W poszukiwaniu sensu w świecie widzialnym

Andrzej Śluzek

Nanyang Technological University
Singapore

Uniwersytet Mikołaja Kopernika
Toruń

Podziękowania

Przedstawione wyniki powstały prawie w całości w ramach grantu 072 134 0052 **"Framework for Visual Information Retrieval and Building Content-based Visual Search Engines"** przyznanego przez A*STAR Science & Engineering Research Council (Singapore).

Członkowie zespołu:

Dr Mariusz Paradowski

Dr Yang Duanduan

Dr Md Saiful Islam

Ms Elahe Farahzadeh

Wstęp

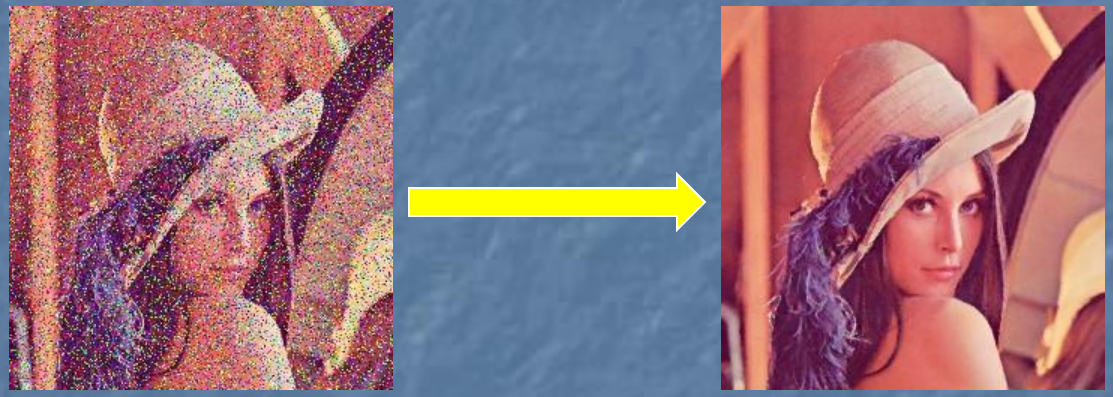
- O czym nie będę (i będę) mówił?
- Podobieństwo wizualne
  - *Cechy globalne*
  - *Cechy lokalne (punkty kluczowe)*
- Wykrywanie fragmentów podobnych
  - *Podejście geometryczne*
  - *Podejście topologiczne*
  - *Implementacja w czasie rzeczywistym*
- Automatyczne tworzenie pojęć „obiektów”
  - *Klasteryzacja podobnych fragmentów*
  - *Precyzyjna segmentacja „obiektów”*
- Co dalej?

O czym nie będę ~~(i będę)~~ mówił:

Pobieranie informacji wizualnej ze świata zewnętrznego jest procesem w pełni zautomatyzowanym, który nie wymaga żadnej wiedzy o otaczającym świecie ani szczególnej „inteligencji”.

O czym nie będę ~~(i będę)~~ mówił:

Przetwarzanie (np. poprawa jakości) informacji wizualnej też oparte jest na w pełni zalgorytmizowanych technikach.



O czym nie będę ~~(i będę)~~ mówił:

Rozumienie treści obrazów wymaga (znacznej) wiedzy „pozawizualnej”.

O czym ~~nie będę~~ (i będę) mówił:

Próba zdefiniowania „rozumienia” obrazów na podstawie ich formy wizualnej. Żadna wiedza na temat (fizycznej) treści i znaczenia obrazów nie jest zakładana.

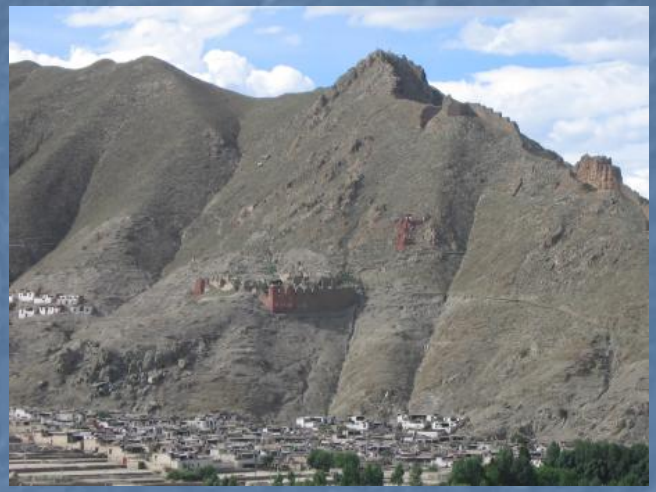
Dwa podstawowe podejścia:

- Ja już coś takiego widziałem, choć nie wiem, co to jest.
- Ja już to widziałem i potrafię to nazwać (sklasyfikować).

Jedynym założeniem jest istnienie w jakiś sposób zdefiniowanego **podobieństwa wizualnego**.

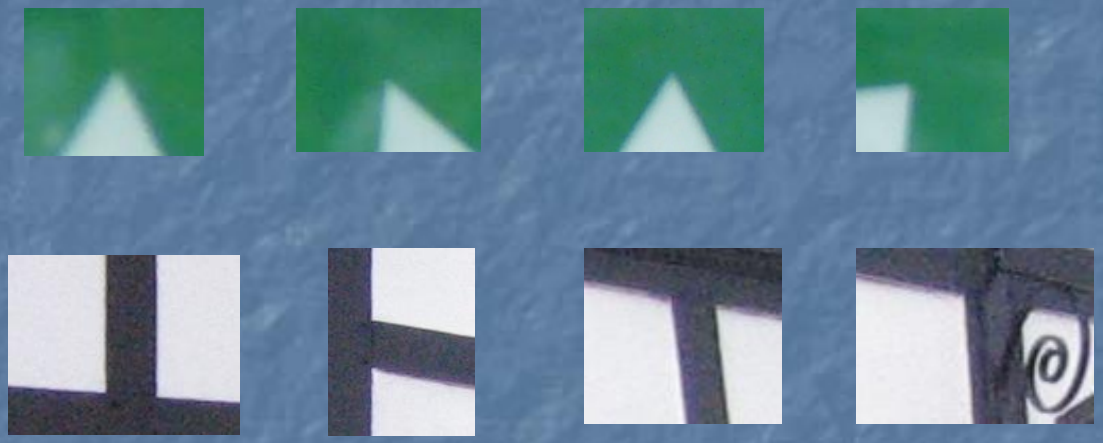
Podobieństwo wizualne

Podejście globalne/semi-globalne



Podobieństwo wizualne

Podejście lokalne

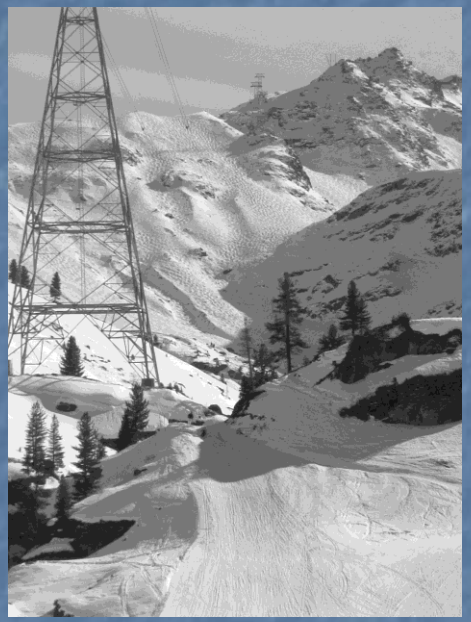


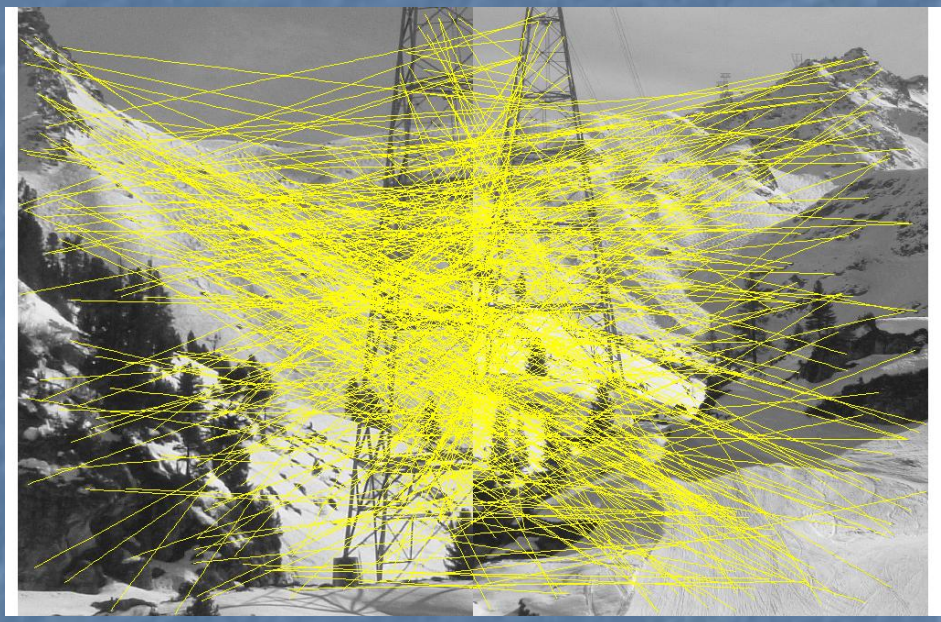
Problem:

Podobieństwo lokalne (jak je zdefiniować?) rzadko przekłada się na rzeczywiste podobieństwo między większymi (autentycznie podobnymi) fragmentami obrazów lub całymi obrazami.

Podobieństwo wizualne

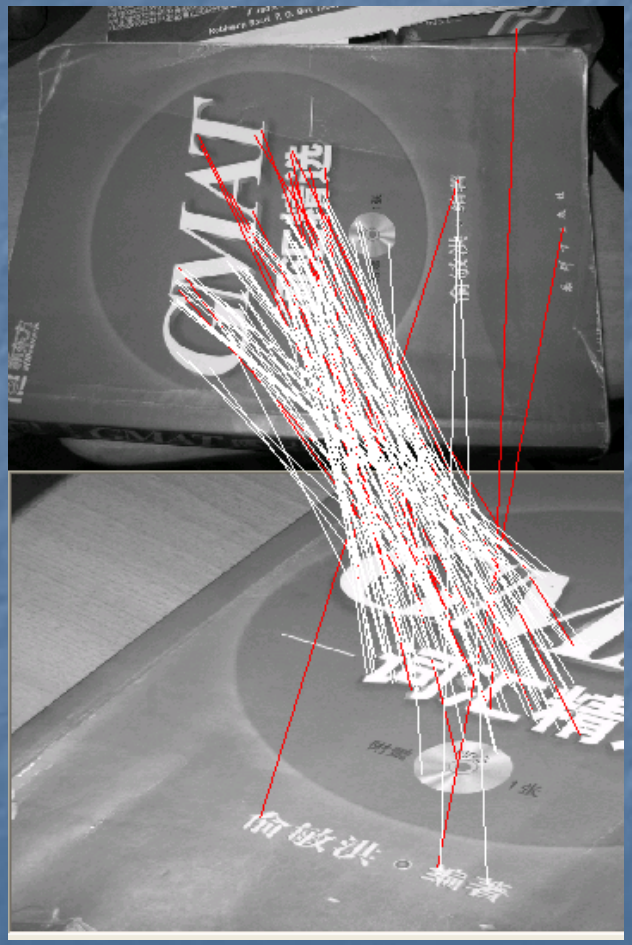
Podejście lokalne

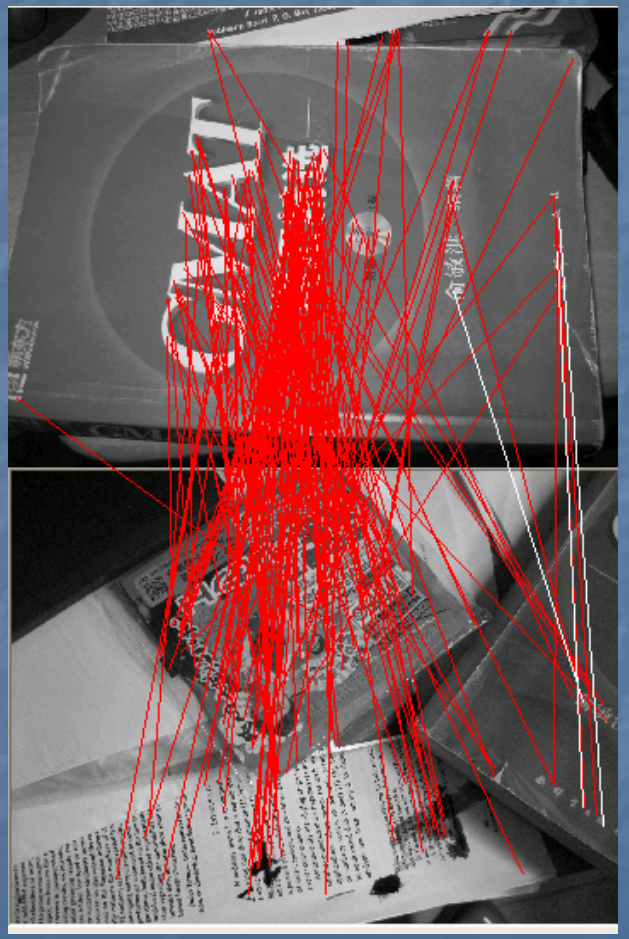




Podobieństwo wizualne

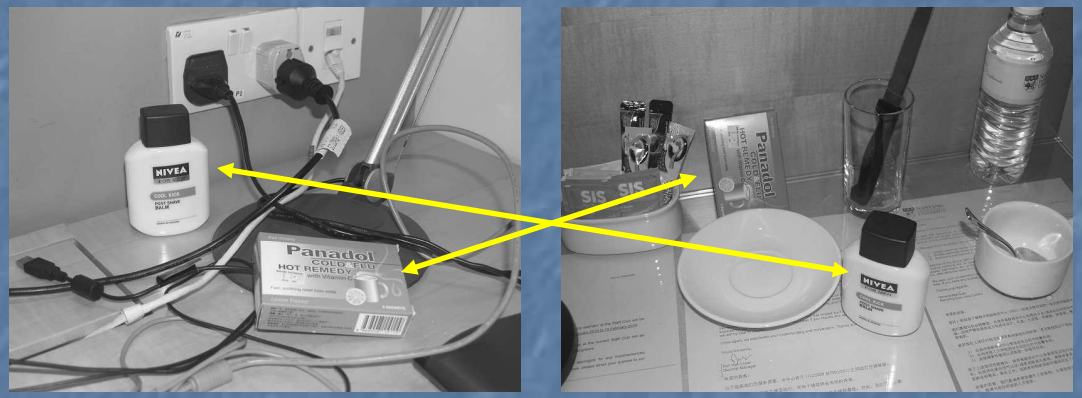
Podejście lokalne





Zdefiniowanie problemu (z wykorzystaniem podobieństwa lokalnego).

Mając parę przypadkowych obrazów, poszukujemy w nich (być może istniejących, a być może nie) *prawie-duplikatów*.



Zadanie polega nie tylko na znalezieniu takich podobnych fragmentów, ale również na (w miarę dokładnym) określeniu kształtów tych *prawie-duplikatów*.

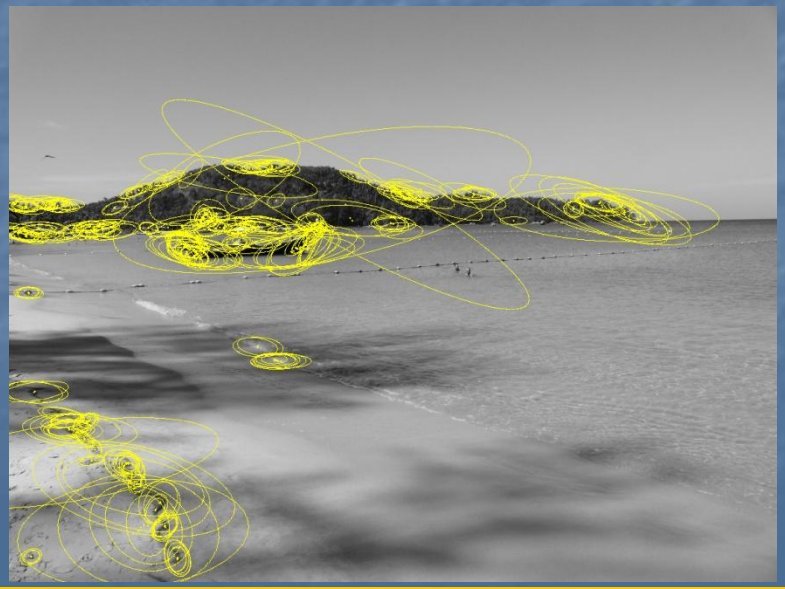
Czysto wizualna definicja podobieństwa lokalnego

Punkty kluczowe definiują fragmenty obrazu (punkty wraz z pewnym otoczeniem) o tak unikalnych własnościach wizualnych, że jeśli ta sam scena przedstawiona będzie na innym obrazie, to większość punktów kluczowych będzie ponownie wykryta. Punkty kluczowe opisywane są *n*-wymiarowymi wektorami w pewnej przestrzeni deskryptorów. W ten sposób lokalne podobieństwo między dwoma obrazami są zdefiniowane jako podobieństwo między parą punktów kluczowych.

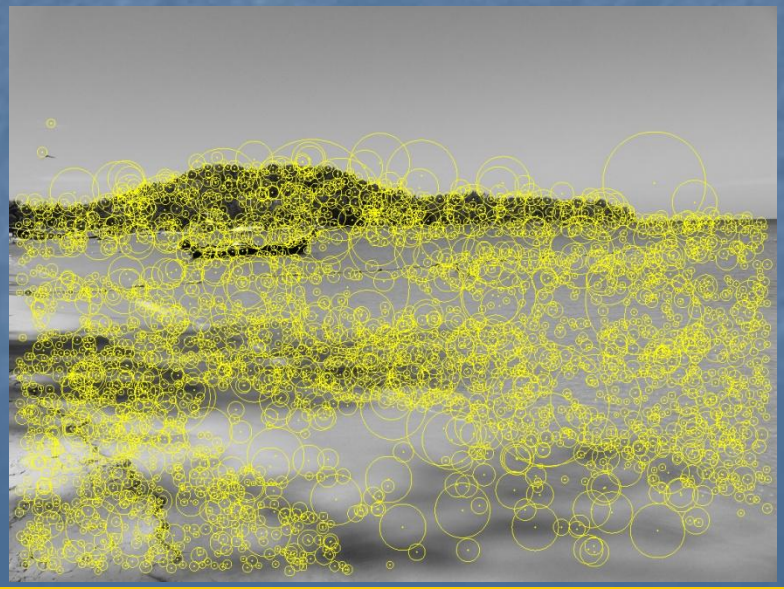
Istnieją różne **detektory** punktów kluczowych (np. DoG, Harris-Affine, Hessian-Affine, MSER, SURF, itd.) oraz różne **deskryptory** wykrytych punktów (np. SIFT, SURF, C2I, itd.).

Przykłady detekcji punktów kluczowych

Harris-Affine



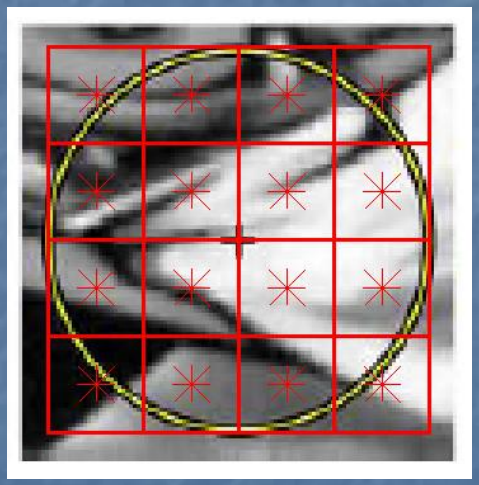
SURF



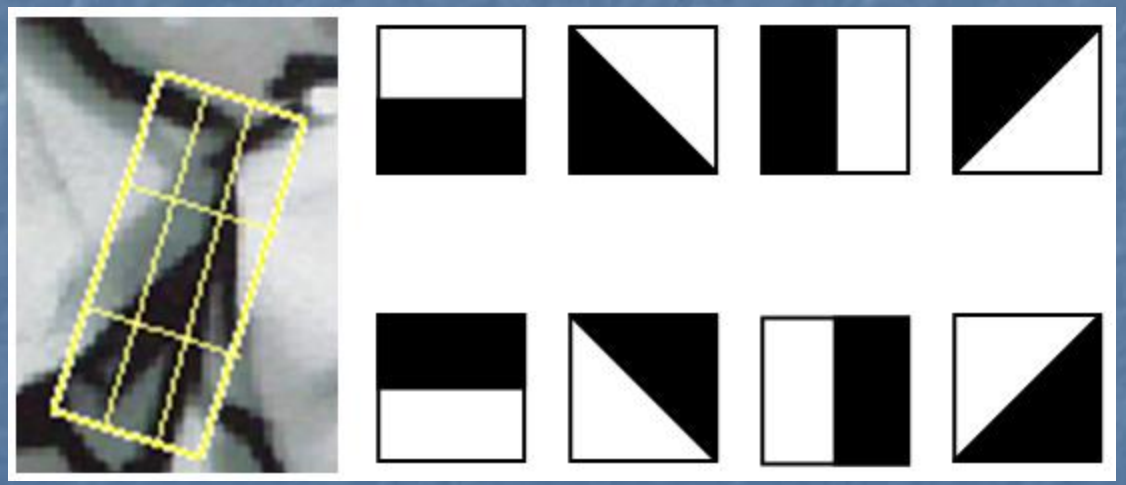
Detekcja i opis punktów kluczowych

Rozmiar (i kształt) kołowego (eliptycznego) obszaru wokół punktu kluczowego definiuje optymalną lokalną skalę obrazu (oraz lokalną anizotropię funkcji jasności).

Wykryte punkty kluczowe (zwykle po normalizacji do okręgu) opisywane są różnymi deskryptorami.

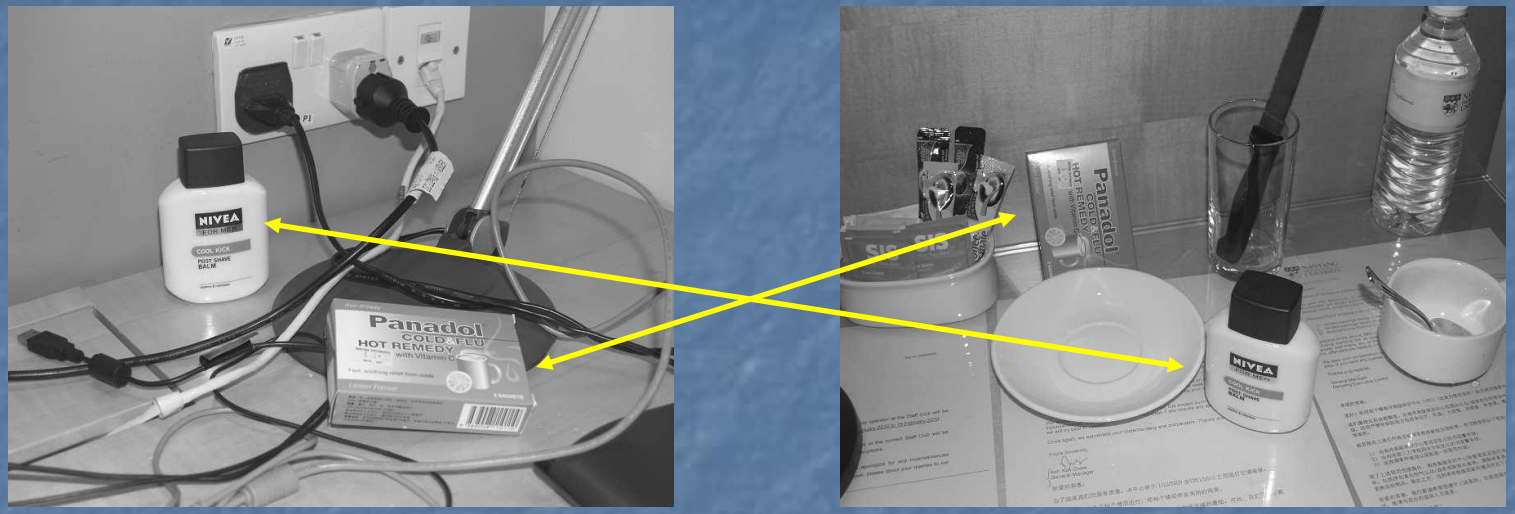


SIFT (128-wymiarowy)

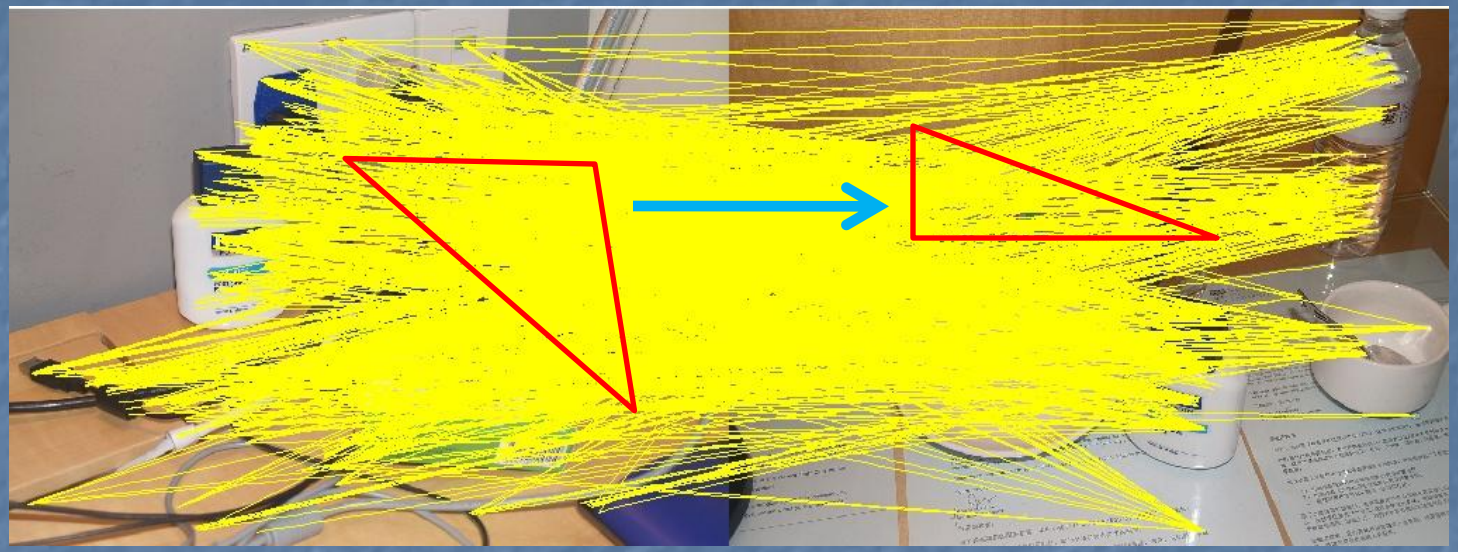


C2I (36-wymiarowy)

Wykrywanie prawie-duplikatów przez porównywanie punktów kluczowych



Wykrywanie prawie-duplikatów przez porównywanie punktów kluczowych



Geometria **afiniczna**

Dowolna para trójkątów zbudowanych z podobnych punktów kluczowych (z dwóch obrazów) wyznacza odwzorowanie afiniczne między tymi dwoma obrazami.

Wykrywanie prawie-duplikatów przed porównywanie punktów kluczowych

1. Detekcja *punktów kluczowych* w obu obrazach;
2. Znalezienie *par podobnych punktów kluczowych*;
3. Budowa *trójkątów* z par podobnych punktów kluczowych;
4. Wyznaczenie *transformacji afinicznej* dla *par trójkątów*;
5. *Dekompozycja* transformacji afinicznych;
6. Budowa 6-wymiarowego *histogramu* parametrów transformacji afinicznych;
7. Detekcja *lokalnych maksymów* histogramu;
8. Maksyma histogramu definiują parę *prawie-duplikatów*.

Dekompozycja odwzorowań afinicznych

SVD-dekompozycja

$$\mathrm{A} = \begin{bmatrix} A & B & C \\ D & E & F \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} K & P \\ 0^T & 1 \end{bmatrix} \qquad K = \begin{bmatrix} A & B \\ D & E \end{bmatrix} \qquad P = \begin{bmatrix} p_x \\ p_y \end{bmatrix} = \begin{bmatrix} C \\ F \end{bmatrix}$$

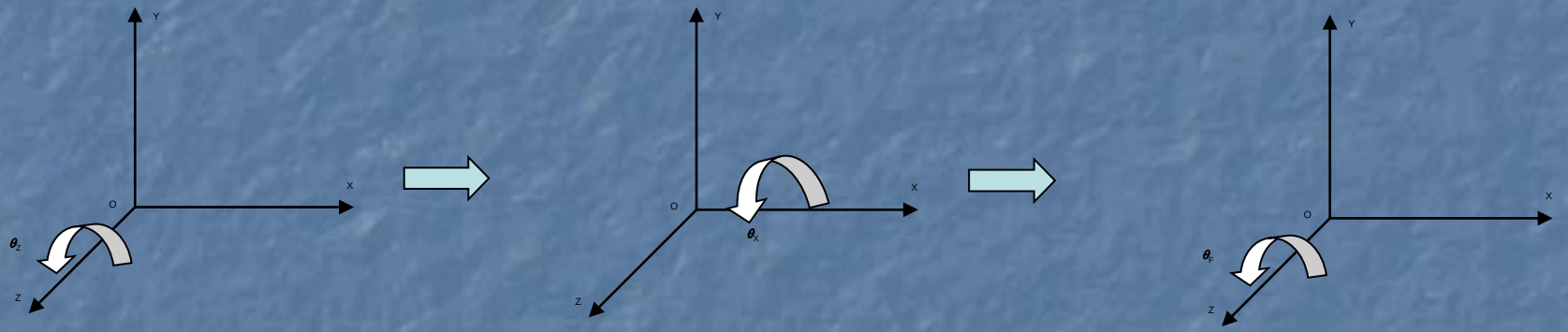
gdzie

$$K = Rot(\gamma) \cdot Rot(-\theta) \cdot S \cdot N \cdot Rot(\theta)$$

$$S = \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix} \qquad N = \begin{bmatrix} 1 & 0 \\ 0 & \pm 1 \end{bmatrix}$$

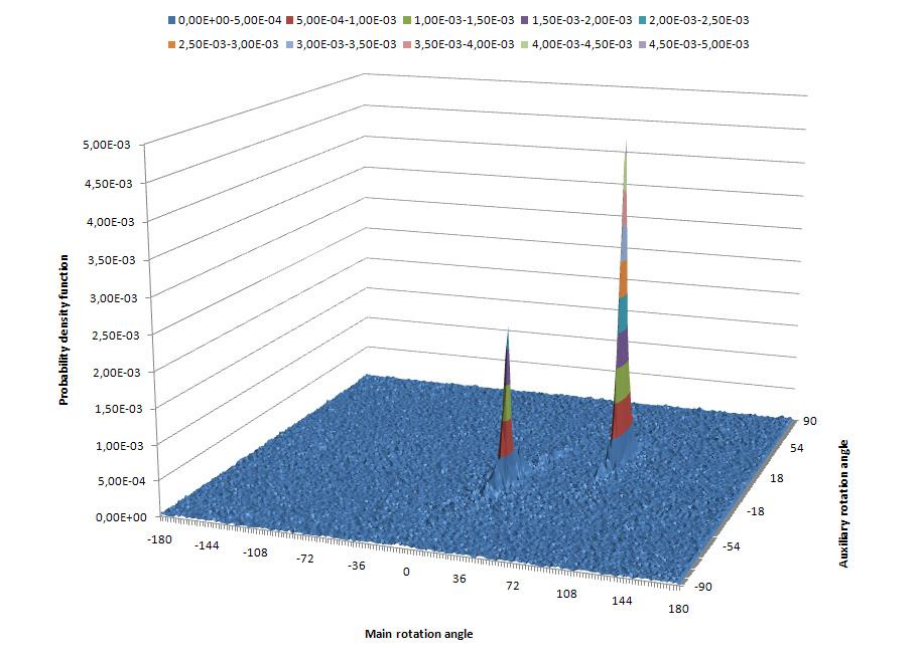
Dekompozycja odwzorowań afinicznych

3D dekompozycja



$$HT = \begin{bmatrix} c_F c_Z - s_F c_X s_Z & -c_F s_Z - s_F c_X c_Z & s_F s_X & p_X \\ s_F c_Z + c_F c_X s_Z & -s_F s_Z + c_F c_X c_Z & -c_F s_X & p_Y \\ -s_X s_Z & s_X c_Z & c_X & p_Z \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$PR = \begin{bmatrix} K & 0 & 0 & 0 \\ 0 & K & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$



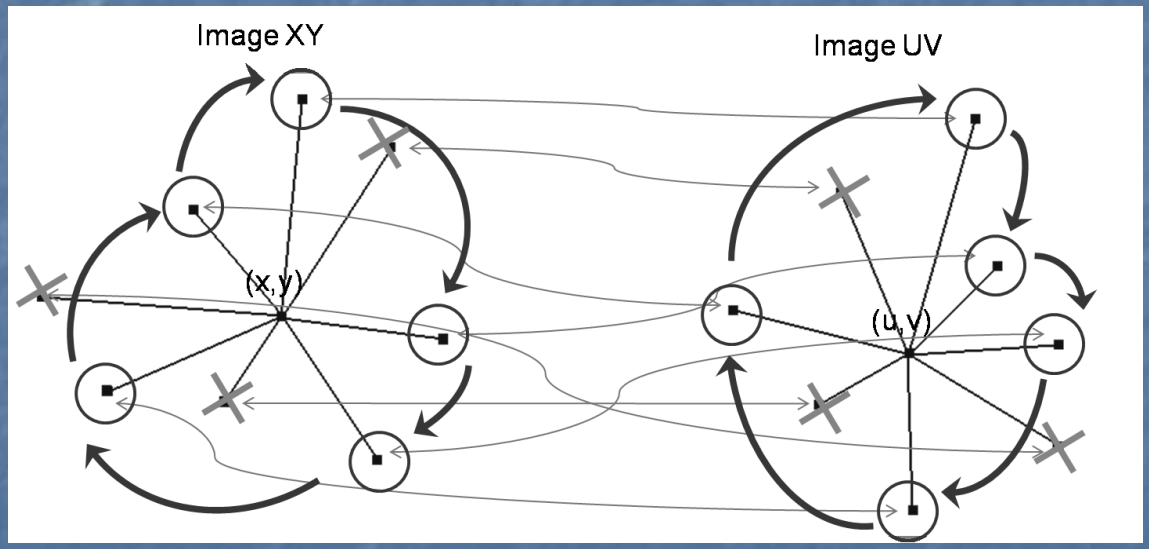
Histogram (2D podprzestrzeń) parametrów afinicznych dla danej pary obrazów.

Przykładowe wyniki

Przykładowe wyniki

Podejście topologiczne

Wokół każdej pary podobnych punktów kluczowych tworzymy wektory do innych (odpowiednio podobnych) punktów kluczowych. Pary, wokół których wystarczająca liczba wektorów zachowuje uporządkowanie orientacji uznawane są za należące do prawie-duplikatów (wraz z sąsiednimi parami, dla których zachowana jest orientacja).
Dalszy etap to zwykła analiza spojności grafów.



Przykładowe wyniki

Przykładowe wyniki

Afinicznie

Topologicznie

Afinicznie

Topologicznie

Zastosowanie

Mając bazę danych wzorcowych obrazów, można stwierdzić, czy obraz bieżący (zapytanie) zawiera fragmenty podobne do ***czegoś już znanego***.

Przy pewnych założeniach upraszczających złożoność obliczeniową, system może działać w czasie rzeczywistym (bieżące obrazy są kolejnymi klatkami sygnału wideo) jeśli baza danych zawiera kilka obrazów.

Real-time Multiple Object Matching, ver. 1.0

Input signal setup

Capture source: Camera capture 640x480 Video file:

Start input signal | Stop input signal | Record input signal

Matching image setup

Capture from signal | Load from image file | Clear image | Matching setup

Image file: Captured from signal

Input signal and matching

Image to match

Exit | About...

Ostatnio powstał system wstępnego wyszukiwania, który umożliwia znalezienie (w czasie rzeczywistym) wśród obrazów z bazy kandydatów potencjalnie zawierających podobne fragmenty (bez wskazywania tych fragmentów). Baza danych może liczyć tysiące obrazów.

zapytanie

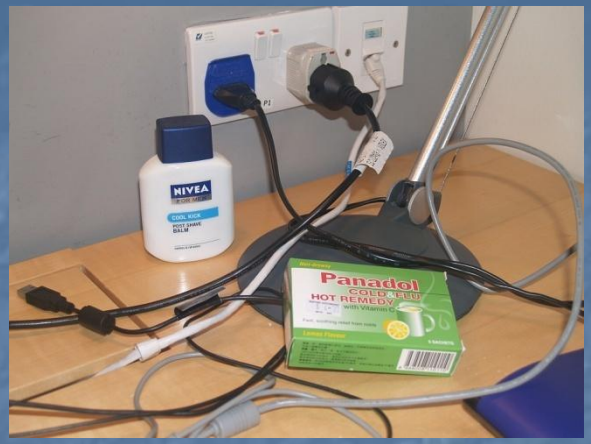
wstępne wyszukiwanie

zapytanie

wstępne wyszukiwanie

Automatyczne tworzenie pojęć „obiektów”

1. Liczba obrazów w bazie danych nie może rosnąć w nieskończoność (nawet wstępne wyszukiwanie może stać się zbyt kosztowne).
2. Klastry wzajemnie podobnych prawie-duplikatów uznawane są za definicje „obiektów” (zazwyczaj odpowiadają one fizycznym obiektom z eksplorowanego środowiska).


NIVEA
Panadol
COLD + FLU
HOT REMEDY


SIS
SIS
Panadol
NIVEA


NIVEA
Panadol
Dlaczego kapitalizm
zawsze zwycięża


JVC
NIVEA

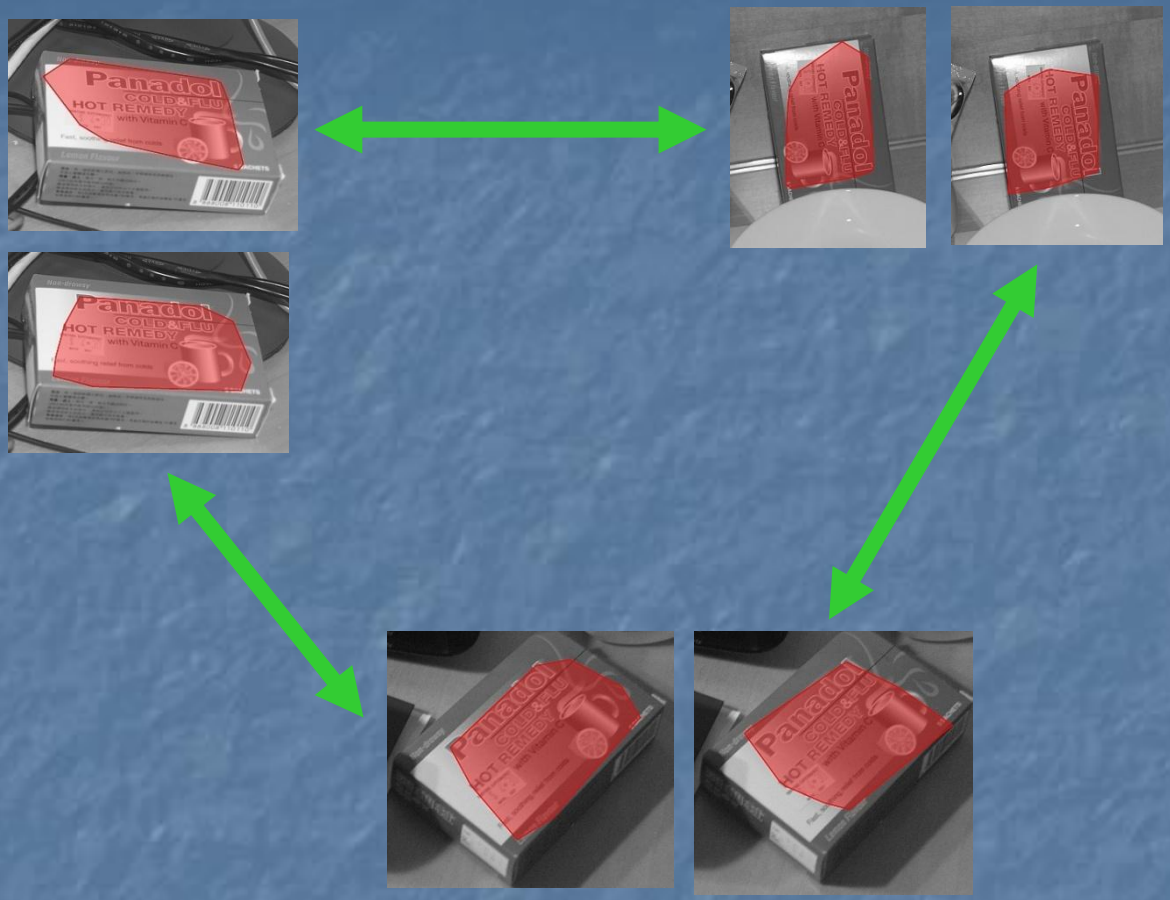

22:52
SONY
NIVEA
Panadol


NIVEA







Zaczątek nowej „klasy obiektu” (*n* wzajemnie podobnych prawie-duplikatów).

Kolejne wzorce danego „obiektu” mogą być dodawane jeśli:

- Są prawie-duplikatami co najmniej *n* już istniejących wzorców w tej klasie (wystarczające podobieństwo do „klasy obiektów”).
- Są prawie-duplikatami nie więcej niż *m* wzorców w tej klasie (unikanie nadmiernej redundancji w gromadzonej informacji wizualnej).

Klasy ”obiektów” (klastry prawie-duplikatów) tworzone są w pełni automatycznie, wyłącznie na podstawie wizualnej treści obrazów. Użytkownik może nadawać tym klasom „obiektów” semantycznie znaczące nazwy.

W przypadku penetrowania w miarę ograniczonego środowiska może się w pewnym momencie zdarzyć, że (nieomal) wszystkie prawie-duplikaty znajdowane w kolejnych pobieranych obrazach są prawie-duplikatami wzorców z powstałych klas „obiektów”. Wówczas bazę danych obrazów można w pełni zastąpić zestawem wzorców automatycznie stworzonych klas „obiektów”.

Przykładowe wyniki działania zrealizowanego systemu.

Klasa „obiektów”

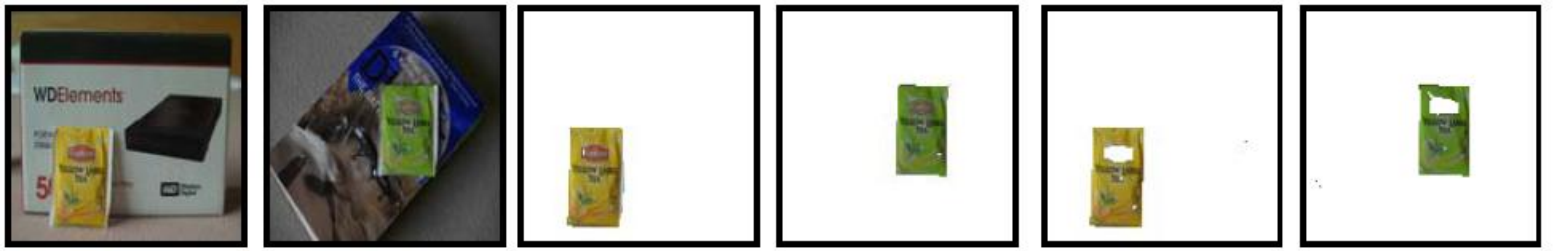
Przykładowe wzorce wybranych klas „obiektów”.

Precyzyjna segmentacja „obiektów”

Dokładne kształty poszczególnych wzorców klas „obiektów” można uzyskać metodami ko-segmentacji.

Zaproponowana metoda przewyższa algorytm uznany obecnie za najlepszy.

Precyzyjna segmentacja „obiektów”



Input pair | Our cosegmentation, err: 0.6% | Hochbaum et al.’s cosegmentation, err: 1.2%

Input pair | Our cosegmentation, err: 3.36% | Hochbaum et al.’s cosegmentation, err: 23.7%

Input pair | Our cosegmentation, err: 3.5% | Hochbaum et al.’s cosegmentation, err: 22.2%

Co dalej?

1. Doskonalenie wydajności (czasowej i pamięciowej) istniejących algorytmów.
2. Próby zasklepienia „semantycznej szczeliny” pomiędzy metodami ręcznej anotacji obrazu i metodami anotacji w pełni automatycznej (w jaki sposób porównywać kategorie semantyczne z automatycznie tworzonymi kategoriami „wizualnymi”?).

Dziękuję za uwagę!