6.1 Wpływ jakości urządzeń mobilnych na końcowy wynik precyzji systemu Snap2Tell (źródło: Chevallet, Lim, Leong 2007).

Model	Rozmiar matrycy CCD (Mp)	Rok produkcji	Ścisła precyzja (ang. strict precision) [%]
Telefon komórkowy	0.01	2004	46.1
Pocket PC	0.3	2004	16.4
Kamera 1	2.1	2004	72.4
Kamera 2	3.2	2003	15.3
Kamera 3	3.3	2000	22.2
Kamera 4	3.3	2001	16.9
Kamera 5	5.2	2002	60.0
Kamera 6	6.3	2003	46.7

Analizując powyższe wyniki można stwierdzić, że rozmiar matrycy CCD wyrażony w mega pikselach oraz rok produkcji urządzenia nie mają większego znaczenia. Stosunkowo dobrze poradził sobie telefon komórkowy w porównaniu do przenośnego komputera PC. Jak podają autorzy testów najlepszy wynik należy do taniej kamery, co sugeruje, że w przyszłości trzeba będzie kalibrować urządzenia, bądź bazować na ekstrakcji cech niskiego poziomu, które są mniej wrażliwe na cechy charakterystyczne kamery.

Podsumowując swoją pracę Chevallet, Lim, Leong (2007) stwierdzili, że ich prototypowy system Snap2Tell, pomimo swojej funkcjonalności i stosunkowo dobrych wyników, nie nadaje się do użytku jako produkt komercyjny. Wykorzystanie tylko dystrybucji kolorów obrazu, jako głównej cechy porównawczej jest niewystarczające w większości rzeczywistych sytuacji. Stąd też prace nad ulepszeniem powyższego systemu będą zakładały dodanie nowych metod detekcji cech charakterystycznych. Więcej informacji na temat systemu Snap2Tell można znaleźć na oficjalnej stronie internetowej:
http://ipal.i2r.a-star.edu.sg/snap2tell.htm.

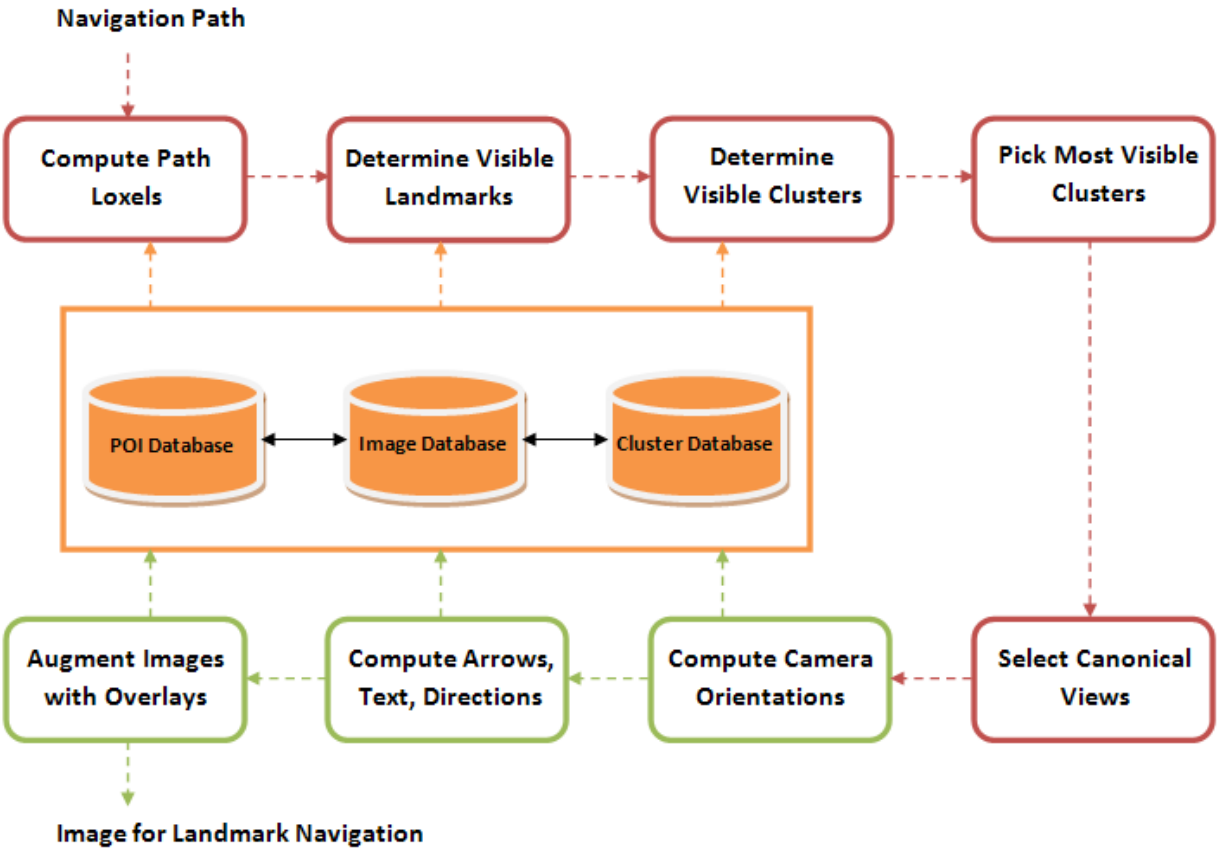
6.2 Landmark-Based Pedestrian Navigation System

Nowy system o nazwie Landmark-Based Pedestrian Navigation jest kolejnym przykładem umiejętnego wykorzystania metod Content-Based Image Retrieval w telefonii komórkowej. Jest to dzieło grupy Nokia Research Center, która razem z trzema uniwersytetami w USA: Stanford University, University of Washington i George Mason University opracowuje system, którego głównym celem jest automatyczne generowanie wskazówek nawigacyjnych dla pieszych, które są bezpośrednio wyświetlane na ekranie urządzenia mobilnego. Do tego celu wykorzystuje istniejącą bazę danych tzw. geotagged photographs tzn. zdjęć zrobionych za pomocą zwykłych telefonów komórkowych, które oprócz zawartości wizualnej są uzupełnione dodatkowymi informacjami np. o położeniu punktów (Hile, Vedantham, Cuellar i in. 2008).

Metoda nawigacji oparta na wyświetlaniu zdjęć punktów charakterystycznych danej trasy (budynków, mostów, parków itp.) jest uważana za efektywną w momencie, gdy dane obiekty są wyraźnie widoczne na ekranie telefonu komórkowego, a drogowe wskazówki są jednoznaczne. Dodatkowo powyższy system nawigacji wyposażony jest w standardową mapę, która służy do sprawdzania poprawności działania aplikacji.

Jak wspomniano powyżej każde zdjęcie w bazie danych jest uzupełnione informacjami o lokalizacji GPS oraz nazwami obiektów, które znajdują się w pobliżu miejsca przedstawionego na obrazie. Naturalną własnością takiej bazy jest to, że interesujący obiekt posiada większą liczbę skojarzonych z nim zdjęć i dzięki temu może być klasyfikowany ze względu na swoją popularność. W celu lepszej organizacji danych baza została podzielona na grupy (ang. clusters) obliczane za pomocą grafu, który określa relacje pomiędzy parami zdjęć. Grupy składają się z tzw. kanonicznych widoków obiektów (and. canonical views). Dodatkowo całkowity obraz mapy świata na telefonie przedstawiony jest w postaci siatki komórek (ang. loxel) o rozmiarze 30×30 metrów.

Poniżej przedstawiamy, w jaki sposób realizowany jest proces selekcji właściwych zdjęć do nawigacji, które są uzupełnione informacjami o pożądanych kierunkach. Dokładnie obrazuje to diagram blokowy.



Rys. 6.2 Schemat blokowy selekcji zdjęć do nawigacji. Pierwsze 5 etapów służy do wyboru obrazu, a następne 3 do jego uzupełnienia o jednoznaczne kierunki (źródło: Hile, Vedantham, Cuellar i in. 2008).

Wybór zdjęć do nawigacji

Etap ten składa się z następujących kroków:

- wyznaczenie ścieżki, którą będzie szedł pieszy – wybór odpowiednich komórek (ang. loxels),

- wyznaczenie widocznych punktów charakterystycznych dla danych komórek. Jest to realizowane poprzez określenie, które obiekty komórki znajdują się wewnątrz jądra o rozmiarze 3×3,

- wybór właściwej grupy zdjęć bazy danych (ang. cluster), która znajduje się wewnątrz jądra,

- obliczenie dwóch ciągów danej grupy zdjęć. Pierwszy porządkuje grupy według ich bliskości do ścieżki nawigacyjnej, a drugi szereguje kierunki,

- dyskretyzowanie każdego ciągu na cztery możliwe wartości: wysoka (ang. high), średnia (ang. medium), niska (ang. low), brak (ang. none). Utworzenie siatki o rozmiarze 4×4, w której każda komórka odpowiada jednoznacznej kombinacji par wartości dla dwóch ciągów,

- przeszukanie siatki i wybór komórki, dla której kanoniczny widok najlepiej obrazuje obraną ścieżkę pieszego.

Uzupełnienie zdjęć kierunkami

Hile, Vedantham, Cuellar i in. (2008) uznali, że najlepszym sposobem zobrazowania kierunku jest wyświetlanie strzałek na ekranie telefonu komórkowego. Dobór właściwego kąta strzałki odbywa się przy wykorzystaniu współrzędnych GPS zdjęcia bazy danych, które ma zostać wyświetlone na ekranie telefonu oraz współrzędnych GPS charakterystycznego obiektu obrazu. Jednak jak się okazało taka technika nie radzi sobie dobrze z bardzo dużymi obiektami, jak np. campus uczelni. Stąd też postanowiono podzielić pojedynczą lokalizację GPS obiektu na kilka punktów. Prowadzi to jednak do błędnego obliczania kierunków strzałek nawigacyjnych. Ostatecznie problem ten został rozwiązany przy wykorzystaniu tzw. geokodowania (ang. geocoding), które polega na przyporządkowaniu punktom specjalnych kodów, które umożliwiają jednoznaczną identyfikację każdego miejsca.

W celu przetestowania poprawności działania systemu Landmark-Based Pedestrian Navigation autorzy wybrali 10 ochotników, którzy mieli za zadanie przebyć drogę z punktu A do punktu B korzystając z powyższego systemu (około 15 minut), który dodatkowo wyposażono w zwykłą mapę. Ochotnicy w wieku od 19 do 50 lat przeszli szybki kurs zapoznawczy z aplikacją na telefonie komórkowym. Wszystkim udało się trafić do docelowego punktu, jednak wskazali oni pewne wady systemu:

- niepewność spowodowana wyświetlanymi zdjęciami charakterystycznych obiektów trasy tzn. część obiektów była przysłonięta drzewami, co znacząco utrudniało poprawne rozpoznawanie budynków,
- strzałki były mylące ze względu na brak głębi i niejednoznaczność wyznaczonego kierunku,
- wyświetlane zdjęcia nie odpowiadały dokładnemu miejscu, w którym znajdował się pieszy, co powodowało trudności w dopasowaniu zdjęcia na ekranie z okolicznymi budynkami.

W celu poprawy skuteczności działania tego prototypowego systemu naukowcy wprowadzili znaczne zmiany, które pozytywnie wpłynęły na końcową efektywność aplikacji. Zostały one dokładnie opisane w pracy Hile, Grzeszczuk, Liu i in. z 2009 r. i obejmowały wykorzystanie technik przestrzennego rozumowania do obliczania właściwej pozycji kamery, użycie strzałek jednoznacznie określających kierunek trasy i możliwość rozwoju systemu poprzez dodawanie nowych informacji, jak np. nazwy ulic,

czy uwypuklanie części zdjęcia zawierającej dany obiekt charakterystyczny. Poniżej przedstawiamy przykładowe obrazy wyświetlane na ekranie urządzenia mobilnego pieszego:

Rys. 6.3 Przykłady sposobu wyświetlania kierunków nawigacyjnych. Obiekty oznaczone A, B, C na ostatnim obrazie odpowiadają budynkom na zdjęciach w kolejności od góry do dołu i od lewej do prawej (źródło: Hile, Grzeszczuk, Liu i in. 2009).

Rozdział 7. Podsumowanie pracy

Praca magisterska miała na celu przybliżenie tematyki związanej z wyszukiwaniem obrazów na podstawie zawartości oraz przeanalizowanie głównych problemów i kierunków rozwoju systemów CBIR. Za główne cele pracy postawiliśmy:

- Dokładne opisanie metod stosowanych w technikach wyszukiwania i rozpoznawania obrazów (ang. pattern recognition), które zostały podzielone na dwie zasadnicze części: metody bez interakcji i z interakcją użytkownika.
- Przedstawienie pierwszych i najpopularniejszych systemów CBIR – Content-Based Image Retrieval z dodatkowym uwzględnieniem obecnych systemów i ich możliwości.
- Przeanalizowanie głównych problemów i kierunków rozwoju technik MIR (ang. Multimedia Information Retrieval).
- Wskazanie zastosowań systemów CBIR w różnych dziedzinach życia publicznego ze szczególnych uwzględnieniem aplikacji przeznaczonych dla telefonów komórkowych i urządzeń PDA (ang. Personal Digital Assistant).

Powyższe cele zostały zrealizowane w kolejnych rozdziałach, które dają pełny obraz metod Content-Based Image Retrieval, ich potencjalnych możliwości, ale zarazem i trudności implementacyjnych.

W pracy magisterskiej dokonaliśmy kluczowego podziału technik stosowanych w wyszukiwaniu obrazów na podstawie zawartości wyróżniając dwie grupy: bez interakcji i z interakcją użytkownika. Pierwsza część koncentruje się na dokładnym opisie metod wykorzystujących podstawowe cechy obrazu, czyli jego zawartość wizualną. W szczególności przedstawiono techniki wyboru deskryptorów koloru, tekstury i kształtu. Kolor, jako jedna z najważniejszych i najczęściej wykorzystywanych cech, znalazł szerokie zastosowanie w technikach opisu obrazu. W pracy poświęciliśmy mu wiele uwagi, dokładnie przedstawiając problemy właściwego doboru przestrzeni barw, porównywania histogramów i wykorzystywania momentów i korelogramu koloru. Zagadnienia związane z odpowiednim wyborem cech tekstury opisują początkowe podejścia bazujące np. na cechach teksturowych Tamury, składnikach dekompozycyjnych Wold'a oraz na bardzo popularnych transformatach Gabora i falkowej. W części poświęconej cechom kształtu wyróżniliśmy metody oparte na momentach (w tym na niezmiennikach momentowych), metody aktywnego konturu, obracanego kąta oraz deskryptory Fouriera. Kończąc opis technik bez interakcyjnych przedstawiliśmy zagadnienia identyfikacji podobiektów i określania relacji między nimi oraz dodatkowo podaliśmy stosowane metody indeksowania i redukcji wektora cech (np. PCA – Principal Component Analysis).

Dodatkowo przeanalizowaliśmy problem występowania tzw. „luki semantycznej", czyli różnicy zgodności pomiędzy informacją wyekstrahowaną bezpośrednio z obrazu, a jej

interpretacją, która może się zmieniać w zależności od sytuacji i celu poszukiwań użytkownika. Problem ten wynika z faktu, że żadna pojedyncza cecha lub ich kombinacja nie opisuje obrazu w kontekście jego zawartości. Stąd też w rozdziale tym skoncentrowaliśmy się na analizie przykładowych rozwiązań powyższej sytuacji uwzględniając możliwość interakcji użytkownika z systemem komputerowym. Szczegółowo opisaliśmy temat związany ze sposobem formułowania zapytań oraz z wykorzystaniem technik sprzężenia zwrotnego typu ***relevance feedback***. Na końcu rozdziału podaliśmy interesującą metodę wykorzystującą zasady percepcyjnego grupowania człowieka, jako narzędzie do określania relacji pomiędzy obiektami obrazu.

Dalsza część pracy magisterskiej to pełny przegląd systemów komercyjnych i badawczo - rozwojowych, które miały znaczący wpływ na rozwój obecnych aplikacji CBIR. Opisaliśmy pierwsze systemy wyszukiwania danych multimedialnych na podstawie ich zawartości takie, jak QBIC, VIR Image Engine, Photobook, VisualSEEk i MARS. Załącznik w postaci tabeli zbiorczej systemów daje obraz różnorodności technik stosowanych w powyższych aplikacjach. Następnie podaliśmy obecną sytuację panującą w świecie Content-Based Image Retrieval. Opisaliśmy znane organizacje naukowe, których głównym celem jest rozwój tej dziedziny wiedzy poprzez określenie wspólnych celów i ujednolicenie kryteriów oceny.

Kolejny aspekt, który przedstawiliśmy w pracy, dotyczy głównych problemów metod wyszukiwania obrazów na podstawie zawartości. Wysunęliśmy tezę, że największym problemem powyższych systemów jest brak wspólnych kryteriów oceny skuteczności i efektywności, które pozwoliłyby na porównywanie systemów między sobą i wybór najlepszych spośród nich. Jednocześnie podaliśmy możliwe kierunki rozwoju aplikacji CBIR w narzędziach m. in. do zarządzania multimedialnymi bazami danych, czy ochrony praw autorskich w sieci World Wide Web.

Następnie skoncentrowaliśmy się na opisie praktycznych zastosowań technik wyszukiwania danych na podstawie zawartości w wielu dziedzinach życia publicznego. Metody te zostały rozpowszechnione na skalę światową i są obecnie stosowane np. do zapobiegania przestępczości (detekcja twarzy, odcisków palców), w technikach wojskowych do naprowadzania rakiet samosterujących, w medycynie diagnostycznej, przy projektowaniu wnętrz i ubrań, czy chociażby w reklamie i dziennikarstwie, edukacji, nauce i domowej rozrywce.

Dodatkowo podaliśmy praktyczne rozwiązania metod CBIR w urządzeniach mobilnych. Jest to interesujące zagadnienie, które bazuje na wykorzystaniu telefonów komórkowych i urządzeń PDA do dostarczania pewnych informacji użytkownikowi. Przedstawiliśmy dwa przykłady systemów, których celem jest określanie lokalizacji i wyświetlanie wskazówek nawigacyjnych turystom i pieszym. Są to: Snap2Tell oraz Landmark-Based Pedestrian Navigation System.

W podsumowaniu niniejszej pracy magisterskiej przedstawiliśmy cele, które postawiono sobie przed napisaniem pracy oraz odpowiedzieliśmy na pytanie, w jaki sposób zostały one zrealizowane. Dodatkowo pokrótce opisaliśmy zawartość

merytoryczną poszczególnych części pracy kończąc ją spisem literatury, która posłużyła do opisu i analizy metod i systemów Content-Based Image Retrieval.

Analizując obecne osiągnięcia i możliwości systemów wyszukiwania obrazów na podstawie zawartości należy podkreślić znaczącą rolę interakcji użytkownika z systemem komputerowym w postaci sprzężenia zwrotnego typu ***relevance feedback***. Jest to kluczowy aspekt, dzięki któremu uwzględnia się subiektywną ocenę użytkownika, co bezpośrednio wpływa na końcowy wynik procesu wyszukiwania. Stąd też ważną z praktycznego punktu widzenia kontynuacją dotychczasowych badań wydaje się być implementacja metod sprzężenia zwrotnego, które mogłyby w większym stopniu odpowiadać potrzebom użytkownika. Wymaga to opracowania technik o wysokim stopniu rozumienia bogatego słownictwa człowieka, które byłyby w stanie właściwie interpretować jego zapytania.

Oprócz sprzężenia zwrotnego interesującym kierunkiem dalszych badań może być rozwój lokalnych deskryptorów obrazu takich, jak SIFT, czy SURF. Metody te są obecnie jednymi z najskuteczniejszych i najefektywniejszych, co pozwala sądzić, że w przyszłości mogą być podstawą do konstrukcji technik wyszukiwania informacji jeszcze bardziej uniwersalnych i odpornych na wszelkie zniekształcenia obrazu. W związku z tym przyszłością systemów CBIR jest zatem rozwój organizacji badawczych takich, jak ImageCLEF, ImagEVAL, czy PASCAL, które poprzez zrzeszanie grup naukowych z całego świata sprzyjają osiągnięciu wyżej wymienionych celów.

Literatura

1. Abate A. F., Nappi M., Ricciardi S., Tortora G. 2004. *FACES: 3D FAcial reConstruction from anciEnt Skulls using content based image retrieval,* Journal of Visual Languages and Computing, vol. 15, pp. 373-389
2. Ahmad I., Abdullah S., Kiranyaz S., Gabbouj M. 2005. *Content-based image retrieval on mobile devices*, Proc. of SPIE (Multimedia on Mobile Devices), vol. 5684, pp. 16-20
3. Alt H., Behrends B., Blömer J. 1991. *Approximate Matching of Polygonal Shapes*, Proc. Seventh ACM Symposium on Computational Geometry, pp. 186-193
4. Amini A., Tehrani S., Weymouth T. 1988. *Using dynamic programming for minimizing the energy of active contours in the presence of hard constraints*, ICCV
5. Antani S., Long L. R., Thoma G. R. 2002. *A biomedical information system for combined content-based retrieval of spine X-ray images and associated text information*, Proc. of the 3rd Conference on Computer Vision, Graphics and Image Processing, pp. 242-247
6. Assfalg J., Bimbo A. D., Pala P. 2000. *Using Multiple Examples for Content-based Retrieval,* Proc. Int'l Conference on Multimedia and Expo
7. Bach J., Fuller C., Gupta A., Hampapur A. i in. 1996. *Virage image search engine: an open framework for image management*, In Proc. Of the SPIE, Storage and Retrieval for Image and Video Databases IV, pp. 76-87
8. Bay H., Tuytelaars T., Van Gool L. 2006. *SURF: Speeded Up Robust Features*, In ECCV
9. Bay H., Ess A., Tuytelaars T., Van Gool L. 2008. *SURF: Speeded Up Robust Features*, Computer Vision and Image Understanding (CVIU), vol. 110, pp. 346-359
10. Beckmann N., et al, 1990. *The R*-tree: An efficient robust access method for points and rectangles,* ACM SIGMOD Int. Conference on Management of Data
11. Belongie S., Carson C., Greenspan H., Malik J. 1998. *Color- and Texture-Based Image Segmentation Using EM and Its Application to Content-Based Image Retrieval,* Sixth International Conference on Computer Vision (ICCV'98), pp. 675
12. Berchtold S., Keim D. A., Kriegel H-P. 1996. *The X-tree: an index structure for high-dimensional data*, Proceedings of the 22nd VLDB Conference Mumbai (Bombay), pp. 28-39
13. Berman A. P., Shapiro L. G. 1999. *Triangle-inequality-based pruning algorithms with triangle tries*, Proceedings of the SPIE Conference on Storage and Retrieval for Image and Video Databases
14. Berman A. P., Shapiro L. G. 1999. *Efficient Content-Based Image Retrieval: Experimental Results*, Proc. of the IEEE Workshop on Content-Based Access of Image and Video Libraries, pp. 55-61
15. Brodatz P. 1966. *Textures: A Photographic Album for Artists and Designers*, Dover, New York

16. Bunke H. 2000. *Graph matching: Theoretical foundations, algorithms, and applications*, Proc. Vision Interface, pp. 82-88
17. Burns J. B., Hanson A. R., Riseman E. M. 1986. *Extracting straight lines*, IEEE Trans. Pattern Anal. Mach. Intell., vol. 8, pp. 425-455
18. Carreira M. J., Orwell J., Turnes R., Boyce J. F., Cabello D. i in. 1998. *Perceptual Grouping from Gabor Filter Responses*, Proceedings of the Ninth British Machine Vision Conference BMVC 98, UK
19. Carson C., Thomas M., Belongie S. i in., Springer 1999. *Blobworld: A system for region-based image indexing and retrieval*, In D. P. Huijsmans and A. W. M. Smeulders, ed. *Visual Information and Information System*, Proceedings of the Third Inter. Conference VISUAL'99
20. Chang N. S., Fu K. S. 1980. *Query-by-Pictorial Example*, IEEE Transactions on Software Engineering, vol. 6, pp. 519-524
21. Chang S. K., Shi Q. Y., Yan C. Y. 1987. *Iconic indexing by 2-D strings*, IEEE Trans. On Pattern Anal. Machine Intell., vol. 9, pp. 413-428
22. Chang S. K., Jungert E., Li Y. 1988. *Representation and retrieval of symbolic pictures using generalized 2D string*, Technical Report, University of Pittsburgh
23. Chevallet J-P., Lim J-H., Leong M-K. 2007. *Object identification and retrieval from efficient image matching. Snap2Tell with the STOIC dataset*, Information Processing and Management vol. 43, pp. 515-530
24. Chew M., Tygar J. D. 2004. *Image recognition CAPTCHAs*, In 7th Annual Information Security Conference, pp. 268-279
25. Clark M., Bovik A. 1987. *Texture Segmentation Using Gabor Modulation/Demodulation*, Patt. Recogn. Lett., 6, pp. 261-267
26. Clough P., Müller H., Sanderson M. 2005. *The CLEF Cross Language Image Retrieval Track (Image CLEF) 2004*, Peters C. et al. (Eds.): *CLEF 2004*, LNCS 3491, pp. 597-613
27. Comaniciu D., Meer P., Xu K., Tyler D. 1999. *Retrieval Performance Improvement through Low Rank Corrections*, IEEE Workshop on Content-Based Access of Image and Video Lib., pp. 50-54
28. Cox I. J., Mileer M. L., Omohundro S. M., Yianilos P. N. 1996. *Target testing and the PicHunter Bayesian multimedia retrieval system*, Advances in Digital Libraries, Library of Congress, pp. 66-75
29. Cox I. J., Miller M. L., Minka T. P. i in. 2000. *The Bayesian Image Retrieval System, PicHunter: Theory, Implementation, and Psychophysical Experiments*, IEEE Trans. on Image Processing, vol. 9, pp. 20-37
30. Datta D., Joshi D., Li J., Wang J. Z.2008. *Image retrieval: Ideas, influences, and trends of the new age*, ACM Computing Surveys (CSUR), vol. 40, issue 2, article no.: 5
31. Daubechies I. 1990. *The wavelet transform, time-frequency localization and signal analysis*, IEEE Trans. On Information Theory, vol. 36, pp. 961-1005
32. Deb S. i in. 2004. *Multimedia Systems and Content-Based Image Retrieval*, Idea Group Publishing

33. Del Bimbo A. 1999. *Visual information retrieval.* Morgan Kaufmann Publisher, Inc. San Francisco, CA
34. Eakins J. P. 1993. *Design criteria for a shape retrieval system*, Computers in Industry, vol. 21, pp. 167-184
35. Eakins J. P., Graham M. E. 1999. *Content-based image retrieval*, A report to the JISC Technology Applications Programme. Technical report, Institute for Image Data Research, University of Northumbria at Newcastle, UK
36. Feiner S., MacIntyre B., Hollerer T., Webster A. 1997. *A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment*, Proc. ISWC '97 (First IEEE International Symposium on Wearable Computers), vol. 1, pp. 208-217
37. Feng H., Shi R., Chau T. S. 2004. *A bootstrapping framework for annotating and retrieving WWW images*, Proceedings of the 12th annual ACM International Conference on Multimedia, pp. 960-967
38. Finlayson G. D. 1996. *Color in perspective*, IEEE Trans. On Pattern Analysis and Machine Intelligence, vol. 8, pp. 1034-1038
39. Floriani L., Falcidieno B. 1998. *A Hierarchical Boundary Model for Solid Object Representation*, ACM Transactions on Graphics, vol. 7, no. 1
40. Francos J. M. 1993. *Orthogonal decompositions of 2-d random fields and their applications in 2-D spectral estimation.* N. K. Bose and C. R. Rao, editors, Signal Processing and its Applications, pp. 207-227
41. Francos J., Meiri A., Porat B. 1993. *A Unified Texture Model Based on a 2-D Wold-Like Decomposition*, IEEE Trans. Sig. Process, 41, pp. 2665-2678
42. Francos J., Narasimhan A., Woods J. 1996. *Maximum Likelihood Parameter Estimation of Discrete Homogeneous Random Fields with Mixed Spectral Distributions*, IEEE Trans. Sig. Process, 44(5), pp. 1242-1255
43. Fu H., Chi Z., Feng D. 2006. *Attention-driven image interpretation with application to image retrieval*, Pattern Recognition, vol. 39, pp. 1604-1621
44. Gao Y., Qi Y. 2005. *Robust visual similarity retrieval in single model face databases*, Pattern Recognition, vol. 38, pp. 1009-1020
45. Gonzalez R. C., Woods R. E. 1993. *Digital Image Processing*, Addison Wesley
46. Granlund G. 1972. *Fourier Preprocessing for Hand Print Character Recognition*, IEEE Trans. Computers, vol. 21, pp. 195-201
47. Haindl M. 1991. *Texture Synthesis*, CWI Quarterly 4, pp. 305-331
48. Haralick R. M., Shanmugam K., Dinstein I 1973. *Textural Features for Image Classification*, IEEE Transactions on Systems, Man, and Cybernetics, 3(6), pp. 610-621
49. Haralick R.M. 1979. *Statistical and Structural Approaches to Texture*, IEEE, 67, pp. 786-804
50. Hare J. S., Lewis P. H. 2005. *Content-based image retrieval using a mobile device as a novel interface*, In R. W. Lienhart, N. Babaguchi, E. Y. Chang (Eds.) *Storage and Retrieval Methods and Applications for Multimedia 2005*, pp. 64-75
51. Haritaoglu I. 2001. *InfoScope: Link from Real World to Digital Information Space*, Proc. of the 3rd International Conference on Ubiquitous Computing, pp. 247-255

52. He Q. 1997. *An evaluation on MARS – an image indexing and retrieval system*, Technical report, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign
53. Heidemann G. 2004. *Combining spatial and colour information for content based image retrieval*, Computer Vision and Image Understanding vol. 94, pp. 234-270
54. Hile H., Vedantham R., Cuellar G., Liu A. i in. 2008. *Landmark-Based Pedestrian Navigation from Collections of Geotagged Photos*, Proc. of the 7th International Conference on Mobile and Ubiquitous Multimedia, pp. 145-152
55. Hile H., Grzeszczuk R., Liu A., Vedantham R. 2009. *Landmark-Based Pedestrian Navigation with Enhanced Spatial Reasoning*, Proc. of the 7th International Conference on Pervasive Computing, pp. 59-76
56. Hirata K., Kato T. 1992. *Query by visual example - content based image retrieval*, In Advances in Database Technologies (EDTB '92), vol. 580, pp. 56-71
57. Holt B., Hartwick L. 1994. *Retrieving art images by image content: the UC Davis QBIC project*, Aslib Proceedings, vol. 46, pp. 243-248
58. Hu M. K. 1962. *Visual Pattern Recognition by Moment Invariants*, IRE Transactions on Information Theory, vol. IT-8, pp. 179-187
59. Huang J., i in. 1997. *Image indexing using color correlogram*, IEEE International Conference on Computer Vision and Pattern Recognition, pp. 762-768, Puerto Rico
60. Huang J., Kumar S. R., Metra M. 1997. *Combining supervised learning with color correlograms for content-based image retrieval*, Proc. of ACM Multimedia'95, pp. 325-334
61. Huijsmans D. P., Sebe N. 2005. *How to Complete Performance Graphs in Content-Based Image Retrieval: Add Generality and Normalize Scope*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, pp. 245-251
62. Huttenlocher D. P., Klanderman G. A., Rucklidge W.J. 1993. *Comparing images using the Hausdorff distance*, IEEE Trans. Pattern Anal. Machine Intell., vol. 15, pp. 850-863
63. Hwang W-S., Weng J.J., Fang M., Qian J. 1999. *A Fast Retrieval Algorithm with Automatically Extracted Discriminant Features*, Proc. of the IEEE Workshop on Content-Based Access of Image and Video Libraries, pp. 8-12
64. Iqbal Q., Aggarwal J. K. 1999. *Applying perceptual grouping to content-based image retrieval: Building images*, Proc, of the IEEE International Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 42-48
65. Iqbal Q., Aggarwal J.K. 2002. *Retrieval by classification of images containing large manmade objects using perceptual grouping*, Pattern Recognition, vol. 35, pp. 1463-1479
66. Ishikawa Y., Subramanya R., Faloustos C. 1998. MindReader: *Query Databases Trough Multiple Examples*, Proc. of the 24th VLDB Conference
67. Jiang H., Ngo Ch-W., Tan H-K. 2006. *Gestalt-based feature similarity measure in trademark database*, Pattern Recognition, vol. 39, pp. 988-1001
68. Jolliffe I. 1986. *Principal Component Analysis*, Springer Verlag, New York
69. Kambhatla N., Leen T. K. 1997. *Dimension reduction by local principal component analysis*, Neural Computation, vol. 9, pp. 1493-1516

70. Kang K., Yoon Y., Choi J., Kim J., Koo H., Choi Jong-Ho 2007. *Additive texture information extraction using color coherence vector*, Proceedings of the 7th WSEAS International Conference on Multimedia Systems & Signal Processing
71. Kass M,. Witkin A., Terzopoulos D. 1988. *Snakes: Active contour models*, IJCV 1, pp. 321-331
72. Katsumata N., Matsuyama Y. 2005. *Database retrieval for similar images using ICA and PCA bases*, Engineering Applications of Artificial Intelligence, vol. 18, pp. 705-717
73. Ke Y., Sukthankar R., Huston L. 2004. *Efficient near-duplicate detection and subimage retrieval*, Proc. of the ACM International Conference on Multimedia, pp. 1150-1157
74. Konstantinidis K., Gasteratos A., Andreadis I., IRM Press 2007. *The Impact of Low-Level Features in Semantic-Based Image Retrieval*, In Zhang YJ. ed. *Semantic-Based Visual Information Retrieval*
75. Kunttu I., Lepistö L., Rauhamaa J., Visa A. 2006. *Multiscale Fourier descriptors for defect image retrieval*, Elsevier Science Inc. vol. 27, pp. 123-132
76. Laaksonen J., Koskela M., Laakso S., Oja E. 2000. *PicSOM – content-based image retrieval with self-organizing maps*, Pattern Recognition Letters 21, no. 13-14, pp. 1199-1207
77. Lawrence S., Giles C. L., Tsoi A. C., Back A. D. 1997. *Face recognition: a convolutional neural-network approach*, IEEE Trans. Pattern Anal. Mach. Intell., vol. 8, pp. 98-113
78. Lee C., Ma W. Y., Zhang H. J. 1999. *Information Embedding Based on user's relevance Feedback for Image Retrieval*, Proc. of SPIE International Conference on Multimedia Storage and Archiving Systems, vol. 4, pp. 19-22
79. Lee S. Y., Hsu F. H. 1990. *2D C-string: a new spatial knowledge representation for image database systems*, Pattern Recognition, vol. 23, pp. 1077-1087
80. Lee S. Y., M. C. Yang, J. W. Chen 1992. *2D B-string: a spatial knowledge representation for image database system*, Proc. ICSC'92 Second Int. Computer Sc. Conf., pp. 609-615
81. Lehmann T. M., Güld M. O., Deselaers T. i in. 2005. *Automatic categorization of medical images for content-based retrieval and data mining*, Computerized Medical Imaging and Graphics, vol. 29, pp. 143-155
82. Lemke H. U. 2003. *PACS developments in Europe*, vol. 27, pp. 111-120
83. Lew M. S. (Ed.) 2001. *Principles of Visual Information Retrieval.* Springer-Verlag London
84. Lew M. S., Sebe N., Djeraba C., Jain R. 2006. *Content-based multimedia information retrieval: State of the art and challenges*, ACM Trans. On Multimedia Computing, Communications, and Applications, vol. 2, pp. 1-19
85. Liu F., Picard W. 1996. *Periodicity, directionality, and randomness: Wold features for image modeling and retrieval*, IEEE Trans. On Pattern Analysis and Machine Learning, vol. 18, pp. 722-733
86. Liu Y., Zhang D., Lu G., Ma W-Y. 2007. *A survey of content-based image retrieval with high-level semantics*, Pattern Recognition, vol. 40, pp. 262-282

87. Loncaric S. 1998. *A survey of shape analysis techniques*, Pattern Recognition, 31(8), pp. 983-1001
88. Long F., Zhang H., Dagan Feng D., Springer-Verlag Berlin Heidelberg New York 2003. *Fundamentals of Content-Based Image Retrieval*, In Feng D., Siu W. C., Zhang H. J. (Eds.) *Multimedia Information Retrieval and Management*
89. Lowe D. G. 1985. *Perceptual Organization and Visual Recognition*, Kluwer Academic Publishers
90. Lowe D. G. 1999. *Object Recognition from Local Scale-Invariant Features*, Proceedings of the Seventh IEEE International Conference on Computer Vision, vol. 2, pp. 1150-1157
91. Lowe D.G. 2004. *Distinctive Image Features from Scale-Invariant Keypoints*, International Journal of Computer Vision, vol. 60, pp. 91-110
92. Lu K., He X. 2005. *Image retrieval based on incremental subspace learning*, Pattern Recognition, vol. 38, pp. 2047-2054
93. Ma W. Y., Manjunath B. S. 1997. *Edge flow: a Framework of Bondary detection and image segmentation*, Proc. IEEE International Conference on Computer Vision and Pattern Recognition
94. Ma W. Y., Manjunath B. S. 1999. *Netra: A toolbox for navigating large image databases*, Multimedia Systems, vol. 7, pp. 184-198
95. MacArthur S. D., Brodley C. E., Shyu C.-R. 2000. *Relevance feedback decision trees in content-based image retrieval*, Proc. of IEEE Workshop on Content-based Access of Image and Video Libraries, pp. 68-72
96. Mai W., Doods G., Tweed C. 2003. *A PDA-based system for recognizing building from user-supplied images*, Lecture Notes in Computer Science, Springer-Verlag Heidelberg, pp. 143-157
97. Mikołajczyk K., Schmid C. 2005. *A Performance Evaluation of Local Descriptors*, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 27, pp. 1615-1630
98. Minka T. P., Picard R. W. 1997. *Interactive learning using a "society of models"*, Special issue of Pattern Recognition on Image Databases, 30(4)
99. Moëllic P-A., Fluhr C. 2006. *ImagEVAL 2006 Official campaign*
100. Mori G., Malik J. 2003. *Recognizing objects in adversarial clutter: Breaking a visual captcha*, Proc. IEEE Conf. Computer Vision and Pattern Recognition, vol. 1, pp. 134-141
101. Muneesawang P., Guan L. 2006. *Multimedia Database Retrieval: A Human-Centered Approach*, Springer, Signals and Communication Technology
102. Müller H., Squire D. M., Müller W., Pun T. 1999. *Efficient access methods for content-based image retrieval with inverted files*, Proc. of Multimedia Storage and Archiving Systems IV, vol. 3846, pp. 461-472
103. Müller H., Müller W., Squire D. M., Marchand-Maillet S., Pun T. 2001. *Performance Evaluation in Content-Based Image Retrieval: Overview and Proposals*, Pattern Recognition Letters, vol. 22, pp. 593-601
104. Müller H., Michoux N., Bandon D., Geissbuhler A. 2004. *A review of content-based image retrieval systems in medical applications – clinical benefits and future directions*, International Journal of Medical Informatics, vol. 73, pp. 1-23

105. Müller H., Deselaers T., Lehmann T. i in. 2007. *Overview of the ImageCLEFmed 2006 Medical Retrieval and Medical Annotation Tasks*, Springer Lecture Notes in Computer Science (LNCS 4730), pp. 595-608
106. Müller H., Deselaers T., Kim E. i in. 2008. *Overview of the ImageCLEFmed 2007 Medical Retrieval and Annotation Tasks*, Springer Lecture Notes in Computer Science (LNCS 5152), pp. 473-491
107. Niblack W., Barber R., Equitz W., Flicker M. i in. 1993. *The QBIC project: Querying images by content using color, texture, and shape*, In Proceedings of the SPIE Conference on Storage and Retrieval for Image and Video Databases, pp. 173-187
108. Oliveira M. C., Cirne W., Marques P. M. de Azevedo 2007. *Towards applying content-based image retrieval in the clinical routine*, Future Generation Computer Systems, vol. 23, pp. 466-474
109. Ortega M., Rui Y., Chakrabarti K. i in. 1997. *Supporting similarity queries in MARS*, Proc. of the 5th ACM Inter. Multimedia Conference, pp. 403-413
110. Otterloo P. J. 1992. *A Contour-Oriented Approach to Shape Analysis*, Hemel Hampstead
111. Ozer B., Wolf W., Akansu A. N. 1999. *A Graph Based Object Description for Information Retrieval in Digital Image and Video Libraries*, Journal of Visual Communication and Image Representation, vol. 13, pp. 425-459
112. Papathomas T. V., Conway T. E., Cox I. J. i in. 1998. *Psychophysical studies of the performance of an image database retrieval system*, IS&T/SPIE Conference on Human Vision and Electronic Imaging III, pp. 591-602
113. Pass G., Zabith R. 1996. *Histogram refinement and content-based image retrieval*, IEEE Workshop on Applications of Computer Vision, pp. 96-102
114. Pass G., Zabith R., Miller J. 1997. *Comparing Images Using Color Coherence Vectors*, Proceedings of the fourth ACM International Conference on Multimedia
115. Pawlik P., Mikrut S. 2006. *Wyszukiwanie punktów charakterystycznych na potrzeby łączenia zdjęć lotniczych*, Automatyka, tom 10, zeszyt 3
116. Pentland A., Picard R. W., Sclaroff S. 1996. *Photobook: Content-based manipulation of image databases*, International Journal of Computer Vision, vol. 18, pp. 233-254
117. Prokop R. J., Reeves A. P. 1992. *A survey of moment-based techniques for unoccluded object representation and recognition*, CVGIP: Graphics Models and Image Processing, vol. 54, pp. 438-460
118. Quack T., Mönich U., Thiele L. i in. 2004. *Cortina: a system for large-scale, content-based web image retrieval*, Proceedings of the 12th annual ACM International Conference on Multimedia, pp. 508-511
119. Ratan A. L., Maron O., Grimson W. E. L., Lozano-Perez T. 1999. *A framework for learning query concepts in image classification*, Proc. of the Computer Vision and Pattern Recognition, vol. 1, pp. 1423-1429
120. Robinson J. T. 1981. *The k-d-B-tree: a search structure for large multidimensional dynamics indexes*, Proc. Of SIGMOD Conference

121. Rocchio J. J. 1971. *Relevance Feedback in Information Retrieval*, In G. Salton ed., *The SMART Retrieval System – Experiments in Automatic Document Processing*, pp. 313-323, Prentice Hall
122. Rogowitz B. E., Frese T., Smith J. i in. 1998. *Perceptual image similarity experiments*, IS&T/SPIE Conference on Human Vision and Electronic Imagening, vol. 3299
123. Rotter P. 2003. *Zastosowanie metod optymalizacji wielokryterialnej w interpretacji obrazów*, praca doktorska Kraków
124. Rucklidge W. J. 1997. *Efficiently Locating Objects Using the Hausdorff Distance*, International Journal of Computer Vision, vol. 24, pp. 251-270
125. Rui Y., Huang T. S., Mehrotra S. 1997. *Content-based image retrieval with relevance feedback in MARS*, Proceedings of International Conference on Image Processing, vol. 2, pp. 815-818
126. Rui Y., Huang T. S. , Mehrotra S. 1998. *Relevance Feedback Techniques in Interactive Content-Based Image Retrieval*, Proc. of IS&T and SPIE Storage and Retrieval of Image and Video Databases VI, pp. 25-36
127. Rui Y., Huang S., Ortega M., Mehrotra S. 1998. *Relevance Feedback: A Power Tool for Interactive Content-Based Image Retrieval*, IEEE Trans. on Circuits and Video Technology, vol. 8, pp. 644-655
128. Rui Y., Huang T. S. 1999. *A Novel Relevance Feedback Technique in Image Retrieval*, Proc. of 7th ACM International Conference on Multimedia, pp. 67-70
129. Rui Y., Huang T. S., Chang S.F. 1999. *Image retrieval: Current techniques, Promising Directions, and Open Issues*, Journal of Visual Communication and Image Representation 10, pp. 39-62
130. Salton G., McGill M. J. 1983. *Introduction to Modern Information Retrieval*, McGraw-Hill, Inc.
131. Schalkoff R. J. 1991. *Pattern recognition: statistical, structural and neural approaches*, John Wiley & Sons, Inc., New York
132. Sebe N., Lew M. S., Springer-Verlag London 2001. *Texture Features for Content-Based Retrieval*, In M. S. Lew (Ed.) *Principles of Visual Information Retrieval*
133. Sebe N., Lew M. S. 2002. *Robust Shape Matching*, International Conference on Image and Video Retrieval (CIVR'02). pp. 17-28
134. Setchell C., Campbell N. 1999. *Using colour Gabor texture features for scene understanding*, In 7th International Conference on Image Processing and its Applications, pp. 372-376
135. Shirahatti N. V., Barnard K. 2005. *Evaluating image retrieval*, Proc. of the Computer and Pattern Recognition (CVPR), vol. 1, pp. 955-961
136. Sim DG., Kwon OK., Park RH. 1999. *Object matching algorithms using robust Hausdorff distance measures*, IEEE Trans. Image Process. vol. 8, pp. 425-429
137. Smeulders A. W. M., Worring M., Santini S. i in. 2000. *Content-Based Image Retrieval at the End of the Early Years*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, pp.1349-1380

138. Smith J. R., Chang SF. 1996. *Transform Features for Texture Classification and Discrimination in Large Image Databases*, International Conference on Image Processing, pp. 407-411
139. Smith J. R., Chang SF. 1997. *VisualSEEk: a fully automated content-based image query system*, Proc. of the fourth ACM International Conferene on Multimedia, pp. 87-98
140. Smolka B., Szczepański M., Lukac R, Venetsanoloulos A. N. 2004. *Robust color image retrieval for the World Wide Web*, Proceedings of the ICASSP, vol. 3, pp. 461-464
141. Squire D. M., Müller W., Müller H., Raki J. 1999. *Content-based query of image databases, inspirations from text retrieval: Inverted files, frequency-based weights and relevance feedback*, In the 11th Scandinavian Conference on Image Analysis (SCIA'99), pp. 143-149
142. Stricker M., Orengo M. 1995. *Similarity of Color Images*, SPIE Storage and Retrieval for Image and Video Databases III, vol. 2185, pp. 381-392
143. Su Z., Zhang H. J., Li S., Ma S. 2003. *Relevance Feedback in Content-Based Image Retrieval: Bayesian Framework, Feature Subspaces, and Progressive Learning*, IEEE Trans. On Image Processing, vol. 12, no. 8
144. Sural S., Idea Group Publishing 2004. *Histogram Generation from the HSV Color Space Using Saturation Projection*, In S. Deb ed. *Multimedia Systems and Content-Based Image Retrieval*
145. Tadeusiewicz R., Flasiński M. 1991. *Rozpoznawanie obrazów*. Państwowe Wydawnictwo Naukowe Warszawa
146. Tamura H., Mori S., Yamawaki T. 1978. *Texture features corresponding to Visual perception*, IEEE Transaction On Systems, Man, and Cybernetics, vol. SMC-8, pp. 460-472
147. Teague M. R. 1980. *Image Analysis via the General Theory of Moments*, Journal of the Optical Society of America, vol. 70, pp. 920-930
148. Turner M. 1986. *Texture Discrimination by Gabor Functions*, Biol. Cybern., 55, pp. 71-82
149. Vailaya A., Figueiredo M. A. G., Jain A. K., Zhang H. J. 2001. *Image Classification for Content-based Indexing*, IEEE Transaction on Image Processing, vol. 10, no. 1
150. Vasconcelos N., Lippman A. 1999. *Learning from User Feedback in Image Retrieval Systems*, Prof of Neural Information Processing Systems 12, pp. 843-849
151. Vasconcelos N., Lippman A. 1999. *Probabilistic retrieval: new insights and experimental results*, IEEE Workshop on Content-Based Access on Image and Video Lib., pp. 62-66
152. Veltkamp R. C., Hagedoorn M. 1999. *State-of-the-art in shape matching*, Technical Report UU-CS-1999-27, Ultrecht University
153. Veltkamp R. C., Tanase M. 2000. *Content-Based Image Retrieval Systems: A Survey*, Technical Report UU-CS-2000-34

154. Vendrig J., Worring M., Smeulders A. W. M. 1999. *Filter image browsing: exploiting interaction in retrieval*, Proc. Viusl'99: Information and Information System
155. Venters C. C., Hartley R. J., Hewitt W. T., Idea Group Publishing 2004. *Mind the gap: content-based image retrieval and user interface*, In S. Deb ed. *Multimedia Systems and Content-Based Image Retrieval*
156. Voorhees H., Poggio T. 1988. *Computing texture boundaries from images*, Nature, 333, pp. 364-367
157. Wang J. Z., Boujemaa N., Del Bimbo A. i in. 2006. *Diversity in multimedia information retrieval research*, In Proceedings of the ACM SIGMM International Workshop on Multimedia Information Retrieval (MIR) at the International Conference on Multimedia
158. Weber R., Schek H-J., Blott S. 1998. *A quantitative analysis and performance study for similarity-search methods in high-dimensional spaces*, Proceedings of the 24th VLDB Conference, pp. 194-205
159. Wong E. K. 1992. *Model matching in robot vision by subgraph isomorphism*, Pattern Recognition vol. 25, pp. 287-303
160. Wong WT., Shih F. Y., Liu J. 2007. *Shape-based image retrieval using support vector machines, Fourier descriptors and self-organizing maps*, Information Sciences: an International Journal, vol. 177, pp. 1878-1891
161. Yadav R. B., Nishchal N. K., Gupta A. K., Rastogi V. K. 2007. *Retrieval and classification of shape-based objects using Fourier, generic Fourier, and wavelet-Fourier descriptors technique: A comparative study*, Optics and Lasers in Engineering, vol. 45, pp. 695-708
162. Yang D., Garrett J. H. Jr., Shaw D. S., Rendell L. A. 1994. *An Intelligent Symbol Usage Assistant for CAD Systems*, IEEE Expert: Intelligent Systems and Their Applications, vol. 9, pp. 32-41
163. Yoo H.W., Jang D.S., Jung S.H., Park J.H., Song K.S. 2002. *Visual information retrieval system via content-based approach*, Pattern Recognition, vol. 35, pp. 749-769
164. Yoshino M., Tanniguchi M., Imaizumi K. 2005. *A new retrieval system for database of 3D facial images*, Forensic Science International, vol. 148, pp. 113-120
165. Yu C., Ooi B. C., Tan K-L., Jagadish H. V. 2001. *Indexing the distance: an efficient method to KNN processing*, Proceedings of the 27th VLDB Conference, pp. 421-430
166. Yu K., Ma W-Y., Tresp V. i in. 2003. *Knowing a tree from the forest: Art Image retrieval using a society of profiles*, Proc. of the eleventh ACM International Conference on Multimedia, pp. 622-631
167. Zahn C., Roskies R. 1972. *Fourier descriptors for plane closed curves*, Computer Graphics and Image Processing, pp. 269-281
168. Zhang D., Lu G. 2001. *Shape Retrieval Using Fourier Descriptors*, Proc. Int. Conference on Multimedia and Distance Education (ICMADE'01), pp. 1-9
169. Zhang D., Lu G. 2002. *Shape-based image retrieval using generic Fourier descriptor*, Signal Processing: Image Communication, vol. 17, pp. 825-848

170. Zhang D., Lu G. 2003. *A comparative study of curvature scale space and Fourier descriptors for shape-based image retrieval*, Journal of Visual Communication and Image Representation, vol. 14, pp. 41-60
171. Zhang D-Q., Chang S-F. 2004. *Detecting image near-duplicate by stochastic attributed relational graph matching with learning*, Proc. of the 12th annual ACM International Conference on Multimedia, pp. 877-884
172. Zhang H. J. Springer-Verlag Berlin Heidelberg New York 2003. *Relevance feedback in content-based image retrieval*, In Feng D., Siu W. C., Zhang H. J. (Eds.) *Multimedia Information Retrieval and Management*
173. Zhang H. J., Zhong D. 1995. *A Scheme for visual feature-based image indexing*, Proc. of SPIE Conf. on Storage and Retrieval for Image and Video Databases III, pp. 36-46
174. Zhang Y. J. (Ed.) 2007. *Semantic-Based Visual Information Retrieval*, IRM Press

Strony internetowe:

175. http://www.vision.ee.ethz.ch/~surf/index.html - strona poświęcona metodzie SURF