

動的背景予測モデルによる照明条件変動下での物体検出

島田 敬士[†] 田中 達也[‡] 有田 大作^{*†} 谷口 倫一郎[†]

[†]九州大学大学院システム情報科学研究院

[‡]九州大学大学院システム情報科学府

〒819-0395 福岡県福岡市西区元岡 744 番地

*財団法人九州システム情報技術研究所

〒814-0001 福岡県福岡市早良区百道浜 2 丁目 1-22

E-mail: {atsushi,tatsuya,rin}@limu.is.kyushu-u.ac.jp, arita@isit.or.jp

あらまし 動的背景モデルは、背景画素値の変動を確率的にモデル化することで、移動物体領域を頑健に検出することができる。屋外映像観測のために動的背景モデルを利用する場合、従来型のモデルは、背景画素値の微小変動に対しては頑健に背景推定を行うことができるため、映像内に出現する移動物体を検出することが可能である。しかし、背景画素値が変動する理由には、天候の変化により徐々に照明条件が変化する場合が想定される。この場合、モデルのパラメータ更新が追いつかずに、多くのノイズが検出されてしまう傾向にある。そこで、背景画素値が徐々に変化するという条件下では、画素値の変化を予測し、予測値に基づいた動的背景モデルを構築することで、より頑健に移動物体を検出できることを示す。

キーワード 動的背景モデル, 動的背景予測モデル, 照明条件の変動

Object Detection Based on Gaussian Mixture Predictive Background Model under Varying Illumination

Atsushi Shimada[†], Tatsuya Tanaka[‡], Daisaku Arita^{*†}, and Rin-ichiro Taniguchi[†]

^{†‡} Department of Intelligent Systems, Kyushu University

744, Motoooka, Nishi-ku, Fukuoka, 819-0395 Japan

*Institute of Systems & Information Technologies/KYUSHU

2-1-22, Momochihama, Sawara-ku, Fukuoka, 814-0001 Japan

E-mail: {atsushi,tatsuya,rin}@limu.is.kyushu-u.ac.jp, arita@isit.or.jp

Abstract We propose a new method to create adaptive background models. Traditionally, each pixel has an adaptive background model which consists of Gaussian mixtures. Each model can approximate a change of pixel values and it helps us to detect moving objects which appears in the image sequences. However, traditional adaptive background models cannot adapt to some illumination changes such as gradually varying illumination, precipitously varying illumination and so on. In our approach, therefore, we use a predictive background model which can predict a pixel value in the next frame. To detect moving object, each pixel value is judged as “foreground” or “background” by the predictive background model as well as the traditional adaptive background model.

Key words Adaptive Background Model, Predictive Background Model, Varying Illumination

1. はじめに

近年、背景差分法を用いた物体検出の研究が数多く行われている。事前に用意した背景画像と観測画像との差分を計算することにより、物体に関する事前知識を必要とせずに物体領域を簡単に抽出することができるという利点がある。しかし、屋外などを観察する場合には、木々や水面の揺らぎ

などの微小な変化や、天候による照明条件の変化による影響を直に受けるために、単純な背景画像を用意しただけでは、対象物体以外に様々なノイズが前景として抽出されてしまう。このような背景の変化に柔軟に対応するための様々な背景のモデル化手法が提案されている。

背景モデルの構築手法は、事前に背景モデルを

構築しておく方法と背景モデルを背景の変化に応じて更新していく方法の2つに大別できる。前者は、想定される背景の変化を事前に学習しモデル化しておくことで、急激な背景変化にも対応できる方法である [1, 2]。これらの手法に共通する問題は、事前に用意した背景画像を用いているため、想定外の背景変化が起こった場合に、対応することができないということである。

一方で後者は、背景の変化を動的にモデル化することで、前者の場合のような想定外の背景変化には対応できないという問題が起こらないという利点があり、混合ガウス分布を利用して背景モデルを構築する手法がよく利用される。Staufferらは、分布数を固定して混合ガウス分布を高速に推定する方法を提案している [3]。この方法は、すべての画素について同数のガウス分布を用意し、ある基準を満たす分布を背景モデルとして構築している。パラメータを更新する分布についても、観測画素値にマッチする分布のパラメータのみを更新しているため、高速なパラメータ更新が可能である。しかし、ガウス分布の数を固定しているため、照明変化が小さい画素などに対しては分布数が多すぎたり、また急激な照明変化に対しては分布数が不足し、背景の変化に対応するまでに時間がかかってしまい、柔軟に対応することができない。これに対して、画素単位で分布の数を動的に変化させる方法 [4] や、その高速化手法 [5] が提案されている。一方で、ノンパラメトリックな動的背景モデル構築手法も提案されている。Parzen 推定を利用した背景モデル構築手法 [6] では、背景モデルの推定に要する時間が、混合ガウス分布を利用する場合と比べて安定するという利点があるが、計算時間がかかったり、高速化をするとき多量のメモリを要するという問題がある [7]。

上述の動的背景モデルの特徴は、画素値の微小変動に対しては、過去に観測された画素値の統計的情報を基に、その変動を背景画素値によるもの、あるいは映像内に出現した物体領域の一部であるという判定を行うという点である。過去に多くの事象があるものを背景とするという考え方に基けば、この様な背景推定法は妥当であるが、実際に観測画像に起こり得る画素値の変化には、次に挙げる場合が考えられる。

- (1) 微小に画素値が変化する場合
- (2) 急激に画素値が変化する場合
- (3) 徐々に画素値が変化する場合

微小に画素値が変化するのは、草木の揺らぎや、照明条件の微小変動による場合などであり、上述の

動的背景モデルで対応することが可能である。一方、急激に画素値が変化するのは、移動物体の通過時に起こることがひとつ考えられる。移動物体によって引き起こされる画素値の急激な変化を検出できるのは、動的背景モデルが微小変化のみをモデル化しているためである。しかし、急激な変化には、屋内照明のスイッチのオンオフによる変化なども想定でき、この場合にも従来の動的背景モデルでは移動物体領域として誤検出する傾向にある。この問題の解決法としては、急激な背景変化を想定した事前モデルを構築し、動的背景モデルと併用するという方法が考えられるが、本稿では取り扱わない。最後に、徐々に画素値が変化するのは、天候の変化により、照明条件が徐々に明るくなったり暗くなったりする場合である。上述の動的背景モデルでは、変化の度合いが低い場合は、学習率を上げるなどの調整を行うことで対応可能であるが、変化の度合いが大きくなると、モデルパラメータの更新が追いつかず、結果として照明条件による変化を移動体領域として誤検出してしまふ。そこで本稿では、このように徐々に照明条件が変化する場合への対処法を提案する。徐々に画素値が変化する場合、それまでの観測画素値を基に、次に観測されるであろう画素値を予測することが可能である。予測値を利用して、現在観測された画素値が背景画素値の変化に伴うものであるかどうかを判定することができるため、従来の動的背景モデルの更新が追いつかないという状況は起こりにくくなる。しかし、あくまで予測が正しいという条件下で背景画素値の変化かどうかを推定するものであるため、予測に失敗してしまうと、従来通り誤検出してしまうことが予想される。そこで、本稿ではいくつかの予測モデルを利用して、照明条件の変動が頻繁に起こる屋外映像を利用して、実証的実験とともに各予測モデルの効果について述べる。

2. 混合ガウス分布による背景モデル推定

本節では、混合ガウス分布を利用した背景モデル構築法 [5] について述べる。画素ごとに背景モデルを構築するため、以下では特に断りがない限りひとつの画素 $\{x, y\}$ に焦点を絞って述べることにする。ここで、時刻 t における画素 $\{x, y\}$ の画素値を X_t で表すことにすると、時刻 t までに得られた画素値 $\{X_1, \dots, X_t\}$ は K 個のガウス分布を用いてモデル化することができる。このとき、現在の画素 X_t が持つ確率は、

$$P(X_t) = \sum_{k=1}^K w_{k,t} \eta(X_t | \mu_{k,t}, \Sigma_{k,t}) \quad (1)$$

となる。ここで、 $w_{i,t}$ は時刻 t において i 番目のガウス分布を持つ重み、 $\mu_{i,t}$ は時刻 t における i 番目のガウス分布の平均値、 $\Sigma_{i,t}$ は時刻 t における i 番目のガウス分布の分散共分散行列である。また η は式 (2) で表されるガウス分布の密度関数である。

$$\eta(X_t, \mu_t, \Sigma_t) = \frac{1}{(2\pi)^{\frac{3}{2}} |\Sigma_t|^{\frac{1}{2}}} e^{-\frac{1}{2}(X_t - \mu_t)^T \Sigma_t^{-1} (X_t - \mu_t)} \quad (2)$$

K はガウス分布の数で、利用可能なメモリ領域に応じてあらかじめ設定する必要がある。通常 3 ~ 5 程度が適当である。計算コストを削減するために、分散共分散行列を次式で置き換える。

$$\Sigma_{k,t} = \sigma_{k,t}^2 \mathbf{I} \quad (3)$$

この式は、例えば X_t が RGB 値で構成される場合、各色成分は独立で同じ大きさの分散を持つことを意味する。各軸の分散を同じにして、さらに共分散成分を無視しているため精密さには欠けるが、逆行列の計算などの複雑な計算を回避できるため高速にガウス分布の計算を行うことができる。

次に、背景モデルの推定方法について述べる。

Step 1 新たに観測された画素値 X_t に対して、 K 個の分布の中からマッチする分布を探す。画素値 X_t が各分布の平均値から標準偏差 2.5 以内であれば、その分布にマッチしたとみなす。

Step 2 Step1 でマッチする分布が見つかった場合、その分布が背景モデル（背景モデルの定義は後述の Step7 を参照）であればその画素を背景とし、そうでなければ前景とする。

Step 3 K 個のガウス分布の重み $w_{i,t}$ を次式で更新する。

$$w_{k,t} = (1 - \alpha)w_{k,t-1} + \alpha M_{k,t} \quad (4)$$

ここで、 α は学習率であり、この値が大きいと新たな観測画素値 X_t にマッチする分布の重みがすぐに大きくなるので新たな背景モデルが比較的早い段階で構築される。また、 $M_{k,t}$ はマッチした分布については 1、それ以外の分布では 0 とする。各分布の重みを更新した後、重みの総和が 1 になるように正規化する。

Step 4 Step1 でマッチする分布が見つかった場合、その分布の平均値と分散値を式 (5) と (6) で更新する。また、マッチしない分布の平均値 μ および分散値 σ については更新を行わない。

$$\mu_t = (1 - \rho)\mu_{t-1} + \rho X_t \quad (5)$$

$$\sigma_t^2 = (1 - \rho)\sigma_{t-1}^2 + \rho(X_t - \mu_t)^T (X_t - \mu_t) \quad (6)$$

ここで、 ρ は学習率であり、次式で表される。

$$\rho = \alpha \eta(X_t | \mu_t, \sigma_t) \quad (7)$$

Step 5 Step1 でマッチする分布がひとつも見つからなかった場合、新たにガウス分布 η_{k+1} を作成する。ここで、

$$w_{k+1,t} = W \mu_{k+1,t} = X_t \sigma_{k+1,t} = \sigma_{k,t} \quad (8)$$

とし、 W は新たな分布に対してあらかじめ設定しておく重みである。 W の値が大きいほど、その分布は長い間背景モデルとして登録されることになる。最後に新たな分布の重みを含めた全ガウス分布の重みの総和が 1 になるように正規化する。

Step 6 照合度が最も小さいガウス分布の重みがある閾値以下である場合、その分布を削除する。削除後、残りの分布が持つ重みの総和が 1 になるように正規化する。また、2 つのガウス分布 η_a と η_b の平均値の差が各成分において閾値以下であれば分布を統合する。統合後のガウス分布を η_c とすると、

$$w_{c,t} = w_{a,t} + w_{b,t} \quad (9)$$

$$\mu_{c,t} = \frac{w_{a,t}\mu_{a,t} + w_{b,t}\mu_{b,t}}{w_{a,t} + w_{b,t}} \quad (10)$$

$$\sigma_{c,t} = \frac{w_{a,t}\sigma_{a,t} + w_{b,t}\sigma_{b,t}}{w_{a,t} + w_{b,t}} \quad (11)$$

によって、重み、平均値、分散値をそれぞれ計算する。

Step 7 ガウス分布を照合度 w/σ^2 の大きい順に並べ替える。照合度が高いほど頻繁に背景モデルとして利用されていることになる。

Step 8 次式を満足する B 個の分布を新たな背景モデルとする。

$$B = \operatorname{argmin}_b \left(\sum_{k=1}^b w_{k,t} > T \right) \quad (12)$$

この式は、ガウス分布の重みを照合度の高い順に足し合わせていき、その重みの合計が閾値 T を初めて越えるときの分布までを背景モデルとして利用することを意味している。 T が小さいと、背景モデルは単一分布で表現されることになる。一方、 T が大きいと混合分布で背景モデルが表現されることになる。

3. 動的背景予測モデル

3.1 カルマンフィルタによる動的背景予測モデル

カルマンフィルタ [8] は、離散的な観測値から時々刻々と変化する量を推定するために利用される。カルマンフィルタは予測と更新の 2 ステップから構成され、予測の手続きでは前時刻の推定状態から現時刻の状態を推定する。更新手続きでは現時刻の観測に基づいて推定値を補正してより正確な状態を推定する。以下に、予測と更新式を掲載する。

予測手続き

$$\hat{x}_{t|t-1} = F_t \hat{x}_{t-1|t-1} + B_t u_t + w_t \quad (13)$$

$$P_{t|t-1} = F_t P_{t-1|t-1} F_t^T + Q_t \quad (14)$$

更新手続き

$$\tilde{y}_t = z_t - H_t \hat{x}_{t|t-1} \quad (15)$$

$$S_t = H_t P_{t|t-1} H_t^T + R_t \quad (16)$$

$$K_t = P_{t|t-1} H_t^T S_t^{-1} \quad (17)$$

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t \tilde{y}_t \quad (18)$$

$$P_{t|t} = (I - K_t H_t) P_{t|t-1} \quad (19)$$

ここで、 $\hat{x}_{t|t}$ は時刻 t における状態の推定量、 F_t は状態遷移モデル、 B_t は制御入力 u_t に対する制御入力モデルである。また、 w_t は共分散行列 Q_t の多変数正規分布に従う時間遷移に関する雑音、 $P_{t|t}$ は現時刻の誤差行列である。 z_t は観測量と呼ばれ、状態 $\hat{x}_{t|t-1}$ を状態空間から観測空間に写像する観測モデル H_t と観測雑音 v_t を利用して、以下の関係式（観測方程式）が成り立つ。

$$z_t = H_t x_t + v_t \quad (20)$$

また、 \tilde{y}_t は観測残差、 S_t は残差共分散、 K_t はカルマンゲインと呼ばれる。

動的背景予測モデルは、従来の動的背景モデルとは別に各画素に持たせる。カルマンフィルタは (R, G, B, V) の 4 次元ベクトルを状態 (V は速度) とし、観測は (R, G, B) の 3 次元ベクトル (画素値) とし、以下の手順で予測モデルを構築する。

Step K-1 現在の観測画素値に対して、カルマンフィルタの状態を更新する。

Step K-2 カルマンフィルタで次時刻の状態を予測する。

Step K-3 予測画素値を利用して、背景予測モデルを更新する。

このように、動的背景予測モデルは、カルマンフィルタの予測手続きで得られた画素値のみを利用して構築・更新されるモデルである。なお、背景予測モデルの構築・更新方法は 2 節で述べた動的背景モデルと同様である。すなわち、動的背景モデルでは、その入力が現在の観測画素値であるのに対して、動的背景予測モデルでは、予測値が入力となる点が異なるだけである。

3.2 簡易予測モデルによる動的背景予測モデル

予測モデルにカルマンフィルタを利用する場合、計算コストが増大し、背景モデルの構築を実時間でやることは難しくなる。そこで本節では、予測手続きを簡略化した簡易予測モデルによる動的背景予測モデル構築法について述べる。

まず、連続するフレーム間の画素値の変化量 v_t を次のように定義する。

$$v_t = z_t - z_{t-1} \quad (21)$$

ここで、 z_t はカルマンフィルタの説明時に利用した観測量である。簡易予測モデルでは、過去に観測された数フレーム分（ここでは L フレームとする）の変化量を利用して、時刻 t における状態の推定量を次のように計算する。

$$\hat{x}_{t|t-1} = z_{t-1} + \frac{1}{L} \sum_{l=1}^L v_{t-l} \quad (22)$$

動的背景予測モデルの構築は次のように行う。

Step S-1 現在の観測画素値と過去の画素値の変化量を基に次時刻の状態を予測する。

Step S-2 予測画素値を利用して、背景予測モデルを更新する。

3.3 動的背景モデルと動的背景予測モデルによる背景推定法

ここでは、上記で紹介した動的背景予測モデルと従来の動的背景モデルを組み合わせた背景画素値の推定法について述べる。まず、処理の流れは図 1 のようになる。従来の背景推定法と並列に予測モデルによる背景推定を行い、各モデルでの前景背景の判定結果を表 1 に基づいて最終判定を行う。

4. 実験

実験に利用した画像は、PETS2001¹の屋外撮影画像で、320 × 240 の画像サイズに変換して利用した。解析した画像は、約 5000 フレームで構成され、後述の適合率ならびに再現率は、500 フレーム目ま

¹Benchmark data of International Workshop on Performance Evaluation of Tracking and Surveillance. From <http://pets.rdg.ac.uk/PETS2001/> available.

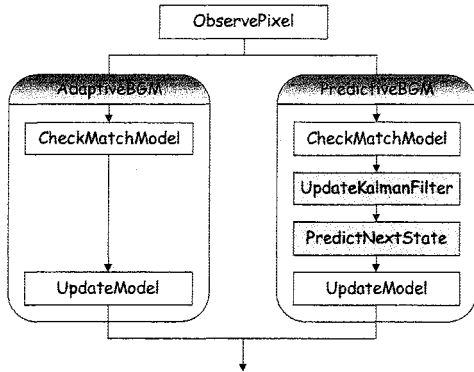


図 1: 動的背景予測モデルの処理の流れ

表 1: 前景背景の最終判定

動的背景モデル	背景予測モデル	最終判定
前景	前景	前景
前景	背景	背景
背景	前景	背景
背景	背景	背景

での解析が終了後に計測を開始した。これは、最初の 500 フレームでモデルの初期化を充分に行わせるためである。また、実験に利用した計算機は、Xeon 2.0GHz の CPU が 2 基と 4GB のメモリが搭載されている。

背景モデルを構築する際のパラメータは、学習率 0.3、分布の重みの初期値 0.3 とした。これらの値は、動的背景モデル、動的背景予測モデル、比較手法すべてで共通とした。また、簡易予測モデルにおいては $L = 10$ とした。

図 2 に照明条件が徐々に変化しているフレームでの移動物体領域検出結果例を示す。同図 (b) の ground truth は、PETS2001 の画像を基に我々が独自に作成したもの²であり、各フレームで移動物体領域として抽出すべき領域を示している。その領域には、移動物体はもちろんのこと移動物体によって作られる影領域も含まれる。また、照明条件の変化に影響されて現れる影領域や明るさが変化する領域については、抽出すべきでない領域としている。従って、移動物体のみによって変化する領域を抽出することが本実験の目的である。Stauffer らの手法で、分布数を 3 に固定した場合は、照明条件

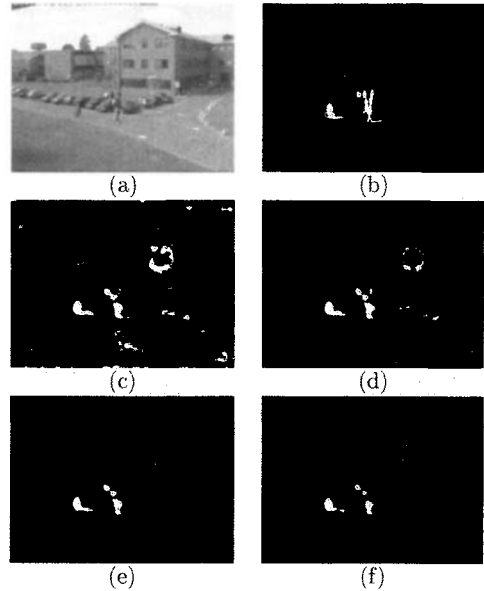


図 2: 照明条件が変化しているフレームにおける移動物体検出結果: (a) 観測画像, (b) ground truth, (c) Stauffer らの手法 (分布数 3 に固定), (d) 予測モデルなし (分布数可変), (e) カルマンフィルタによる予測あり, (f) 簡易予測モデルによる予測あり。

の変化にモデルパラメータの更新が追い付かず、建物の壁の一部も前景画素として検出されている (図 2(c))。また、それ以外にも複数のノイズが検出されている。動的背景予測モデルを使わずに、2 節で述べた分布数可変型の動的背景モデルのみを利用した場合も、ノイズの数は減っているものの建物領域が前景として検出されている (図 2(d))。一方で、カルマンフィルタあるいは簡易予測モデルを利用した場合は、建物領域が前景領域として検出されることはなく、さらにノイズも軽減されている (図 2(e), (f))。この画像に対する適合率・再現率とすべてのフレームの画像を解析後の適合率・再現率を表 2 に示す。適合率・再現率は次の式により計算される。

$$\text{適合率 (\%)} = \frac{\text{正しく抽出した画素数}}{\text{検出した画素数}} \times 100 \quad (23)$$

$$\text{再現率 (\%)} = \frac{\text{正しく抽出した画素数}}{\text{抽出すべき画素数}} \times 100 \quad (24)$$

表の最左列のアルファベットは、図 2 のアルファベットと対応している。Stauffer らの方法や予測を利用しない場合には、適合率・再現率ともに低い

²この画像系列の ground truth を含め数種類を Web で公開している。http://limu.is.kyushu-u.ac.jp/dataset/

表 2: 適合率 (%), 再現率 (%), 1 フレームあたりの平均処理時間 (msec) の比較

	図 2		全フレーム		
	再現	適合	再現	適合	時間
(c)	67.7	25.0	61.3	58.2	65.5
(d)	67.3	48.5	55.7	83.7	50.6
(e)	60.5	74.4	50.7	90.1	742.1
(f)	50.9	74.4	50.4	93.5	129.7

値となっている。一方、予測を利用する場合は、予測を利用しない場合と比べて再現率がやや低下するが、適合率が大幅に向上していることがわかる。再現率が低下した要因は、移動物体を検出できた数が減ったのではなく、移動物体領域の一部の画素が予測モデルにより次フレームで背景モデルとして学習され、検出された画素数が減少したことにある。カルマンフィルタを予測モデルに利用した場合と、簡易予測モデルを利用した場合では、ほぼ同程度の結果が得られている。処理時間は、簡易予測モデルを利用した方が短時間でモデル構築が可能であるため、予測の厳密性には欠けるが簡易予測モデルの方が動的背景予測モデルには適していることがわかる。

5. おわりに

本稿では、動的背景予測モデルを従来の動的背景モデルと併用することにより、照明条件の変化により画素値が徐々に変化していく場合にも頑健に移動物体を抽出する方法を提案した。観測画素値の変化を基に、次のフレームで観測され得る画素値を予測し、その予測値を動的背景予測モデルとして構築することで、従来法ではモデルパラメータの更新が追い付かず照明変動による影響を受けた画素を誤検出してしまいう問題を解決できた。しかし、動的背景予測モデルと動的背景モデルを併用しているために、従来用よりもモデル構築に要する時間が増大してしまうという問題が浮き彫りになった。今後の課題として、動的背景予測モデルと動的背景モデルを単純に併用するのではなく、併合するような仕組みを考える必要がある。

謝辞

本研究の一部は、財団法人服部報公会 2007 年度工学研究奨励援助金「課題名：長期記憶と短期記憶による背景モデルの時空間統合に関する研究」の助成を受けたものである。

参考文献

- [1] 浮田宗伯, 加藤文和. “背景差分と色検出の統合によるターゲット色の自動学習と背景変動に頑健な実時間対象検出”. 画像の認識・理解シンポジウム (MIRU2004), Vol. II, pp. 24-29, 2004.
- [2] 川端聡, 日浦慎作, 佐藤宏介. “動的背景の学習による未知物体領域の実時間抽出法”. 電子情報通信学会論文誌, Vol. J89-D, No. 4, pp. 826-835, 2006.
- [3] Chris Stauffer and W.E.L. Grimson. “Adaptive background mixture models for real-time tracking”. *Computer Vision and Pattern Recognition*, Vol. 2, pp. 246-252, 1999.
- [4] Jian Cheng, Jie Yang, and Yue Zhou. “A Novel Adaptive Gaussian Mixture Model for Background Subtraction”. *2nd Iberian Conference on Pattern Recognition and Image Analysis*, pp. 587-593, 2005.
- [5] 島田敬士, 有田大作, 谷口倫一郎. 適応的な分布数の増減法を利用した混合ガウス分布による高速な動的背景モデル構築. 電子情報通信学会論文誌 D, Vol. J90-D, No. 9, pp. 2606-2614, 09 2007.
- [6] Ahmed Elgammal, Ramani Duraiswami, David Harwood, and Larry Davis. “Background and Foreground Modeling Using Non-parametric Kernel Density Estimation for Visual Surveillance”. *Proceedings of the IEEE*, Vol. 90, pp. 1151-1163, 2002.
- [7] Tatsuya Tanaka, Atsushi Shimada, Daisaku Arita, and Rin-ichiro Taniguchi. A Fast Algorithm for Adaptive Background Model Construction Using Parzen Density Estimation. In *CD-ROM Proc. of IEEE International Conference on Advanced Video and Signal based Surveillance*, 09 2007.
- [8] Kalman R. E. “A New Approach to Linear Filtering and Prediction Problems”. *Transactions of the ASME - Journal of Basic Engineering*, Vol. 82, pp. 35-45, 1960.