


 Zachodniopomorski
 Uniwersytet Technologiczny
 w Szczecinie

 Wydział
 Informatyki

Widzenie Komputerowe

wykład #4: algorytmy modelowania tła

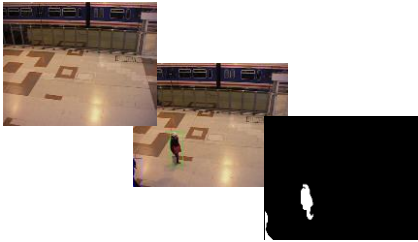
dr hab. inż. Paweł Forczmański, prof. ZUT


Semestr letni 2020/2021

1

Spis treści:


- Modelowanie tła
 - Obiekt pierwszoplanowy, tło
 - Modelowanie rozkładów kolorów
 - Klasyfikacja tło/obiekt pierwszoplanowy
 - Usuwanie cienia
- Przykłady





 Wydział
 Informatyki

2


Idea





 Wydział
 Informatyki

3

Idea






 Wydział
 Informatyki

4

Cele modelowania tła

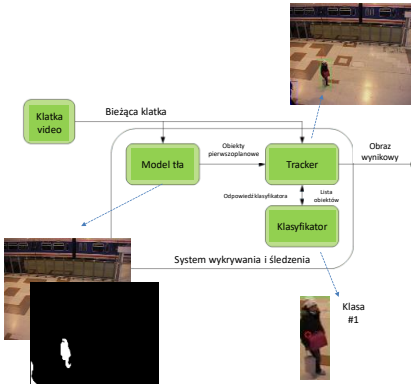
- Zmniejszenie złożoności problemu do dalszego przetwarzania
- Organicznie obszar przetwarzania tylko części obrazu, która zawiera odpowiednie informacje
- Segmentacja obrazu na pierwszy plan i tło
- Wprowadzenie wirtualnego tła





 Wydział
 Informatyki

5

Schemat przetwarzania




 Wydział
 Informatyki

6

Podział metod

- **Algorytmy na poziomie pikseli.**
 - używają tylko informacji zebranych z pojedynczych pikseli
 - są bardzo szybkie, ale nie używają żadnego rodzaju relacji między pikselami.
- **Algorytmy na poziomie bloków**
 - dzielą obraz na bloki i obliczają relacje pomiędzy blokami opisującymi tło.
 - są zwykle bardziej odporne na szum niż podejścia na poziomie pikseli i zapewniają lepsze wykrywanie obiektów pierwszoplanowych ale są kosztowne obliczeniowo.
- **Algorytmy na poziomie regionów**
 - dzielą obraz na zbiór regionów, którymi następnie są sklasyfikowane jako tło lub pierwszy plan
 - Wykorzystują relacje przestrzenne i kryteria spójności obszarów
 - Są kosztowne obliczeniowo



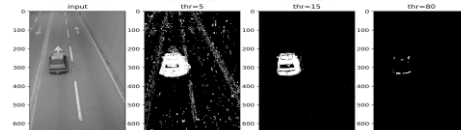
7

Wprowadzenie : odejmowanie tła

- Większość technik BS (background subtraction) ma wspólny mianownik: zakładają, że obserwowana sekwencja wideo składa się ze statycznego tła, przed którym obserwuje się poruszające się obiekty.
- Zakładając, że każdy poruszający się obiekt ma kolor (lub rozkład kolorów) inny niż ten obserwowany w B_s , wiele metod BS można podsumować następującym wzorem

$$\mathcal{X}_t(s) = \begin{cases} 1 & \text{dla } d(I_{s,t}, B_s) > \tau \\ 0 & \text{w przeciwnym przypadku} \end{cases}$$

gdzie τ to próg (threshold), \mathcal{X}_t to etykieta ruchu w momencie t (maska ruchu), d to odległość pomiędzy $I_{s,t}$ (kolor w czasie t piksela s) a B_s to model tła dla piksela s

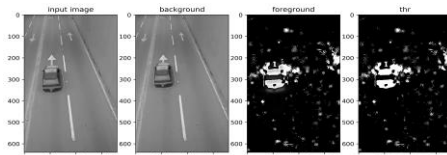


8

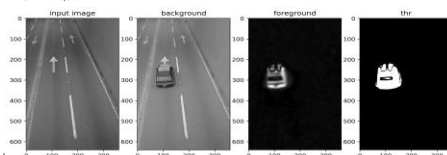
Wprowadzenie : odejmowanie tła

- Odejmowanie sąsiednich klatek lub średniej z początkowych klatek sekwencji

$$B_s = I_{s,t-1}$$



$$B_s = \sum I_{s,0:t-1}$$



9

Przegląd metod

- Odejmowanie mediany - alternatywa dla użycia średniej wartości pikseli w sekwencji wideo do modelowania tła
- Jako tło sceny można użyć wartości mediany ostatnich N klatek.
- Główna zaleta jest, że obraz tła nie jest degradowany przez pojawiające się obiekty, pod warunkiem, że tło jest widoczne przez więcej niż $N/2$ klatek.
- Z drugiej strony, obliczenia wymagają bufora do przechowywania ostatnich N ramek.
- Metoda iteracyjnego przybliżania wartości mediany, unikająca zatem potrzeby stosowania bufora ramek, przedstawiono w [McFarlane i Schofield, 1995].
- Dla każdej przychodzącej ramki obraz tła jest aktualizowany w następujący sposób:

$$B_{s,t+1} = \begin{cases} B_{s,t} + 1 & \text{jeżeli } I_{s,t} > B_{s,t} \\ B_{s,t} - 1 & \text{jeżeli } I_{s,t} < B_{s,t} \end{cases}$$



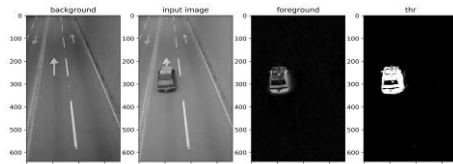
[McFarlane and Schofield, 1995]

10

- Odejmowanie mediany z początkowych klatek sekwencji

Wprowadzenie : odejmowanie tła

$$B_s = \text{median}(I_{s,0:t-1})$$

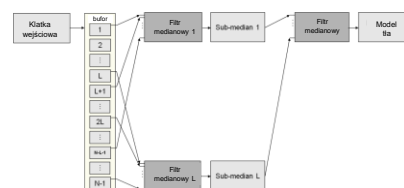


Yannick Berachet, Pierre-Marc Jodoin, Bruno Emile, Hélène Laurent, Christophe Rosenberg. Comparative study of background subtraction algorithms. Journal of Electronic Imaging, SPIE and IS&T 2010, 19

11

Wprowadzenie : odejmowanie mediany

- bardziej solidny model tła w przypadku często pojawiających się obiektów pierwszoplanowych
- Poprzez zwiększenie odporności na ww. obiekty można zmniejszyć rozmiar bufora, a tym samym przyspieszyć obliczenia
- Klatki wejściowe dzieli się na L grup po $(N-1)/L$ klatek
- W każdej grupie oblicza się medianę, które następnie są (poddawane kolejnej) filtracji medianowej
- Zmniejsza to wymagania pamięciowe i obliczeniowe



[Karaman, 2010]

12

KDE

- Aby poradzić sobie ze zmianami o wysokiej częstotliwości i dowolnymi rozkładami, można stosować nieparametryczne modele tła
- Prawdopodobieństwo zaobserwowania danej wartości piksela X_t w czasie t przy użyciu estymatora jądrowego K można oszacować nieparametrycznie na podstawie próbkę pikseli $X = \{X_1, X_2, \dots, X_N\}$

w następujący sposób:

$$p(X_t) = \sum_{i=1}^N \alpha_i K(X_t - X_i)$$

gdzie α_i to współczynniki wagowe (zwykle wybrane z rozkładu jednorodnego $\alpha_i = 1/N$).

- Prawdopodobieństwo w ww. równaniu można skutecznie obliczyć, przyjmując funkcję rozkładu normalnego $N(0, \Sigma)$ jako estymator jądra, zakładając niezależność między różnymi kanałami kolorów, i używając wstępnie obliczonych tablic LUT dla funkcji jądra, biorąc pod uwagę różnice wartości intensywności ($X_t - X_i$)

13

KDE

- Zastosowanie nieparametrycznych modeli tła zostało po raz pierwszy zaproponowane w [Elgammal et al., 2000] i [Elgammal et al., 2002].
- Metoda łagodzi wysokie wymagania dotyczące pamięci narzucone przez konieczność przechowywania całego przykładowego zestawu ramek branych pod uwagę przy estymacji rozkładu gęstości,
- Technika estymacji oparta na algorytmie mean-shift
- Podejście wykorzystujące zmienną wielkość jądra
- Ma wysoki koszt obliczeniowy.
- Ponadto, w [Zivkovic i van der Heijden, 2006] wykazano, że GMM wydaje się być lepszym modelem dla prostych scen, zapewniając jednocześnie bardziej zwartą reprezentację, która nadaje się do dalszych etapów przetwarzania, jak np. wykrywanie cieni.

14

Modelowanie ze słownikiem (książka kodowa)

- Jako alternatywę dla modeli statystycznych zaproponowano również modele książki kodowej
- Wygląd tła opisuje się poprzez środki słów kodowych.
- Zbiór słów kodowych opisujących piksel stanowi jego książkę kodową.
- Każde słowo kodowe składa się z wektora koloru $v = (R, G, B)$ i kilku parametrów pomocniczych.
- Wartości każdego piksela są porównywane z odpowiednią książką kodową w celu sklasyfikowania go jako tło lub pierwszy plan.

$M = \{c_m | c_m \in C \wedge \lambda_m \leq T_{\lambda}\}$

Algorithm for Background Subtraction

- $x = (R, G, B)$, $l = R + G + B$
- For all codewords in M in Eq. 1, find the codeword c_m matching to x based on two conditions:
 - $colorDist(x, v_m) \leq \epsilon_2$
 - $brightness(l, (l_m, l_m)) = true$
- $BGS(x) = \begin{cases} foreground & \text{if there is no match} \\ background & \text{otherwise.} \end{cases}$

15

Modelowanie ze słownikiem (książka kodowa)

Most of the time, the pixel shows sky colors

The tree shows up quasi-periodically with an acceptable λ

A pixel on the top of the tree was sampled

The person occupied the pixel after this period

time

16

Elgenbackground

- Kompensowanie zmian oświetlenia na poziomie kadru
- Uwzględnia korelacje przestrzenne poprzez tzw eigenspace
- Model eigenspace jest obliczany ze zbioru N klatek i średniego obrazu tła oraz macierzy kowariancji
- Macierz kowariancji jest rozkładana za pomocą SVD.
- PCA jest wykorzystywane do redukcji wymiarowości
- Zachowane są wektory własne M odpowiadające największym wartościom własnym.
- Wektory własne są przechowywane w macierzy Φ_{Mb} o rozmiarze $M \times p$, gdzie p to liczba pikseli w klatce.
- Dla każdej ramki wejściowej I_t , średni znormalizowany wektor obrazu jest rzutowany do przestrzeni własnej i rzutowany wsteczne w przestrzeń obrazu przy użyciu macierzy wektorów własnych Φ_{Mb} i jej transpozycji.

17

Elgenbackground

- Ponieważ przestrzeń własna zapewnia stabilny model tła, ale jedynie dla nieporuszających się obiektów, obraz wejściowy B_t poddany projekcji wstecznej nie powinien zawierać poruszających się obiektów.
- Dlatego progując różnicę euklidesową pomiędzy obrazem wejściowym I_t i obrazem wstecznie rzutowanym B_t , można wykryć poruszające się obiekty.

(a) Training image (b) Original test image (c) Reconstructed background

(a) Sample training images

(b) BGS result at $t = 277$ (left) $t = 2991$ (right)

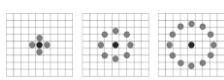
18

Metody teksturalne

- Tekstury można stosować w celu uniezależnienia się od zmiennego oświetlenia
- Podejście blokowe zaadaptowane z [Heikkilä i Pietikäinen, 2006].
- Zamiast używać funkcji koloru lub intensywności, metody te wykorzystują tzw. dyskryminacyjne miary tekstury do obliczania statystycznych cech tła
- Obliczane przy użyciu lokalnego wzorca binarnego (LBP).

$$LBP = \sum_{p=1}^P s(X_i - X_c)2^{p-1}, \quad s(x) = \begin{cases} 1 & x \geq 0, \\ 0 & x < 0. \end{cases}$$

3	4	5
1	3	4
3	0	2



M. Heikkilä and M. Pietikäinen, "A texture-based method for modeling the background and detecting moving objects," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 4, pp. 657-662, April 2006.

19

Metody teksturalne

- Model tła dla piksela składa się z grupy K histogramów LBP, gdzie K jest wybierane przez użytkownika
- Każdy histogram modelu ma wagę od 0 do 1, tak że wagi histogramów modelu K sumują się do jednego.
- Na pierwszym etapie przetwarzania porównuje się histogram każdego piksela z modelem za pomocą przecięcia

$$\cap(\vec{a}, \vec{b}) = \sum_{n=0}^{N-1} \min(a_n, b_n),$$

- Następnie wybiera się najbliższy histogram i koryguje się go uwzględniając aktualną wartość piksela i otoczenia
- Segmentacja tła odbywa się poprzez progowanie pikseli odpowiednio daleko oddalonych od grupy histogramów



M. Heikkilä and M. Pietikäinen, "A texture-based method for modeling the background and detecting moving objects," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 28, no. 4, pp. 657-662, April 2006.

20

Wprowadzenie : odedmowanie tła

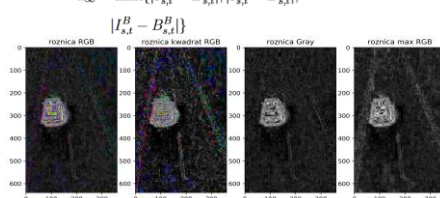
Główną różnicą między metodami BS jest to, w jaki sposób modeluje się B i jaką metrykę odległości d się stosuje.

$$d_0 = |I_{s,t} - B_{s,t}|$$

$$d_1 = |I_{s,t}^R - B_{s,t}^R| + |I_{s,t}^G - B_{s,t}^G| + |I_{s,t}^B - B_{s,t}^B|$$

$$d_2 = (I_{s,t}^R - B_{s,t}^R)^2 + (I_{s,t}^G - B_{s,t}^G)^2 + (I_{s,t}^B - B_{s,t}^B)^2$$

$$d_\infty = \max\{|I_{s,t}^R - B_{s,t}^R|, |I_{s,t}^G - B_{s,t}^G|, |I_{s,t}^B - B_{s,t}^B|\}$$



Yannick Benzeath, Pierre-Marc Jodan, Bruno Emis, Helene Laurent, Christophe Rosenberg, "Com-parative study of background subtraction algorithms," Journal of Electronic Imaging, SPIE and ISAT, 2010, 19.

21

Wprowadzenie : odedmowanie tła

- Zalety:
 - bardzo proste w implementacji i użyciu
 - Bardzo szybkie
 - Odpowiednie modele tła nie są stałe w czasie (ulegają zmianom)
- Wady
 - Dokładność odedmowania klatek zależy od szybkości obiektów i klatek/sek (framerate)
 - Modele bazujące na średniej i medianie mają stosunkowo wysokie wymagania pamięciowe
 - Można je obniżyć w przypadku średniej za pomocą średniej ruchomej...

Yannick Benzeath, Pierre-Marc Jodan, Bruno Emis, Helene Laurent, Christophe Rosenberg, "Com-parative study of background subtraction algorithms," Journal of Electronic Imaging, SPIE and ISAT, 2010, 19.

22

Wprowadzenie : odedmowanie tła

- Wady
 - Problem z doбором progu t
 - Proóg jest globalny dla wszystkich pikseli obrazu
 - Proóg nie jest funkcją czasu
- Dlatego wyniki stosowania ww. modeli nie dadzą dobrych rezultatów w nast. sytuacjach:
 - Tło jest bi-modalne (zmienia cyklicznie swoje wartości)
 - Jeśli scena zawiera obiekty, które poruszają się bardzo wolno (średnia i mediana)
 - Jeśli scena zawiera obiekty bardzo szybkie a framerate jest niskie (odedmowanie klatek)
 - Jeśli oświetlenie zmienia się w czasie

Yannick Benzeath, Pierre-Marc Jodan, Bruno Emis, Helene Laurent, Christophe Rosenberg, "Com-parative study of background subtraction algorithms," Journal of Electronic Imaging, SPIE and ISAT, 2010, 19.

23

Wprowadzenie : odedmowanie tła

Korzenie odedmowania tła sięgają XIX wieku, kiedy wykazano, że obraz tła można uzyskać wystawiając film na okres znacznie dłuższy niż czas wymagany do poruszania się obiektu w polu widzenia. Zatem w najprostszej postaci obraz tła jest średnim obrazem długoterminowym:

$$B(x, y, t) = \frac{1}{t} \sum_{t'=1}^t I(x, y, t')$$

gdzie $I(x, y, t)$ jest chwilową wartością piksela dla (x, y) .

Można to obliczyć przyrostowo:

$$B(x, y, t) = \frac{(t-1)}{t} B(x, y, t-1) + \frac{1}{t} I(x, y, t)$$

Wariancję można również obliczyć przyrostowo, a poruszające się obiekty można zidentyfikować, progując odległość Mahalanobisa między $I(x, y, t)$ i $B(x, y, t)$.

N.Friedman, S.Russell, "Image segmentation in video sequences: A probabilistic approach, Proceedings Thirteenth Conf. On Uncertainty in Artificial Intelligence, 1997

24

Wprowadzenie: odejmowa nie tła

Jednym z problemów związanych z tym podejściem jest to, że warunki oświetleniowe zmieniają się z czasem.

Można sobie z tym poradzić za pomocą ruchomej średniej lub, bardziej efektywnie, stosując zapomnianie wykładnicze.

W tym drugim schemacie każdy udział obrazu w tle jest ważony, aby wykładniczo zmniejszyć się w miarę cofania w przeszłość. Jest to realizowane za pomocą równania aktualizacyjnego:

$$B(x, y, t) = (1 - \alpha)B(x, y, t - 1) + \alpha I(x, y, t)$$

gdzie $1/\alpha$ jest stałą czasową procesu zapomniania.



N.Friedman, S.Russell, Image segmentation in video sequences: A probabilistic approach, Proceedings Thirteenth Conf. On Uncertainty in Artificial Intelligence, 1997

25

Wprowadzenie: odejmowa nie tła

Typowe problemy:

- Cienie (duchy) – nawet do 50% błędów
- Obiekty wolno poruszające się lub okresowo zatrzymujące się
- Kamera, która nie jest w pełni stacjonarna (wahania, drgania)



N.Friedman, S.Russell, Image segmentation in video sequences: A probabilistic approach, Proceedings Thirteenth Conf. On Uncertainty in Artificial Intelligence, 1997

26

Pojedynczy rozkład Gausa

- Modelowanie B z pojedynczym obrazem wymaga rygorystycznie ustalonego tła, wolnego od szumów i artefaktów.
- Wymaganie jest trudne do spełnienia w rzeczywistości więc modeluje się każdy piksel tła za pomocą funkcji gęstości prawdopodobieństwa (PDF) obliczonej dla sekwencji ramek treningowych.
- Wtedy problem BS staje się problemem progowania PDF, w przypadku którego piksel o niskim prawdopodobieństwie może odpowiadać obiektowi poruszającemu się na pierwszym planie.
- Na przykład, w celu uwzględnienia szumu, Wrenet i wsp. [33] modelują każdy piksel tła z rozkładem Gausa $N(\mu_{s,t}, \Sigma_{s,t})$, gdzie $\mu_{s,t}$ i $\Sigma_{s,t}$ oznaczają odpowiednio średni kolor tła i macierz kowariancji dla piksela s w czasie t.
- W tym kontekście metrykę odległości może obliczać tak:

$$d_G = \frac{1}{2} \log((2\pi)^3 |\Sigma_{s,t}|) + \frac{1}{2} (I_{s,t} - \mu_{s,t})^T \Sigma_{s,t}^{-1} (I_{s,t} - \mu_{s,t})$$

$$d_M = |I_{s,t} - \mu_{s,t}| \Sigma_{s,t}^{-1} |I_{s,t} - \mu_{s,t}|^T$$

Wydział
Informatyki

27

Pojedynczy rozkład Gausa

- Ponieważ macierz kowariancji zawiera duże wartości w obszarach zaszumionych i niskie wartości w obszarach bardziej stabilnych, Σ sprawia, że próg jest lokalnie zależny od ilości szumu.
- Innymi słowy, im bardziej zaszumiony piksel, tym większy musi być gradient czasowy $|I_{s,t} - \mu_{s,t}|$, aby piksel został oznaczony jako ruch.
- To sprawia, że metoda jest znacznie bardziej elastyczna niż podstawowa metoda wykrywania ruchu.
- Ponieważ oświetlenie często zmienia się w czasie, średnią i kowariancję każdego piksela można również iteracyjnie aktualizować zgodnie z następującą procedurą:

$$\mu_{s,t+1} = (1 - \alpha) \mu_{s,t} + \alpha I_{s,t}$$

$$\Sigma_{s,t+1} = (1 - \alpha) \Sigma_{s,t} + \alpha (I_{s,t} - \mu_{s,t})(I_{s,t} - \mu_{s,t})^T$$

- Jeśli Σ jest z definicji macierzą 3×3 , można założyć, że jest ona diagonalna, aby zmniejszyć koszty pamięciowe i obliczeniowe.

Wydział
Informatyki

28

Wprowadzenie: modele zaawansowane

Na początku powszechnie stosowaną metodą było modelowanie tła za pomocą **mieszanek Gaussa** (GMM) [1] i statystyki [2].

Główną zaletą GMM jest to, że może osiągnąć przetwarzanie w czasie rzeczywistym. Jednocześnie GMM są wrażliwe na mały szum, taki jak zmiana luminancji.

W ostatnich latach stwierdzono, że model **odpornej analizy komponentów głównych** (RPCA) jest skuteczniejszy niż inne najnowocześniejsze metody [3] [4] [5].

Z drugiej strony, jest bardziej kosztowny obliczeniowo.

- [1] N.Friedman, S.Russell, Image segmentation in video sequences: A probabilistic approach, Proceedings Thirteenth Conf. On Uncertainty in Artificial Intelligence, 1997
 [2] Rita Cucchiara, Costantino Giana, Massimo Piccardi, Andrea Prati Detecting Moving Objects, Ghosts, and Shadows in Video Streams IEEE TPAMI, VOL. 25, NO. 10, Oct. 2003
 [3] John Wright - Yigang Peng, Yi Ma, Anand Ganesh, Shankar Rao Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Matrices by Convex Optimization NIPS 2009
 [4] Xiaowei Zhou, Can Yang and Weichuan Yu Moving Object Detection by Detecting Contiguous Outliers in the Low-Rank Representation Pattern Analysis and Machine Intelligence, IEEE Transactions 2012
 [5] Bo Xin Yuan Tian Yizhou Wang Wen Gao Background Subtraction via Generalized Fused Lasso Foreground Modeling CVPR 2015

Wydział
Informatyki

29

GMM

GMM wykorzystuje 3 ~ 5 funkcji Gaussa dla każdego piksela kanału koloru w ramce wideo.

Algorytm sprawdza nowe piksele wejściowe, czy ich wartość jest mniejsza niż odchylenie Gaussa, które jest określone przez poprzednie wartości pikseli tej samej lokalizacji, aby ustalić, czy piksel ten jest na pierwszym planie, czy nie.



Zoran Zivkovic Improved Adaptive Gaussian Mixture Model for Background Subtraction In Proc. ICPR, 2004

Wydział
Informatyki

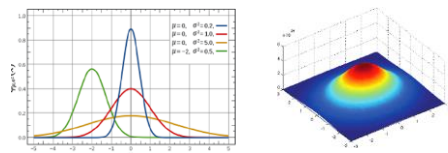
30

Rozkład Gaussa

- W rezultacie piksel, gdy stanie się pierwszym planem, może ponownie stać się tłem tylko wtedy, gdy wartość intensywności zbliży się do tej, jaka była przed zmianą na pierwszy plan.
- Metoda ma jednak kilka wad: działa tylko wtedy, gdy wszystkie piksele są początkowo pikselami tła (lub piksele pierwszego planu są oznaczone jako takie).

$$\mathcal{N}(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

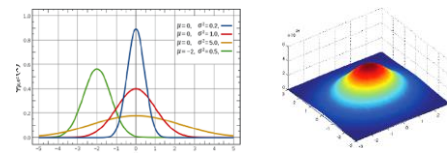
$$\mathcal{N}(\mathbf{x}|\mu, \Sigma) = \frac{1}{(2\pi)^{D/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^T \Sigma^{-1}(\mathbf{x}-\mu)}$$



31

Rozkład Gaussa

- Plik PDF każdego piksela charakteryzuje się średnią i wariancją.
- Aby zainicjować wariancję, możemy na przykład użyć wariancji w x i y z małego okienka wokół każdego piksela.
- Należy pamiętać, że tło może zmieniać się w czasie (np. z powodu zmian oświetlenia lub niestacystycznych obiektów w tle).
- Aby dostosować się do tej zmiany, w każdej klatce należy zaktualizować średnią i wariancję każdego piksela



32

Rozkład Gaussa

- Plik PDF każdego piksela charakteryzuje się średnią i wariancją.
- Aby zainicjować wariancję, możemy na przykład użyć wariancji w x i y z małego okienka wokół każdego piksela.
- Należy pamiętać, że tło może zmieniać się w czasie (np. z powodu zmian oświetlenia lub niestacystycznych obiektów w tle).
- Aby dostosować się do tej zmiany, w każdej klatce należy zaktualizować średnią i wariancję każdego piksela

$$\hat{p}(\vec{x}|\mathcal{X}_T, BG+FG) = \sum_{m=1}^M \hat{\pi}_m \mathcal{N}(\vec{x}; \hat{\mu}_m, \hat{\sigma}_m^2 I)$$



33

GMM



'traffic' sequence

average processing time per frame Old: 19.1ms New: 13.0ms



selected number of modes M

average processing time per frame Old: 19.1ms New: 13.0ms



'lab' sequence

average processing time per frame Old: 19.3ms New: 15.9ms



selected number of modes M

average processing time per frame Old: 19.3ms New: 15.9ms



'trees' sequence

average processing time per frame Old: 19.7ms New: 19.3ms



selected number of modes M

average processing time per frame Old: 19.7ms New: 19.3ms

Zoran Zivkovic Improved Adaptive Gaussian Mixture Model for Background Subtraction In Proc. ICPR, 2004

34

GMM: algorytm

Rozważmy pojedynczy piksel $i_{x,y}$ i rozkład jego wartości w czasie:

- Przez określony czas będzie on w „normalnym” stanie (tło) - na przykład na niewielkim obszarze powierzchni drogi.
- Przez inny okres może znajdować się w cieniu poruszających się pojazdów, a czasem może być częścią pojazdu.

Tak więc, w przypadku monitoringu wizyjnego, możemy założyć rozkład wartości $i_{x,y}$ piksela (x, y) jako ważoną sumę trzech rozkładów: $r_{x,y}$ (droga), $s_{x,y}$ (cienie) i $v_{x,y}$ (pojazd):

$$i_{x,y} = \mathbf{w}_{x,y} \cdot (r_{x,y}, s_{x,y}, v_{x,y})$$

Ważne jest określenie różnych modeli dla każdego piksela, ponieważ różne fragmenty obrazu mogą odpowiadać różnym fizycznym obiektom (obszarom).

Wagi są również indeksowane, ponieważ niektóre piksele mogą „spędzać” więcej czasu w cieniu lub pojeździe niż inne

N.Friedman, S.Russell, Image segmentation in video sequences: A probabilistic approach, Proceedings Thirteenth Conf. On Uncertainty in Artificial Intelligence, 1997



35

GMM: algorytm

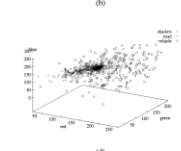
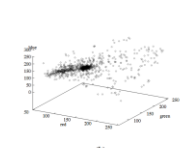
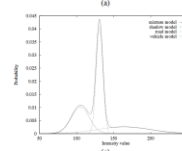
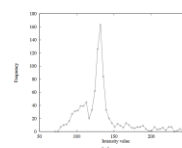


Figure 3: (a) Empirical distribution of intensity values for pixel (160,170) over 1000 frames. (b) Scatter plot of RGB values for the same pixel. (c) Fitted three-component Gaussian mixture model for the data in (a). (d) Scatter plot of 1000 randomly-generated data points from a fitted three-component Gaussian mixture model for the data in (b).

N.Friedman, S.Russell, Image segmentation in video sequences: A probabilistic approach, Proceedings Thirteenth Conf. On Uncertainty in Artificial Intelligence, 1997



36

GMM: algorytm

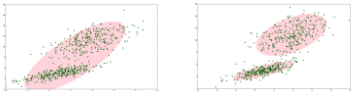
Model piksela (x, y) jest parametryzowany parametrami:

$$\Theta = \{w_l, \mu_l, \Sigma_l : l \in \{r, s, v\}\}$$

$$w_{x,y} = (w_r, w_s, w_v) \quad r_{x,y} \sim N(\mu_r, \Sigma_r)$$

Model budowany jest w dwóch trybach. W pierwszym, badany jest poziomy intensywności a μ i Σ to skalary.

W drugim, badane są wartości RGB i μ jest wektorem 3x1, a Σ to macierz 3x3.



N. Friedman, S. Russell, "Image segmentation in video sequences: A probabilistic approach, Proceedings Thirteenth Conf. On Uncertainty in Artificial Intelligence, 1997"

37

GMM: algorytm

Niech i będzie wartością piksela (intensywność lub wektor wartości RGB). Niech L jest zmienną losową oznaczającą etykietę piksela na obrazie.

Model określa prawdopodobieństwo, że $L = l$ oraz $I(x, y, t) = i$ takie, że

$$P(L = l, I(x, y, t) = i \mid \Theta) = w_l \cdot (2\pi)^{-\frac{d}{2}} |\Sigma_l|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(i - \mu_l)^T \Sigma_l^{-1} (i - \mu_l)\right\}$$

gdzie d jest wymiarem każdej wartości piksela (1 lub 3).

Biorąc pod uwagę te prawdopodobieństwa, można sklasyfikować wartość piksela.

W tym celu wybiera się klasę o największym prawdopodobieństwie a posteriori $P(L = l \mid I(x, y, t))$.

PROBLEM: aktualizacja wag i parametrów rozkładu...

N. Friedman, S. Russell, "Image segmentation in video sequences: A probabilistic approach, Proceedings Thirteenth Conf. On Uncertainty in Artificial Intelligence, 1997"

38

GMM: algorytm

- Potrzebnych jest wiele adaptacyjnych rozkładów Gaussa.
- W tym celu używa się mieszanki Gaussianów, aby aproksymować proces dostosowania się modelu do zmian.
- Za każdym razem parametry Gaussianów są aktualizowane.
- Analizie podlega zmiana wartości określonego piksela w czasie. Jest to szereg wartości pikseli, np. skalary dla obrazów w odcieniach szarości i wektory dla obrazów kolorowych.
- W dowolnym momencie t , znana jest „historia” danego piksela,

$$\{X_1, \dots, X_t\} = \{I(x_0, y_0, i) : 1 \leq i \leq t\}$$
- Wartość każdego piksela, $\{X_1, \dots, X_t\}$, jest modelowana przez mieszaninę K rozkładów Gaussa.
- Prawdopodobieństwo zaobserwowania bieżącej wartości piksela wynosi

$$P(X_t) = \sum_{i=1}^K \omega_{i,t} * \eta(X_t, \mu_{i,t}, \Sigma_{i,t})$$

C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," Proceedings, 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1999, pp. 246-252 Vol. 2.

39

GMM: algorytm

Wartość każdego piksela, $\{X_1, \dots, X_t\}$, jest modelowana przez mieszaninę K rozkładów Gaussa. Prawdopodobieństwo zaobserwowania bieżącej wartości piksela wynosi

$$P(X_t) = \sum_{i=1}^K \omega_{i,t} * \eta(X_t, \mu_{i,t}, \Sigma_{i,t})$$

- $\omega_{i,t}$ jest oszacowaniem wagi i -tego Gaussianu w mieszance w czasie t (jaka część danych jest uwzględniana przez ten Gaussian),
- $\mu_{i,t}$ jest średnią wartością i -tego Gaussianu w mieszance w czasie t ,
- $\Sigma_{i,t}$ jest macierzą kowariancji i -tego Gaussianu w mieszance w czasie t ,
- η jest funkcją gęstości prawdopodobieństwa Gaussa:

$$\eta(X_t, \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(X_t - \mu)^T \Sigma^{-1} (X_t - \mu)}$$

C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," Proceedings, 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1999, pp. 246-252 Vol. 2.

40

GMM: algorytm

- Wartość K zależy od dostępnej pamięci i mocy obliczeniowej.
- Obecnie stosuje się wartości od 3 do 5.
- Ponadto ze względów obliczeniowych przyjmuje się, że macierz kowariancji ma postać

$$\Sigma_{k,t} = \sigma_k^2 \mathbf{I}$$

Zakłada się, że wartości R, G i B piksela są niezależne i mają te same wariancje. *Chociaż z pewnością tak nie jest, założenie pozwala nam uniknąć kosztownego odwracania macierzy kosztem pewnej dokładności*

C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," Proceedings, 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1999, pp. 246-252 Vol. 2.

41

GMM: algorytm

Dla usprawnienia obliczeń dokonuje się przybliżenia on-line za pomocą K -średnich.

Każda nowa wartość piksela, X_t , jest sprawdzana względem istniejących K rozkładów Gaussa, aż do znalezienia dopasowania.

Dopasowanie jest zdefiniowane jako wartość piksela w obrębie 2,5 standardowych odchylen rozkładu.

Określa się indywidualne progi dla każdego rozkładu, ponieważ jednolity próg często powoduje znikanie obiektów, gdy wchodzi one w zacienione obszary.

Jeśli żaden z K rozkładów nie odpowiada bieżącej wartości piksela, rozkład najmniej prawdopodobny jest zastępowany rozkładem o wartości średniej równej wartości danego piksela, początkowej dużej wariancji i niskiej wadze.

C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," Proceedings, 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1999, pp. 246-252 Vol. 2.

42

**GMM:
algorytm**

Wcześniejsze wagi rozkładów K w czasie $t \rightarrow \omega_{k,t}$ są dostosowywane w następujący sposób:

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

gdzie α jest współczynnikiem uczenia a $M_{k,t}$ wynosi 1 dla modelu, który pasował i 0 dla pozostałych modeli.

Po tym przybliżeniu wagi są ponownie znormalizowane. $1/\alpha$ definiuje stałą czasową, która określa szybkość, z jaką zmieniają się parametry rozkładu.

$\omega_{k,t}$ jest w rzeczywistości filtrowaną wartością prawdopodobieństwa, że wartości pikseli odpowiadają modelowi k , biorąc pod uwagę obserwacje z czasu od 1 do t .

C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 246-252 Vol. 2.

43

**GMM:
algorytm**

Parametry μ i σ dla niedopasowanych rozkładów pozostają takie same. Parametry rozkładu, które pasują do nowej obserwacji, są aktualizowane w następujący sposób:

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

$$\sigma_t^2 = (1 - \rho)\sigma_{t-1}^2 + \rho(X_t - \mu_t)^T(X_t - \mu_t)$$

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

C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 246-252 Vol. 2.

44

**GMM:
algorytm**

Rozkłady Gaussa są posortowane według wartości ω/σ . Wartość ta rośnie zarówno w miarę, jak rozkład zyskuje większe znaczenie, oraz ze spadkiem wariancji.

Po ponownym oszacowaniu parametrów mieszaniny wystarczy posortować dopasowane rozkłady od najbardziej prawdopodobnego.

Ta kolejność modelu jest uporządkowaną listą, w której najbardziej prawdopodobne rozkłady tła pozostają na górze, a mniej prawdopodobne rozkłady tła przesuwają się w dół i ostatecznie zostają zastąpione nowymi rozkładami.

Następnie pierwsze rozkłady B są wybierane jako model tła, gdzie

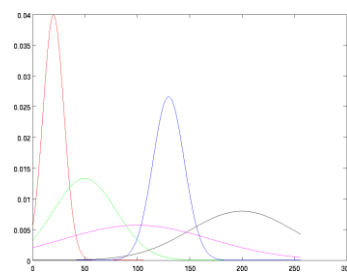
$$B = \operatorname{argmin}_b \left(\sum_{k=1}^b \omega_k > T \right)$$

C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 246-252 Vol. 2.

45

**GMM:
algorytm**

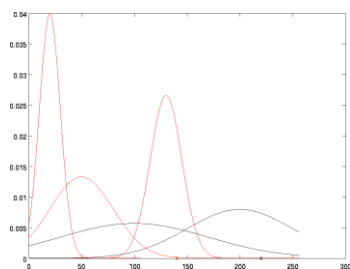
- Jeśli założymy obraz w odcieniach szarości i $K=5$, historia pojedynczego piksela może wyglądać tak:



46

**GMM:
algorytm**

- Po estymacji modelu tła, **czerwone** rozkłady należą do modelu tła, a **czarne** – obiektu pierwszoplanowego



47

**GMM:
algorytm**

- Jeśli T jest małe, model tła jest zwykle jednomodowy
- Wielomodowy rozkład tła oznacza, że może zawierać on kilka różnych kolorów i w ten sposób prawidłowo reagować na cykliczne zmiany w tle
- Jeśli bieżący kolor piksela nie pasuje do żadnej z pierwszych B rozkładów, to jest uznawany za **obiekt ruchomy**

48

Usuwanie cieni

Cień rzucany przez poruszające się obiekty może zakłócać lokalizację i wykrywanie. Używa się więc przestrzeni barw HSV i założenia, że rzucony cień przyciemnia piksele (zmniejsza ich luminancję), podczas gdy chrominancja obszarów w cieniu i dobrze oświetlonych nie różni się znacznie:

$$SP(x, y) = \begin{cases} 1 & \text{if } \alpha \leq \frac{I^V(x, y)}{I^V(x, y)} \leq \beta \wedge (I^S(x, y) - B^S(x, y)) \leq \tau_S \\ & \wedge |I^H(x, y) - B^H(x, y)| \leq \tau_H \\ 0 & \text{otherwise} \end{cases}$$

gdzie SP jest nową binarną maską bloba na pierwszym planie; $\alpha, \beta, \tau_H, \tau_S$ to parametry dobrane empirycznie, I to bieżąca ramka, B to obraz tła, a indeksy górne H, S i V wskazują, który składnik piksela HSV powinien zostać użyty, x, y są współrzędnymi piksela na obrazie.

Cucchiara, R., Grana, C., Piccardi, M., Prati, A., Sisti, S.: Improving shadow suppression in moving object detection with huecolor information. IEEE Intelligent Transportation Systems, 334-339 (2001)

49

Usuwanie cieni



C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1999, pp. 246-252 Vol. 2.

50

RPCA: algorytm

Metoda polega na przekształceniu klatek wideo w wektory i połączeniu wektorów w jedną dużą macierz.

Wynika z warunków, że B jest macierzą tła, która ma niski rząd (rank), a F reprezentuje macierz pierwszego planu, która jest rzadka. Połączenie tych dwóch macierzy daje oryginalną klatkę.

$$\operatorname{argmin}_{B, F} \operatorname{rank}(B) + \gamma \|F\|_1$$

przy założeniu, że $D = B + F$

F : macierz pierwszego planu
 B : macierz tła
 D : klatka wejściowa
 $\| \cdot \|_1$: norma l_1
 γ : stała



John Wright, Yigang Peng, Yi Ma, Anind Ganesh, Shankar Rao Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Matrices by Convex Optimization NIPS 2009

51

RPCA: algorytm

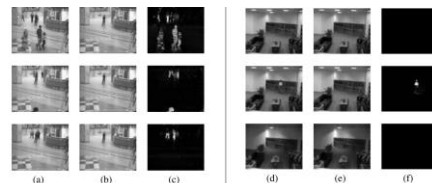


Figure 2: Background modeling. (a) Video sequence of a scene in an airport. The size of each frame is 72×88 pixels, and a total of 200 frames were used. (b) Static background recovered by our algorithm. (c) Sparse error recovered by our algorithm represents activity in the frame. (d) Video sequence of a lobby scene with changing illumination. The size of each frame is 64×80 pixels, and a total of 550 frames were used. (e) Static background recovered by our algorithm. (f) Sparse error. The background is correctly recovered even when the illumination in the room changes drastically in the frame on the last row.

John Wright, Yigang Peng, Yi Ma, Anind Ganesh, Shankar Rao Robust Principal Component Analysis: Exact Recovery of Corrupted Low-Rank Matrices by Convex Optimization NIPS 2009

52

Porównanie

Metoda	Szybkość	pamięć	Dokładność
Running Gaussian Average	1	1	L-M
Temporal Median Filter	n_s	n_s	L-M
Mixture of Gaussians	m	m	H
Kernel Density Estimation	n	n	H
Sequential KD Approximation	$m + 1$	m	M-H
Co-occurrence of Image Variance	$8n/N^2$	nK/N^2	M
Eigenbackgrounds	M	n	M

MoG – m = no. of gaussian distributions used (3-5)
 KDE – n is typically as high as 100
 SKDA – m = no. of modes of approximated pdf
 COIV – n = nearest neighbours, N^2 = spreads the cost over pixels
 EBG – M = no. of eigenvectors

53