



Wyszukiwanie multimedialnych

Multimedia to dowolne dane cyfrowe video, dźwięk, zwykłe teksty, zazwyczaj bez struktury, używane do komunikacji i przechowywania informacji

Wyszukiwanie multimedialnych wymaga także ich rankingu według stopnia podobieństwa do zapytania

Dla celów wyszukiwania użytkownik może opisać scenę video np.: „Keanu Reeves unikający pocisków podczas zderzenia helikopterów w filmie Matrix”

Wyszukiwanie informacji multimedialnych obejmuje następujące tematy:

- reprezentacja treści i obiektów multimedialnych,
 - wydobywanie cech,
 - formułowanie pytań odwzorowujących semantykę wysokopoziomową na cechy niskiego poziomu
-



Wyszukiwanie multimediów

zapytania przez przykłady

sprzężenie informacyjne, pytania interaktywne,

indeksowanie i katalogowanie cech,

zintegrowane wyszukiwanie typów searching i browsing,

przeszukiwanie multimediów w oparciu o ich treści.

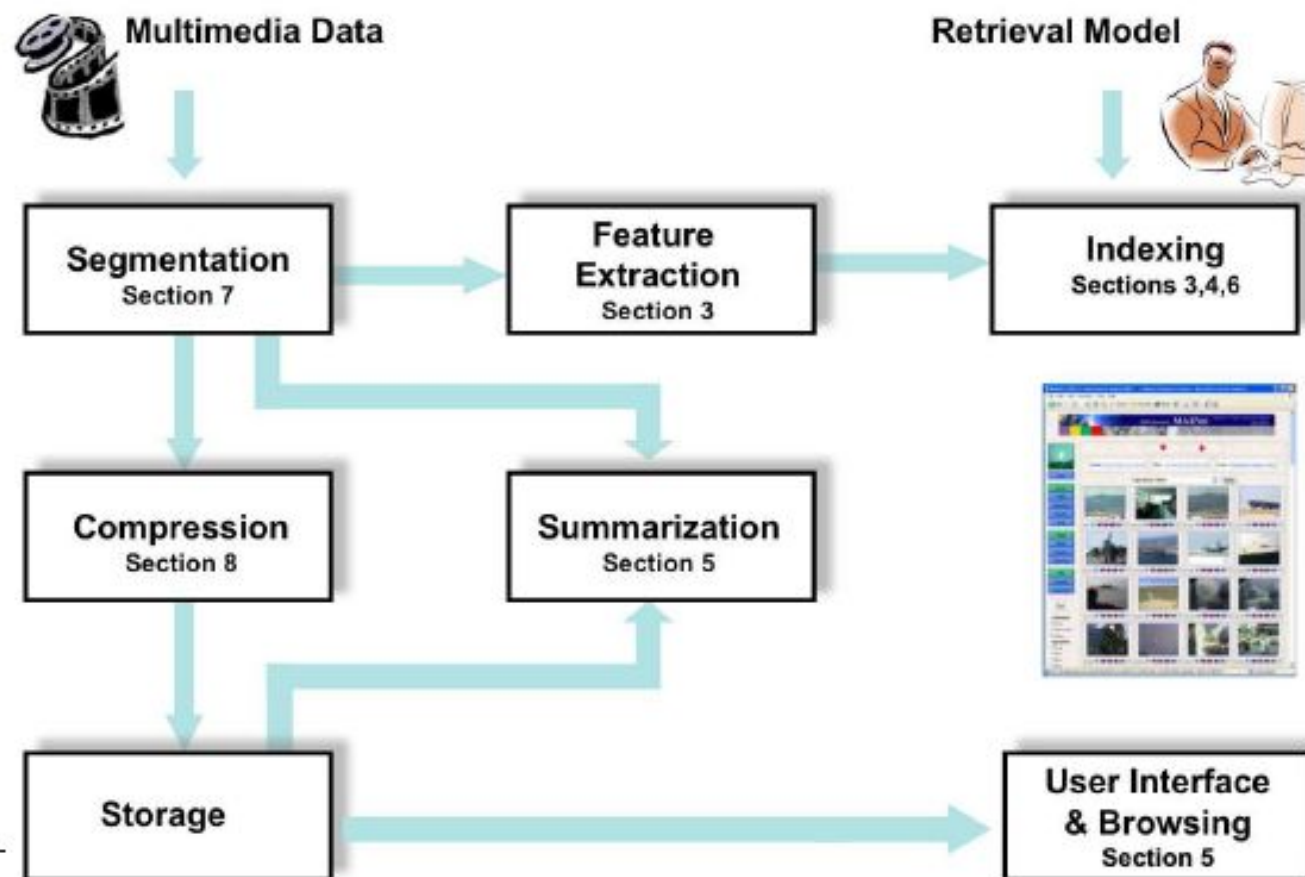
W tekście słowa, znaki przestankowe i paragrafy tworzą jego strukturę

W multimediach dane są zazwyczaj jednym, nieprzerwanym strumieniem; pożądane jest zdefiniowanie jednostek semantycznych

Wyszukiwanie multimedialnych

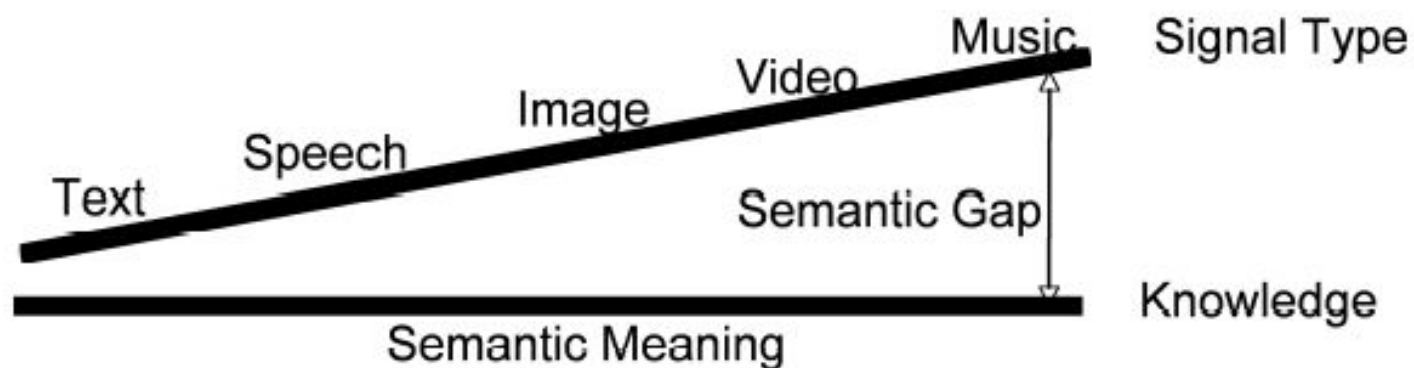
Przepływ informacji przy wyszukiwaniu multimedialnych:

High-Level Multimedia IR Software Architecture



Wyszukiwanie multimedialne

Między treścią sygnału multimedialnego i jego znaczeniem występuje duża luka semantyczna.



Wyszukiwanie multimediów

Identyfikacja obiektów: istotny problem w przetwarzaniu obrazu i dźwięku

ludzie patrząc na obraz rozpoznają twarze i inne obiekty
automatyczne etykietowanie obiektów w obrazach lub
rozpoznawanie dźwięków w sygnałach audio są
problemami nadal nie rozwiązanymi globalnie

Dlatego systemy wyszukiwania multimediów często
wykorzystują opisy tekstowe wytworzone przez człowieka

Systemy opisu multimediów często wykorzystują
interpretacje emocjonalne

W przypadku mowy informacja niesemantyczna jest
przekazywana przez tzw. prozodia sygnału – akcent,
intonacja i iloczyn.

Np. prozodia pozwalają rozróżnić pomiędzy zdaniami „Nie
zatrzymuj się” i „Nie zatrzymuj się!”



Wyszukiwanie multimediów

Apertura:

Ze względu na duże ilości danych ruch w sekwencji obrazów testuje się na małej porcji każdego z obrazów określonej przez aperturę; utrudnia to interpretację ruchu na obrazach video

Wyszukiwanie obrazów z treści

Polega na identyfikowaniu i wydobywaniu najważniejszych cech opisujących treść obrazu

Zapytanie przez przykład (*Query By Example* - QBE):

użytkownik posługuje się przykładowym obrazem jako zapytaniem o podobne do niego i ignoruje informację sementyczną związaną z obrazem

Najlepsze rankingi bazują na cechach obrazu takich jak: poza, ogniskowa kamery, oświetlenie, pozycja kamery, charakter ruchu.

Typowe cechy obrazów wyszukiwane przy zapytaniach QBE

średni kolor i jego rozkład na powierzchni obrazu,

histogramy składowych koloru dla zadanej ilości słupków

Są to cechy niezależne od rozdzielczości i kąta pochylenia obrazu

Wyszukiwanie obrazów z treści

- Dla takich cech nie ma konieczności segmentacji obiektów i tła obrazu
- **Histogram koloru** c_i w obrazie I :

$$h_I(c_i) = P(\text{color}(p) = c_i | p \in I)$$

gdzie $P(\text{color}(p) = c_i | p \in I)$ – prawdopodobieństwo, że piksel p losowo wybrany z obrazu I posiada kolor c_i .

- **Autokorelogram** uzupełnienia histogram informacją o rozłożeniu koloru w polu obrazu

$$h_I(c_i, c_j, r) = P(\text{color}(p_1) = c_i \wedge \text{color}(p_2) = c_j | r = d(p_1 - p_2))$$

- gdzie p_1, p_2 – piksele losowo wybrane z obrazu I ,
 - $d(p_1 - p_2)$ – odległość pikseli p_1, p_2 na obrazie.
-



Miary teksturalne

Problem:

człowiek rozpoznaje kolory obiektów niezależnie od padającego na nie światła – np. jabłko rozpoznaje się jako czerwone w świetle dziennym i przy oświetleniu sztucznym; histogramy kolorów nie spełniają tego warunku

Tekstura – obszar o powtarzającym się wzorze kolorów łatwo rozpoznawalny przez człowieka

Miary tekstury nie powinny zależeć od jasności ani orientacji obrazu

Macierz kookurencji poziomów szarości (GLCM) – rejestruje zmiany jasności pomiędzy parami pikseli w obrazie; zapisuje w ten sposób informację o teksturze jasności obrazu

Miary teksturalne

- Przy obliczaniu macierzy GLCM rozważa się wszystkie v-związane pary pikseli p_1, p_2 odległe o wektor $\vec{v} = \vec{p}_2 - \vec{p}_1$
- Wyznacza się:

$$P_I(c_i, c_j, \vec{v}) = P(\text{color}(p_1) = c_i, \text{color}(p_2) = c_j | \vec{p}_2 - \vec{p}_1 = \vec{v})$$

gdzie $P_I(c_i, c_j, \vec{v})$ – prawdopodobieństwo znalezienia v-związanych par pikseli związanych z kolorami c_i, c_j w obrazie I .

- Globalne wskaźniki tekstury obliczane na podstawie GLCM: **energia, entropia, kontrast, jednorodność.**
-

Miary teksturalne

- **Energia** – miara jasności v-związanych pikseli

$$\mathcal{E}_I(c_i, c_j, \vec{v}) = \sum_i \sum_j P_I(c_i, c_j, \vec{v})^2$$

- **Entropia** – miara niejednorodności v-związanych pikseli

$$\Psi_I(c_i, c_j, \vec{v}) = \sum_i \sum_j P_I(c_i, c_j, \vec{v}) \log P_I(c_i, c_j, \vec{v})$$

- **Kontrast** – miara różnic jasności ϕ_i dla v-związanych par pikseli

$$\mathcal{C}_I(c_i, c_j, \vec{v}) = \sum_i \sum_j (\phi_i - \phi_j)^2 P_I(c_i, c_j, \vec{v})$$

Miary teksturalne

Jednorodność – miara podobieństwa v-związanych pikseli

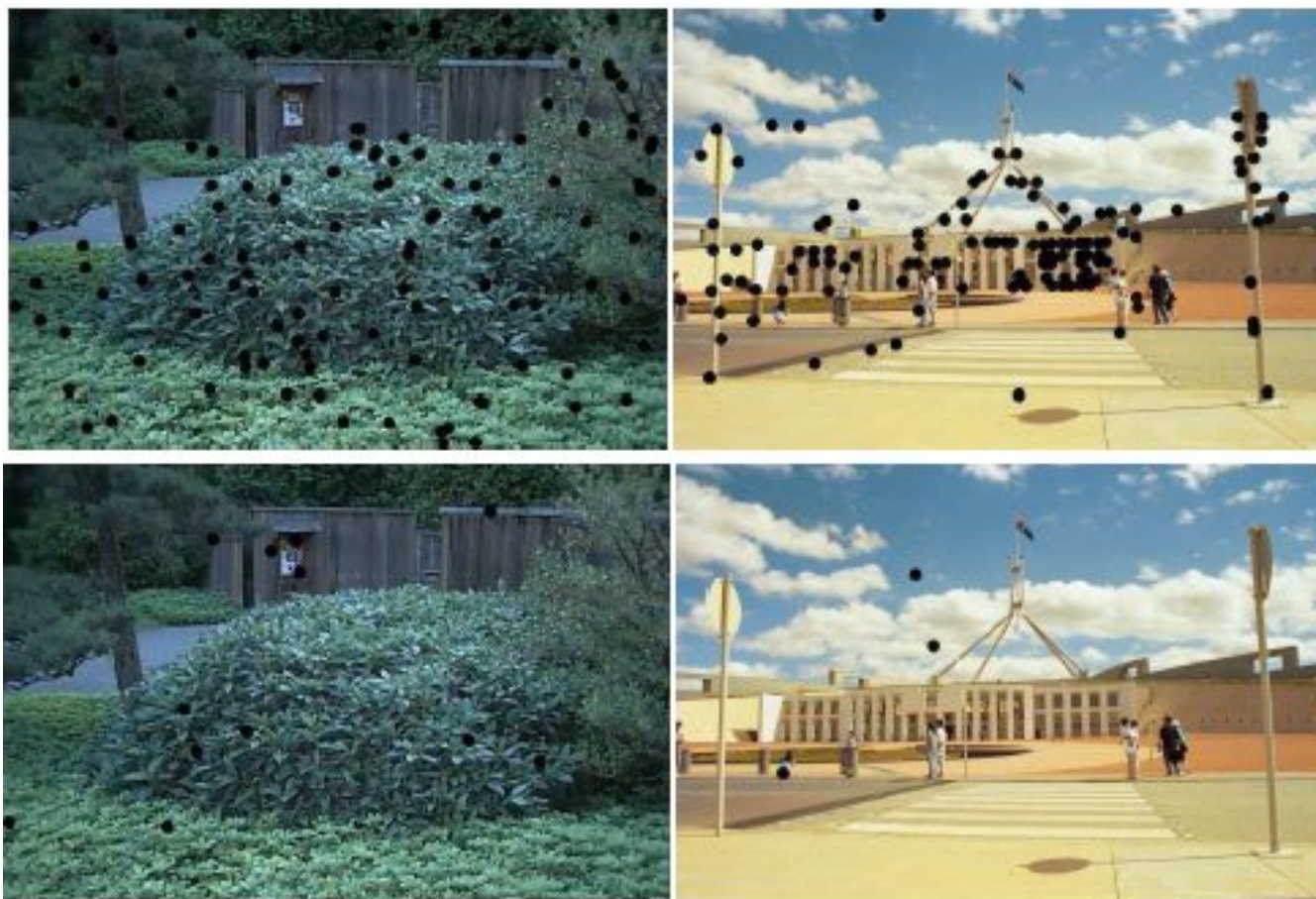
$$\mathcal{H}_I(c_i, c_j, \vec{v}) = \sum_i \sum_j \frac{P_I(c_i, c_j, \vec{v})}{1 + |\phi_i - \phi_j|}$$

Inteligentniejsze metody oparte są o modele cech obrazu – np. kombinacje koloru i częstotliwości przestrzennych wybranych regionów

Metoda punktów (regionów) odniesienia (*sailent points*):
zbieranie cech obrazu niezależnych od skali, oświetlenia, pozycji kamery, rotacji obiektów; wykorzystuje się punkty odniesienia, unifikację orientacji i lokalną geometrię dla tekstury

Wyszukiwanie obrazów z treści

Punkty odniesienia są związane z rogami obrazu lub jego szczególnymi obszarami





Wyszukiwanie obrazów z treści

Podobieństwo obrazów oblicza się za pomocą sumarycznych statystyk punktów (obszarów) odniesienia

W regionach odniesienia obrazy są specjalnie filtrowane

Wartości charakterystyk teksturalnych są klastrowane metodą k-średnich aby określić słowa języka

Do dopasowania tak przetworzonych obrazów używa się algorytmu pLSA (*probabilistic Latent semantic analysis*)
