AKADEMIA GÓRNICZO-HUTNICZA
IM. STANISŁAWA STASZICA W KRAKOWIE

Segmentacja obiektów pierwszoplanowych

Tomasz Kryjak
Laboratorium Systemów Wizyjnych,
Katedra Automatyki i Robotyki,
AGH w Krakowie

Wstęp

Odejmowanie tła – jedna z najbardziej rozpowszechnionych metod wykrywania obiektów pierwszoplanowych (przy założeniu, że obraz rejestrowany jest statyczną kamerą – a przynajmniej trajektoria ruchu kamery jest znana – przypadek PTZ).

Prosta wersja to stałe tło – (np. puste skrzyżowanie, pusty korytarz, jednorodne tło stosowane w TV). Problem to podatność na szum oraz brak mechanizmu kompensacji zmian na scenie, przykładowo oświetlania, przesunięcia krzesła itp.

Wstęp cd.

Idea generacji tła wywodzi się wprost z metody ze stałym tłem.

Na wstępnie tworzony jest **model tła** i następnie w trakcie działania algorytmu model ten jest na bieżąco aktualizowany.

Najprostsze rozwiązanie to „podmienianie” modelu tła: np. w przypadku nagłej zmiany oświetlenia lub co określony czas.

W tym przypadku nie ma gwarancji, że w nowym modelu tła nie znajdą się obiekty ruchome (np. samochody na skrzyżowaniu).

Dlatego też, stosuje się bardziej złożone metody …

Przykład

SCENA

TŁO

RÓŻNICA

MASKA

Wyzwania związane z generacją tła

- szum na obrazie
- potencjalne drżenie kamery
- niewrażliwość na automatyczną zmianę nastaw kamery (np. balans bieli)
- ***pora dnia (w ogólności płynne zmiany oświetlenia)***
- ***nagłe zmiany oświetlenia (np. słoneczno - pochmurny dzień, włączenie światła w biurze)***
- ***obecność obiektów ruchomych w sekwencji inicjalizacyjnej (bootstrap)***
- ***tło multimodalne (ruszające się liście, fontanna)***
- kamuflaż – obiekt ruchomy podobny do tła
- cienie
- poruszone obiekty w tle (np. przesunięte krzesło)
- obiekty wstawione w tło (wtapianie się obiektów)
- obiekty początkowo nieruchome zaczynają się poruszać
- „śpiące” obiekty

Problem

W pracy (Shireen, 2008) można znaleźć następujące twierdzenie:

Istnieją metody, które są w stanie poradzić sobie z wymienionymi problemami (lepiej lub gorzej) ale są one złożone obliczeniowo.

Stoi to zazwyczaj w sprzeczności w wymaganiem działania w czasie rzeczywistym (realny system nadzoru wizyjnego).

Czym nie powinien być moduł generacji tła

Generacja tła nigdy nie jest jedynym i ostatnim elementem systemu nadzoru wizyjnego.

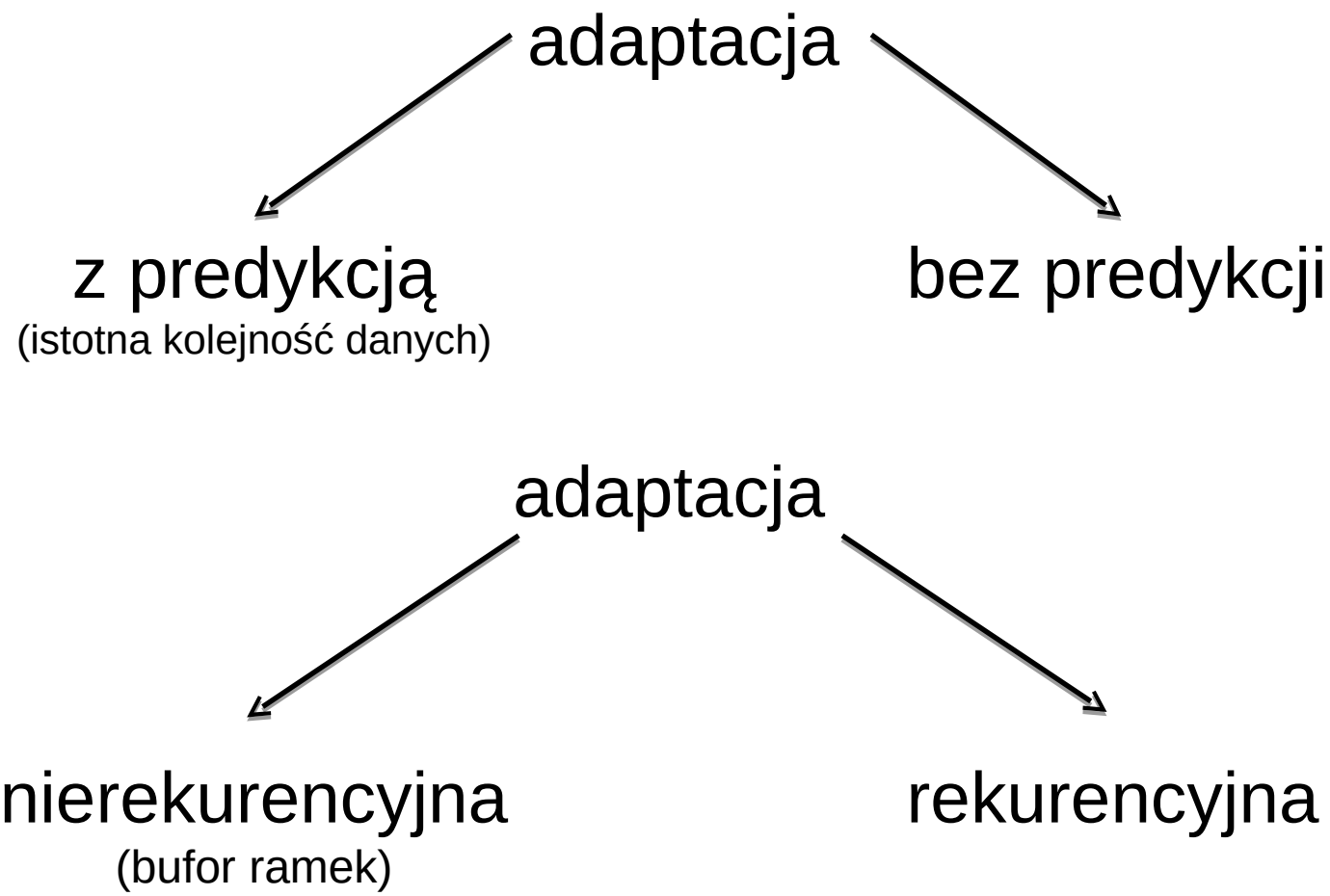
Zatem nie powinno się oczekiwać, że na poziomie generacji tła realizowany będzie opis semantyczny sceny (czyli rozwiązywany problem np. śpiącej osoby).

Z drugiej strony konieczna jest ścisła współpraca „wyższych” modułów z generacją tła – selektywna aktualizacja.

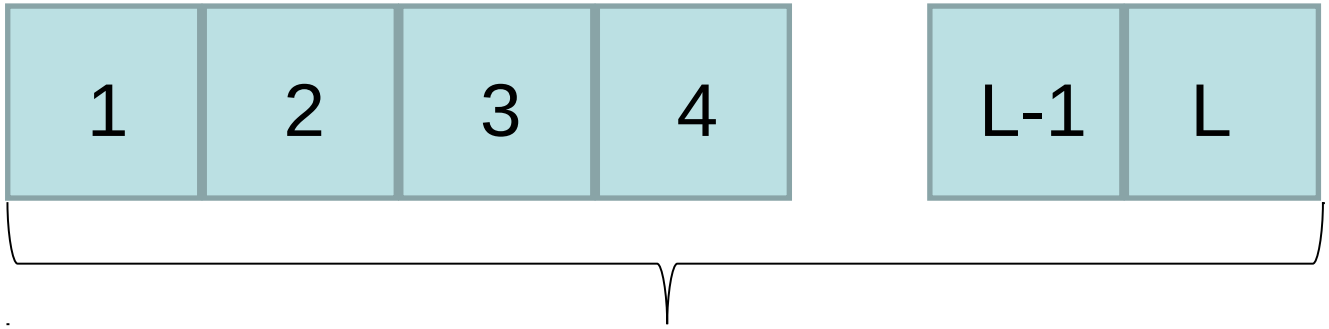
Model tła - reprezentacja

- stałe tło
- pojedynczy rozkład Gaussa
- kilka rozkładów Gaussa
- KDE – estymacja nieparametryczna
- model oparty o przepływ optyczny
- model oparty o filtrację Winera
- HMM – ukryte modele Markova
- model oparty o klastry lub książkę kodową (codebook)
- model oparty o bufor próbek (uporządkowany lub losowy)

Model tła - adaptacja



Techniki nierekurencyjne



- średnia (również wersja ważona)
- mediana, maksimum,
- predykcja liniowa (Wiener)
- KDE

Średnia z bufora

- wymaga zapamiętania L ramek
- obliczenia wykonywane są tylko na elemencie nowym i ostatnim (na zasadzie bufora cyklicznego) -> niska złożoność obliczeniowa metody

$$S_{t+1} = S_t + \frac{I_{t+1}}{N} - \frac{B_{t-N}}{N}$$

Mediana, maksimum

- tworzony jest histogram wartości pikseli
- wartość mediany i maksimum wyznaczana jest na podstawie histogramu
- dla koloru, w celu uproszczenia obliczeń, zakłada się niezależność składowych - co dla przestrzeni RGB jest założeniem błędnym

Techniki nierekurencyjne

Zalety:

- wysoka adaptacyjność,
- brak wypływu na model tła historii sprzed L ramek

Wady:

- złożoność pamięciowa,
- dla wolno poruszających się obiektów, konieczny duży bufor,
- mała odporność na nagłe zmiany oświetlenia

Techniki rekurencyjne (1)

Jednowariantowe:

- występuje jeden wariant tła,
- aktualizacja:
  - aproksymacja średniej (*Single Gaussian*):

$$\mu_t = \alpha I_t + \left(1-\alpha\right)\cdot \mu_{t-1}$$

  - aproksymacja mediany (sigma-delta)

$$M_t = M_{t-1} + \operatorname{sgn}\left(I_t - M_{t-1}\right)$$

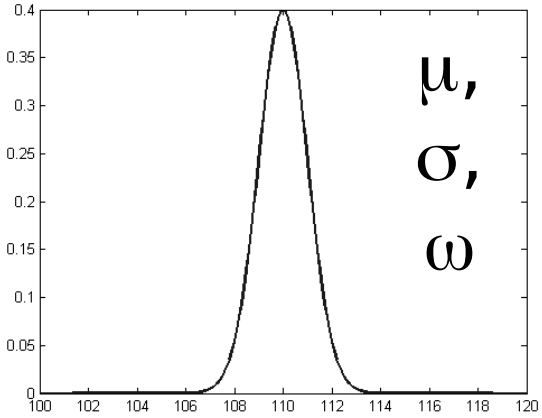
Techniki rekurencyjne (2)

Wielowariantowe:

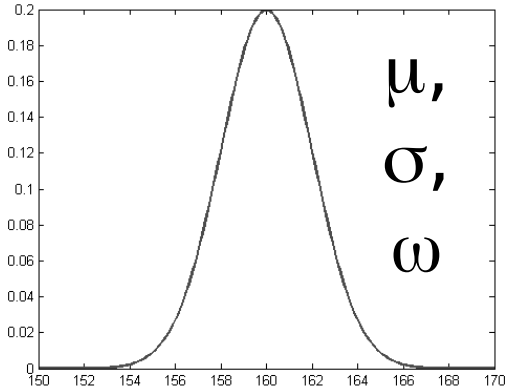
- istnieje więcej niż 1 wariant tła, najczęściej od K=3-5 (model zawsze tylko jeden)
- pierwsza w literaturze propozycja to MOG (Stauffer, Grimson, 1999)
- pojedynczy wariant składa się z:
  - średniej μ,
  - odchylenia standardowego σ,
  - wagi ω.

MOG (1)

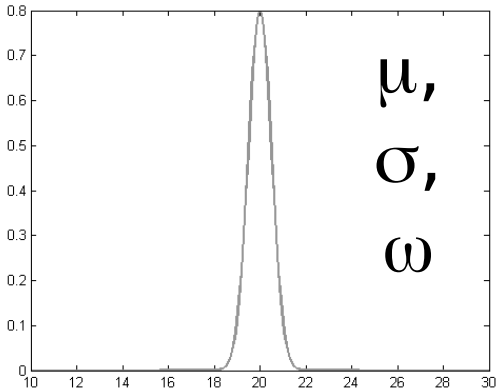
1. Dla nowego piksela liczona jest odległość od każdego z wariantów – zwykle Euklidesowa
2. Warunek przynależności: $||x-\mu|| < 2.5\ \sigma$



K=1



K=2



K=3

MOG (2)

Jeżeli znaleziony zostanie „pasujący” wariant - aktualizacja:

$$\mu_t = (1-\rho)\mu_{t-1} + \rho X_t$$

$$\sigma^{2}{}_t = (1-\rho)\sigma^{2}{}_{t-1} + \rho (X_t - \mu_t)^T (X_t - \mu_t)$$

$$\rho = \alpha\eta(X_t | \mu_k, \sigma_k)$$

$$\omega_{i,t} = (1-\alpha)\omega_{i,t-1} + \alpha$$

Dla pozostałych wariantów redukcja wagi:

$$\omega_{i,t} = (1-\alpha)\omega_{i,t-1}$$

MOG (3)

Brak dopasowanego wariantu:

- ostatni (o najmniejszym współczynniku ω/σ) jest usuwany i jego miejsce tworzony nowy:
  - mała waga ω
  - duże wstępne odchylenie σ
  - średnia taka jak wartość piksela

MOG (4)

Detekcja obiektów:

- warianty, w każdej iteracji, są sortowane względem ω/σ (stacjonarne tło występuje częściej i ma małą wariancję)
- za należące do tła uznaje się warianty spełniające warunek:

$$B = \operatorname{argmin}_b \left(\sum_{k=1}^{b} \omega_k > T \right)$$

- pozostałe to obiekty ruchome

MOG (5)

Uwagi do metody (literatura):

- rozkład pikseli tła nie zawsze ma charakter rozkładu Gaussa,
- wyniki przeprowadzonych badań wskazują na to, że algorytm pozwala na poprawne śledzenie tylko średniej, estymowanie odchylenia standardowego i wag jest niestabilne,
- kolejnym słabym punktem jest metoda klasyfikacji obiekt/tło i wynikająca z niej metoda aktualizacji tła: uaktualniania jest tylko ta cześć rozkładu, która położona jest w okolicach średniej (spełnienie warunku na odchylenie standardowe). W innym przypadku tworzony jest nowy rozkład, co skutkuje niedoszacowaniem wartości odchylenia standardowego. Ponadto taki rozkład klasyfikowany jest błędnie jako obiekt,
- wybór przestrzeni RGB oraz poczynione założenie o niezależności składowych jest nieprawdziwe.

Metoda szeroko badana – 200 artykułów.

Selektywna aktualizacja tła

- przenikanie (niekontrolowane) obiektów ruchomych do tła to przyczyna większości problemów – oczywiście są sytuacje w których takie zachowanie może być pożądane (np. „przesunięte krzesło”).
- sytuacja idealna: każdy wykryty obiekt (a raczej piksele, które wchodzą w jego skład) jest klasyfikowany jako obiekt ruchomy lub tło i na tej podstawie następuje aktualizacja modelu tła – w praktyce raczej nierealne ...

Kryterium selektywnej aktualizacji

W literaturze różne podejścia:

- na podstawie maski obiektów ruchomych z generacji tła (brak odporności na błędy w tle, powstawanie ghost’ów),
- zliczanie liczby klasyfikacji pikseli jako obiekt,
- analiza czy piksel nie jest naprzemienne klasyfikowany jako obiekt i jako tło,
- itp.

ViBE (Visual Background Extractor) – model tła

Metoda została zaproponowana przez O. Barnich and M. Van Droogenbroeck (University of Liege, Belgium) – 2009. Rozwiązanie jest opatentowane

Model tła jest oparty na zbiorze zaobserwowanych wartości piksela (bufor), nie jest to żaden rozkład prawdopodobieństwa.

$$M(x) = \{v_1, v_2, \ldots, v_N\}$$

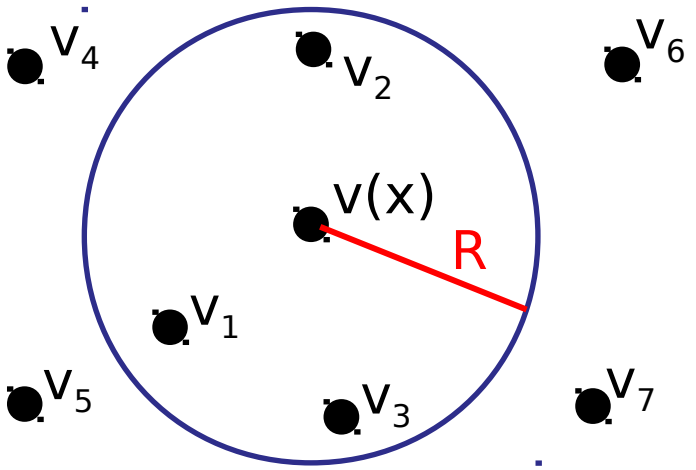
Barnich O., Van Droogenbroeck M.: ViBe: A universal background subtraction algorithm for video sequences. In IEEE Transactions on Image Processing, 20(6):1709-1724, June 2011

ViBE - klasyfikacja

Klasyfikacja obiekt – tło:

-dla danego piksela v(x) definiowana jest sfera $S_R(v(x))$ o promieniu R i środku w punkcie v(x),

-piksel uznawany jest za tło, jeśli co najmniej $\#_{min}$ próbek z modelu leży wewnątrz sfery



ViBE – określanie położenia wewnątrz sfery

Problem obliczania odległości pomiędzy dwoma pikselami:

- dla odcieni szarości proste:

$$dI = |I_A - I_B|$$

- w kolorze niekoniecznie

$$dI = |R_A - R_B| + |G_A - G_B| + |B_A - B_B|$$

$$dI = \sqrt{(R_A - R_B)^2 + (G_A - G_B)^2 + (B_A - B_B)^2}$$

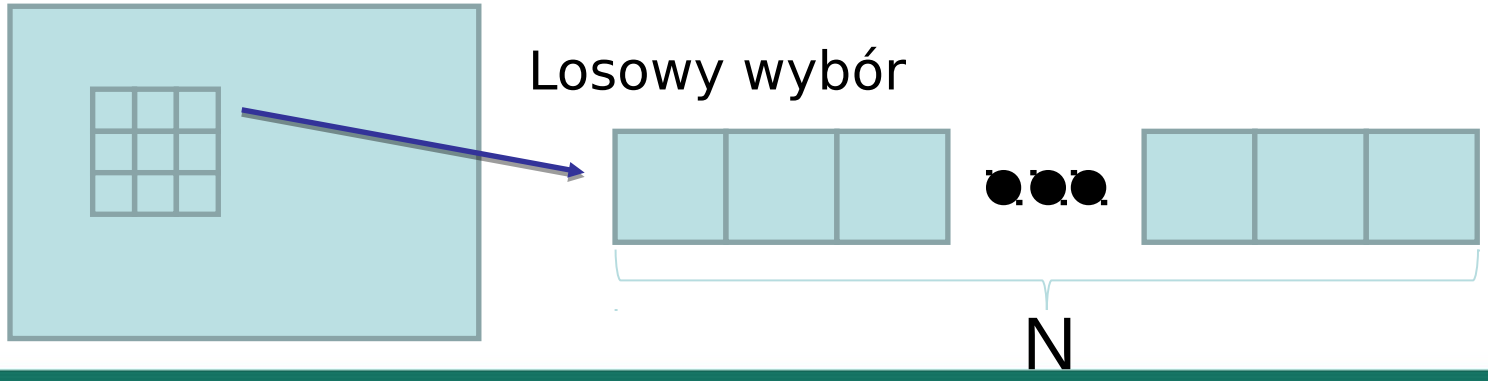
ViBE - inicjalizacja

Inicjalizacja modelu tła:

-na podstawie jednej ramki,

-wykorzystuje założenie o podobieństwie otoczenia – do buforu M trafiają losowe próbki z otoczenia o rozmiarze 3x3,

-wada podejścia to podatność na *ghosty* – przenikanie obiektów pierwszoplanowych do modelu



ViBE - aktualizacja

Aktualizacja modelu tła:

-podejście konserwatywne – aktualizacji podlegają tylko te piksele, które zostały sklasyfikowane jako tło,

-zaleta – odporność na przenikanie obiektów ruchomych do modelu tła,

-wada – podatność na *ghosty* (ogólnie na błędy klasyfikacji) – np. puste miejsce po samochodzie.

ViBE – aktualizacja cd.

Procedura aktualizacji:

a) Piksel sklasyfikowany jako tło ?
b) Czy aktualizować ? (p=1/16) - losowość
c) Wybór losowy próbki do wymiany (1/N)
d) Aktualizacja sąsiedztwa:
    i. Wybór sąsiada (1/8) - losowy,
    ii. Wybór próbki (1/N) - losowy.

Wybrane próbki zastępowane są bieżącą wartością piksela v(x).

ViBE - cechy

Cechy rozwiązania:

-połączenie podejścia rekurencyjnego i nierekurencyjnego: mały bufor próbek, brak analizy zależności czasowych,

-niewielka liczba parametrów:

- N (rozmiar bufora) - 20,
- R (promień sfery) – 30 (odc. szarości),
- $\#_{min}$ (min. liczba próbek wewnątrz sfery)- 2,
- prawdopodobieństwo aktualizacji - 1/16.

PBAS (Pixel Based Adaptive Segmenter) – model tła

Metoda zaproponowana przez M. Hofmann, P. Tiefenbacher i G. Rigoll – 2012

Model oparty o bufor N próbek z sekwencji wideo:

$$B(x_i)=\{B_1(x_i),B_2(x_i),\ldots,B_N(x_i)\}$$

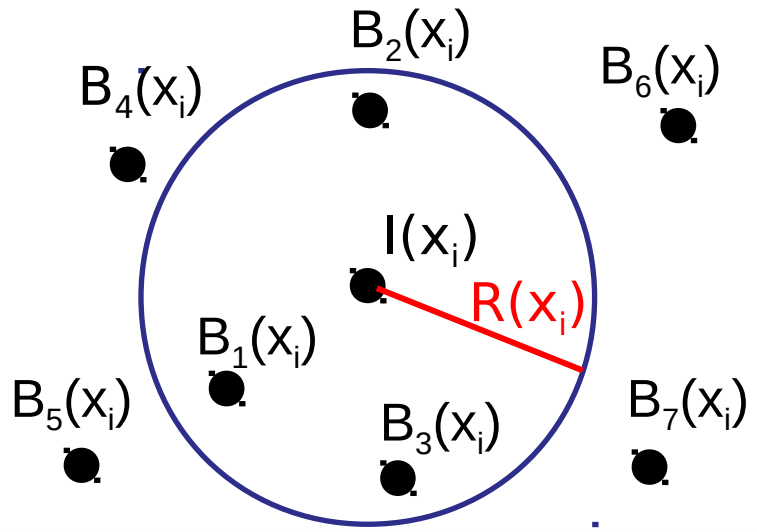
M. Hofmann, P. Tiefenbacher, and G. Rigoll. Background segmentation with feedback: The pixel-based adaptive segmenter. In Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on, pages 38–43, June.

PBAS – klasyfikacja pierwszy plan/tło

$$\left\{ dist\left(I\left(x_i \right), B_k\left(x_i \right) \right) < R\left(x_i \right) \right\} <_{\min} ¿$$

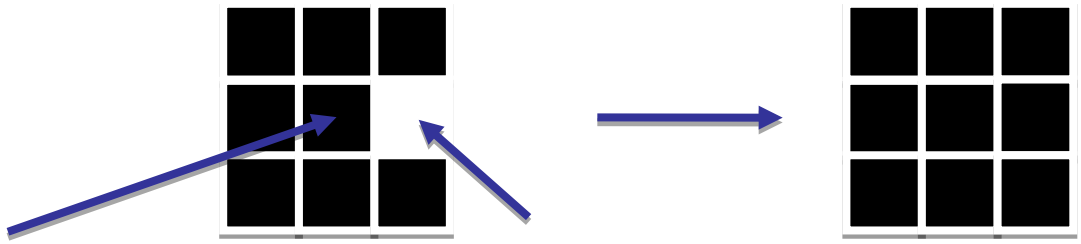
$$¿\, F\left(x_i \right) = \left\{ 1 \;\; if \;\; \sum_{k=0}^{N} \; 0 \right\}$$

dist – odległość pomiędzy bieżącym pikselem a próbką z modelu tła (n.p. L1 or L2),
N – liczba próbek w modelu,
$R(x_i)$ – próg odległości,
$\#_{min}$- próg liczby próbek.



PBAS – aktualizacja modelu tła

- Podejście konserwatywne (tylko piksele klasyfikowane jako tło są uaktualniane),
- Może to prowadzić do nieusuwalnych błędów w przypadku nieprawidłowej klasyfikacji,
- Dlatego aktualizuje się rozpatrywany piksel i jeden losowo wybrany z otoczenia 3x3,
- Prawdopodobieństwo $p = 1 / T(x_i)$,
- R(x) i T(x) zależą od zawartości obrazu

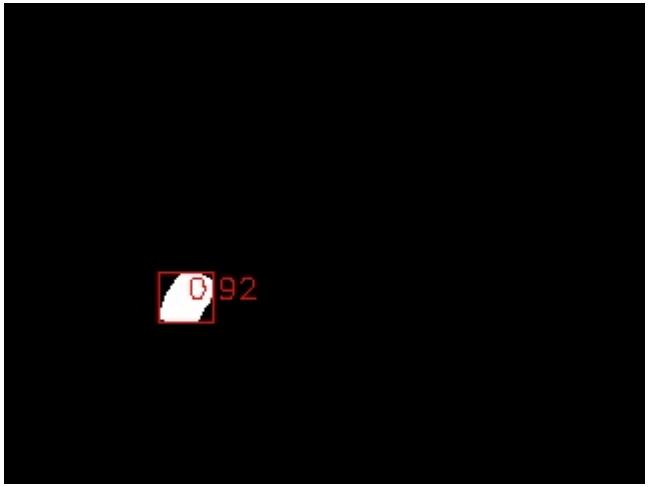


rozpatrywany piksel

sąsiedni piksel do aktualizacji

Ewaluacja metod segmentacji (1)

maska



referencja

Ewaluacja metod segmentacji (2)

Możliwe są 4 sytuacje:

- TP – **obiekt** sklasyfikowany **poprawnie** jako **obiekt**
- TN – **tło** sklasyfikowane **poprawnie** jako **tło**
- FP – **tło** sklasyfikowane **niepoprawnie** jako **obiekt**
- FN – **obiekt** sklasyfikowany **niepoprawnie** jako **tło**

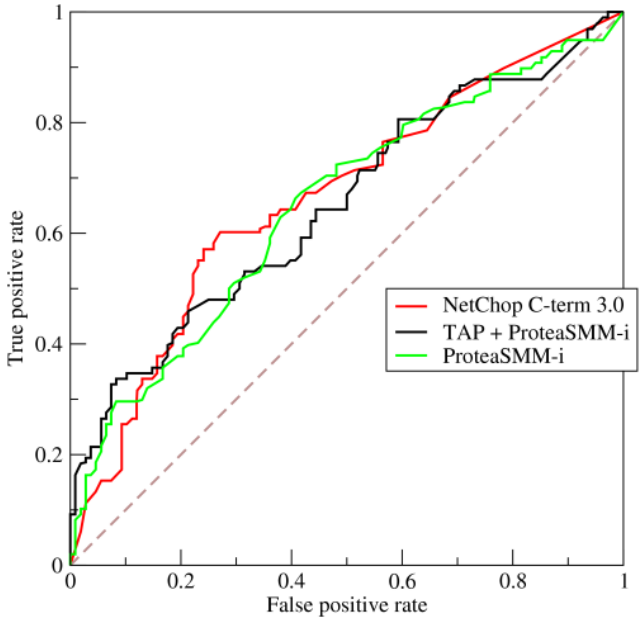
Wynik klasyfikacji binarnej

Ewaluacja metod segmentacji (3)

$$TPR = R = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{TN + FN}$$

$$S = \frac{TP}{TP + FP + FN}$$



$$P = \frac{TP}{TP + FP}$$

$$F1 = 2\frac{PR}{R + R}$$

$$ACC = \frac{TP + TN}{TP + FP + TN + FN}$$

Ewaluacja metod segmentacji (4)

Uwagi:

- ze statystyką trzeba ostrożnie,
- na wyniki trzeba patrzeć krytycznie,
- należy przeanalizować maski,
- ew. ocenić je poprzez pryzmat docelowej aplikacji,
- „optymalizacja” parametrów metody
- metoda dobra, bo względnie obiektywnie pozwala na porównanie różnych algorytmów