

ピクセル状態分析による固定監視映像の圧縮

Fixed Camera Movie Compression Using Pixel State Analysis

西 貴行† 藤吉 弘亘‡ 梅崎 太造†

Takayuki Nishi†, Hironobu Fujiyoshi‡, Taizo Umezaki†

† 名古屋工業大学 ‡ 中部大学

†Nagoya Institute of Technology, ‡Chubu University

E-mail: hf@cs.chubu.ac.jp

Abstract

従来、監視映像の記録には一定間隔に撮影した画像フレームを符号化して保存するタイムラプス方式が用いられている。しかし、タイムラプス方式は監視対象となる物体が存在しないフレームも保存するため、記憶量が膨大となる問題がある。一方、オブジェクトベース符号化は画像から人物等の重要な前景領域を抽出し、背景と区別して情報を扱うので映像を高効率圧縮することが可能である。本稿では、ピクセル状態分析を用いたオブジェクトベース符号化について提案する。本手法は、ピクセル状態分析による物体検出により画像から前景と背景を判別する。さらに、前景である物体領域の各ピクセルを静状態もしくは動状態に判別する。静状態ピクセルは、1フレーム前の同じ位置における輝度値を配置することにより復元可能であるため、各フレームにおいて輝度を保存する必要がない。これにより、記憶量の減少が期待できる。さらに、前景の動ピクセルはLZHで圧縮する。LZHは可逆圧縮なので、圧縮による劣化を受けない。本手法の有効性を評価するために、9つの映像においてMPEG-4と本手法の圧縮率の比較実験を行い、MPEG-4よりも圧縮率が15%高く、移動領域の画質が良いという結果を得た。

1 はじめに

従来、監視映像の記録には一定間隔に撮影した画像フレームを符号化して保存するタイムラプス方式が用いられている。しかし、タイムラプス方式は監視対象となる物体が存在しないフレームも保存するため、記憶量が膨大となる問題がある。この問題を解決するための手法として、オブジェクトベース符号化が提案されている[1][2]。オブジェクトベース符号化は、画像から人物等の重要な前景領域(オブジェクト)をより良い画質で符号化し、それ以外である背景領域を高圧縮率で符号化する。このように前景と背景に対して個別に符号化

するため、注目すべき前景であるオブジェクト領域をより良い画質のまま、映像シーケンス全体を高効率圧縮することが可能となる。Clarkらが提案するObjectVideoは、オブジェクトベース符号化における前景の抽出に背景差分を用いて、動的なオブジェクトを含む前景と背景に分離する[1]。前景領域は、MPEG-4に準拠したビットストリームにより符号化される。Vetroらの手法は、オブジェクト領域の自動抽出に背景差分と動きベクトル分析の2つの処理を利用して注目対象を符号化する[2]。動きベクトル分析では、日照変化、木の葉や枝の揺れ、水面反射の拡散等の画像中で重要でない輝度変化を除去するために、均一性や連続性のある動きを持つ動物体領域を抽出する。これにより、必要でない領域の誤検出を抑制することが可能となる。両者の手法において、前景である物体領域は、動き補償DCT符号化によるMPEG-4オブジェクトベース符号化を用いて圧縮されている。

本稿では、固定カメラから得られる映像をより高い圧縮率で符号化し、注目対象領域をより良い画質で圧縮する手法として、ピクセル状態分析を用いたオブジェクトベース符号化を提案する。本符号化におけるピクセル状態分析は、物体検出により画像から前景と背景に分離する。さらに、前景である物体領域の各ピクセルを静状態もしくは動状態に判別する。車等の移動物体においても、同一のテクスチャや色を持つため輝度値が変化していない静状態ピクセルが存在する。これらの前景に含まれる静状態ピクセルは、1フレーム前の同じ位置における輝度値を配置することにより復元することが可能であり、各フレームにおいてその輝度値を保存する必要がない。これにより、高い圧縮率が期待できる。物体領域の動ピクセルにおいては、LZH符号化により圧縮する。LZH符号化は可逆圧縮であるため、符号化による情報の劣化を防ぐことができ、前景を高画質で保存することが可能となる。

本稿では、2章で新しい手法であるピクセル状態分析

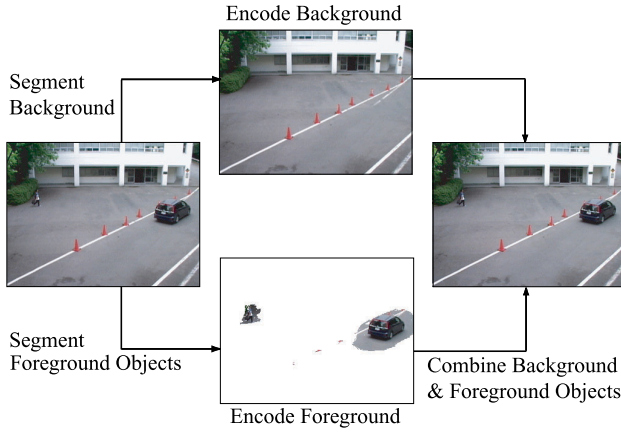


図 1 オブジェクト符号化の概念図

によるオブジェクトベース符号化について示し、3章で提案手法による監視映像の符号化と復号化のアルゴリズムについて述べる。4章では提案手法と MPEG-4 を画質と圧縮率において比較した評価実験結果を示す。

2 ピクセル状態分析によるオブジェクトベース符号化

本章では、オブジェクトベース符号化の概念とピクセル状態分析のアルゴリズムについて示す。

2.1 前景と背景による画像の表現

監視を目的とした固定カメラ画像における多くのピクセルは、監視対象ではない背景に属している。そこで、予め用意した背景と現在の入力画像の差分により、各ピクセルを前景と背景に分離する [2][3]。オブジェクトベース符号化は、図 1 に示すように分離した前景と背景を個別に符号化する。復号化の際には、背景画像上に前景領域を配置して画像を再構成する。

本稿では、オブジェクトベース符号化を基にした高効率の動画圧縮を実現するために、ピクセルの状態遷移に注目する。ピクセルは、図 2 に示す背景 (background)、静状態 (transient)、動状態 (stationary) の 3 つの状態を持つと仮定する。物体が静止したときに頻繁に現れる静状態ピクセルは、ピクセルの輝度値の更新を必要としないため、より圧縮率の高いオブジェクトベース符号化が期待できる。

2.2 ピクセル状態分析

屋外カメラで撮影した画像上のピクセルの輝度変化は、状況に応じて三つの特徴に大別できる。

- 物体がピクセル上を通過するとき、そのピクセルの輝度値は急激な変化を伴う。その後、一時的に不安定な状態が続く、再度急激な変化の後、背景である元の輝度値に戻る。

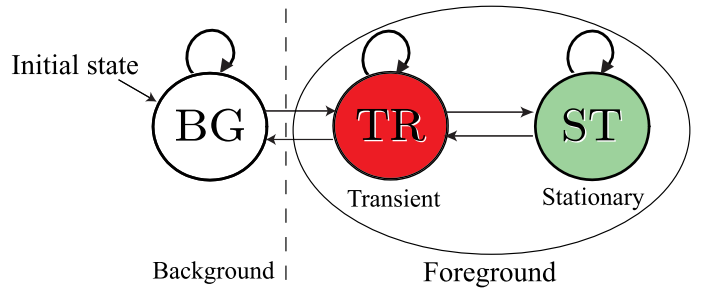


図 2 ピクセルの状態遷移図

- 物体がピクセル上で停止したとき、そのピクセルの輝度値は急激な変化の後、一時的に不安定な状態が続く、最終的には物体の輝度値に安定する。
- 太陽が雲に隠れた等の環境変化が生じたとき、輝度値は緩やかに変化する。

これらの特徴を捉えるためには、以下の二つの変化に着目する必要がある [4]。

1. 輝度値の急激な変化
2. 不安定な状態から安定した状態への遷移

まず最初に、急激な輝度変化を捉えるモーショントリガーを考える。ここで、 I_t を現在のフレームから k フレーム前の輝度値とする。輝度値の変化量 T を求めるには t よりも前 (過去) n フレームに着目し、 I_t と I_{t-j} の差分の絶対値を求める。変化量 T は次式により求められる。

$$T = \max\{|I_t - I_{(t-j)}|, \forall j \in [1, n]\} \quad (1)$$

急激な輝度値の変化がピクセル上に生じたとき、 T の値は大きくなる。次に、ピクセルの安定度を示す S について考える。安定状態の検出には t より後 (未来) k フレームに着目する。安定度 S は次式により求められる。

$$S = \frac{k \sum_{j=0}^k I_{(t+j)}^2 - (\sum_{j=0}^k I_{(t+j)})^2}{k(k-1)} \quad (2)$$

これはフレーム t から $t+k$ までの輝度値の分散であり、安定した状態のとき S は小さくなる。

ここで、各ピクセルの状態を表す M は、背景 (background) のとき bg 、動状態 (transient) のとき tr 、静状態 (stationary) のとき st をとるように、以下に示すアルゴリズムにより決定される。 I は、時刻 t における輝度値とする。 Th_{tr} は急激な変化を判定する閾値、 Th_{st} は安定性を判定する閾値である。background は背景画像として、予め用意しておく。

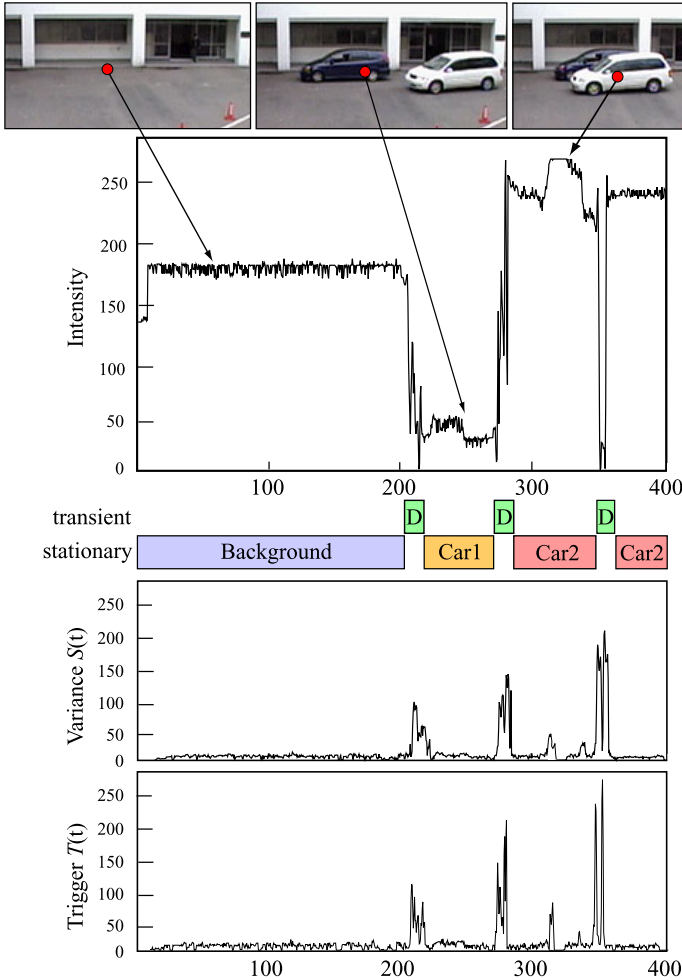


図3 ピクセル状態分析の分析例

```

While all the pixels have not been classified
When M is equal to ST or BG and
T is higher than th_t
    assign TR to M
When M is equal to TR and S is lower than
th_s
    If the intensity I is nearly background
        assign BG to M
    else
        assign ST to M
end

```

この操作を全ピクセルについて行う。時刻 t において、背景と判定されたピクセル $I(t)$ は、次式に示すIIRフィルタにより環境変化に追従するよう適応的に背景 $B(t)$ 更新する [5]。

$$B(t) = \alpha I(t) + (1 - \alpha)B(t - 1) \quad (3)$$

α は、背景と判定されたピクセルを背景画像にどれだけ反映させて更新するかを決定する定数である。

図3に、あるピクセル上に車 (Car1) が侵入後停止し、更に別の車 (Car2) が侵入、停止する場合の輝度変化と分析例を示す。図3より、輝度変化に応じて変化量 T と安定度 S が変化し、背景→動→静 (Car1) →動→静



図4 人が手を振る映像の分析例

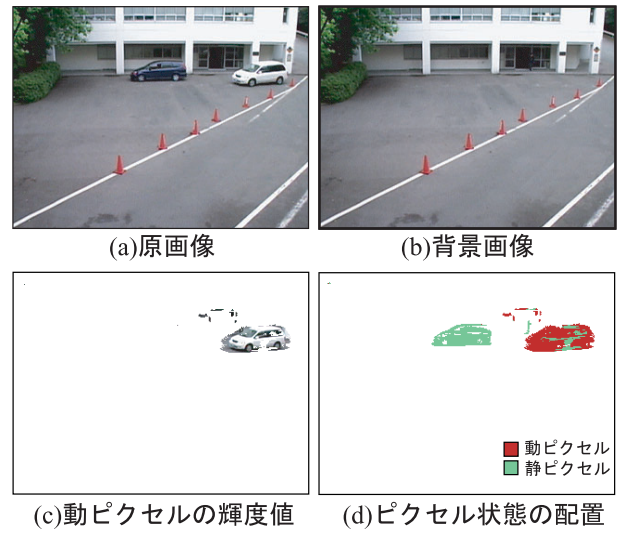


図5 ピクセル状態分析による3要素の分解

(Car2) というピクセルの状態遷移を正確に検出していることがわかる。図4に、人が手を上下に振る映像のピクセル状態分析結果を示す。移動領域である腕の部分のみ動状態と判別され、それ以外は静状態と判別されていることが分かる。

3 オブジェクトベースの符号化と復号化

本章では、提案するピクセルの状態分析によるオブジェクトベース符号化と復号化処理について示す。

3.1 符号化処理

映像の各フレームは、ピクセル状態分析により、図5に示すように背景画像、ピクセル状態の配置、動ピクセルの輝度値という3つの要素に分解される。各要素は以下に示すアルゴリズムによって各々に符号化される。

背景画像:

物体が存在しないときの初期フレームを背景画像として取得する。背景画像は JPEG (Joint Photographic Experts Group) で圧縮し保存する [6]。背景画像は8秒毎に保存する。

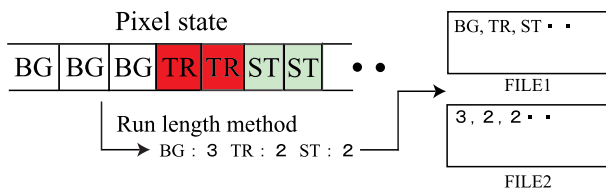


図6 ピクセル状態の配置へのランレングス法の適応

ピクセル状態の配置:

ピクセル状態分析は、各ピクセル毎に、背景、静、動ピクセル状態の配置を出力する。このピクセルの3状態は、2ビットで表現可能である。同一の状態を持つピクセルの連鎖を、図6に示すようにランレングス法により符号化する。さらに、ランレングス法で圧縮したデータをLZH(Lempel-Ziv Huffman)により圧縮を施す[7].

動ピクセルの輝度値:

ピクセル状態分析により判別された動状態ピクセルに対応する輝度値をLZHにより圧縮する。LZHは可逆圧縮なので、前景である対象物体の動ピクセルの色情報を復元することが可能である。

3.2 類似ピクセルの探索

本符号化手法では、ピクセル状態分析により動状態と判定されたピクセルの輝度値を、LZHにより圧縮を施す。大きな物体が侵入した場合や多数の物体が通過する場合、前景に動状態ピクセルが増え圧縮率が低下することが考えられる。そこで、動と判定されたピクセルに対して、図7に示すように1フレーム前の画像における近傍領域(16×16ピクセル)から類似する輝度値を持つピクセルを探索する。類似したピクセルが存在すれば、その位置を示すラベル情報(1 byte)を保存する。復号化の際は、1フレーム前の復元画像から位置ラベルが示す輝度値を用いることで、動ピクセルの輝度を復元することが可能である。これにより、本来ピクセルの更新に必要なRGB(3 byte)のデータ量を3分の1に抑えることができ、動ピクセルにおける圧縮率の低下を防ぐことが可能となる。図8は、類似ピクセルの探索例である。青色で示すピクセルは、1フレーム前の近傍領域内に類似したピクセルが存在したことを表している。図8より、移動体における動ピクセルの輝度値は同一のテクスチャや色のため、近傍領域に多くの類似した輝度値を持つことがわかる。

3.3 復号化処理

画像を復号化する際は、ピクセル状態の配置図を参照して背景画像の上に動ピクセルの色情報を配置する。静ピクセルには、復元化された1フレーム前の復元画像と同じ位置の色情報を配置する。図9に、原画像と本提案手法による復元画像を示す。

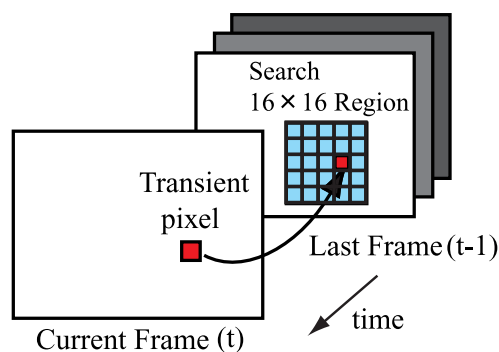


図7 類似ピクセルの探索

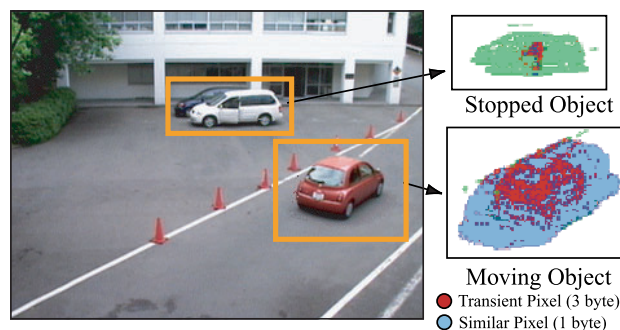


図8 移動物体における動ピクセルの詳細

4 評価実験

固定カメラで撮影された屋外監視映像を用いて、提案手法とMPEG-4の圧縮率・画質に対する評価を行う。評価用映像は、異なる屋外監視場所で撮影した画像サイズ320×240、RGB24ビットの映像である。

4.1 圧縮率の評価

表1に、9つの異なる映像を本手法とMPEG-4により圧縮したファイルサイズと、その圧縮率の割合を示すビット削減率(Bit-saving ratio)を示す。ビット削減率は以下の式により計算される。

$$\text{ビット削減率} = \frac{M - P}{M} \times 100[\%] \quad (4)$$

M はMPEG-4による圧縮後のファイルサイズ、 P は提案手法による圧縮後のファイルサイズである。

このビット削減率は、提案手法による圧縮後のファイルサイズがMPEG-4のファイルサイズより何%圧縮されているかを意味する。表1より、本手法はMPEG-4より約15%圧縮率が高いことがわかる。Seq. Cのビット削減率は20.2%、Seq. Aは39.6%であり評価映像の内容毎にその圧縮率は異なる。

表2は、ピクセル状態分析の結果、前景領域における動状態ピクセルと静状態ピクセルの割合を示す。評価用映像Seq. Cは、自転車等の比較的面積の小さな物体を多く含む映像である(図10)。表2より、Seq. Cは対象物体の面積が小さくかつ移動速度が速いため動状

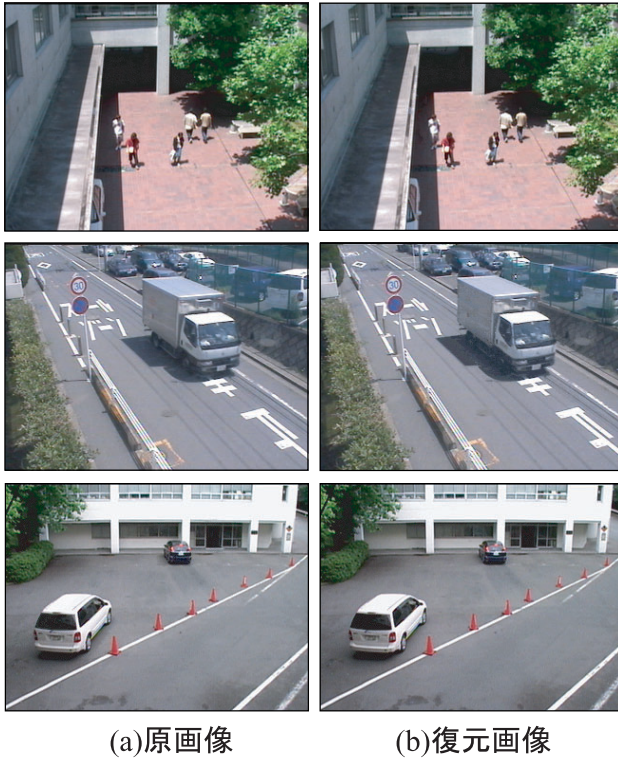


図 9 原画像と提案手法による復元画像の例

態ピクセルの割合が他のシーケンスより多いことが分かる．さらに、動ピクセルに対する類似ピクセルの割合が 61%と低いため、圧縮率の低下を招く原因となる．一方、Seq. A は類似ピクセルの割合が 76.7%と高く大幅に動ピクセルの更新の抑制が可能のため、高いビット削減率の実現が可能である．

4.2 画質の評価

画質の評価として、MPEG-4 と提案手法により復元された画像の PSNR(Peak Signal to Noise Ratio) を計算する．PSNR は以下に示す MSE(Mean Squared Error) の対数として計算される．

$$\text{PSNR} = 10 \log_{10} \frac{255^2}{\text{MSE}} [\text{dB}]$$

$$\text{MSE} = \frac{\sum_{x=0}^{W-1} \sum_{y=0}^{H-1} (f(i, j) - f'(i, j))^2}{WH}$$

W : 画像の幅 H : 画像の高さ

$f(i, j)$: 原画像 $f'(i, j)$: 復元画像

表 3 に、画像全体、物体領域と物体領域における動ピクセルにおける各 PSNR を示す．MPEG-4 と提案手法の PSNR は、画像全体と物体領域においてほぼ同じ値であるという結果を得た．しかし、物体領域の動ピクセルに注目すると、本手法の PSNR の平均値は 37.3[dB] であり、MPEG-4 より良い画質で動ピクセルの色情報を復元することが可能である．人間の視覚は移動物体

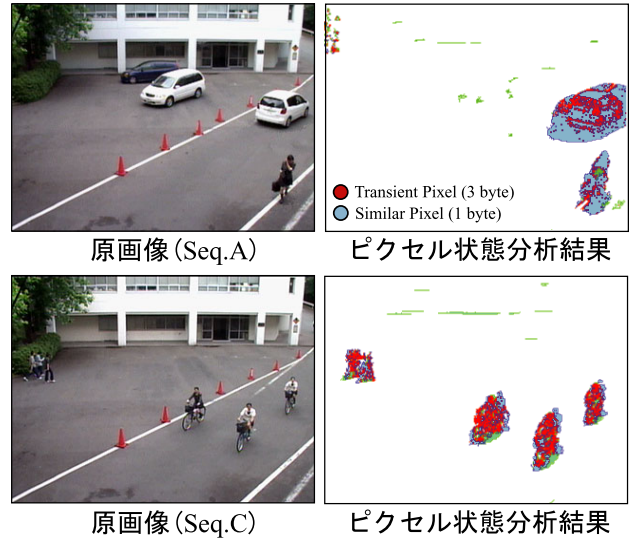


図 10 圧縮率の良い映像と悪い映像の比較

表 1 圧縮したファイルサイズの比較

Seq.	原画像 [MB]	MPEG* [MB]	提案 手法 [MB]	bit-saving ratio [%]
A	184	5.30	3.23	39.06
B	169	4.81	4.42	8.11
C	150	5.05	6.07	-20.20
D	429	12.57	8.90	29.20
E	468	15.29	12.83	16.09
F	521	16.83	14.32	14.91
G	1,930	83.70	61.44	26.59
H	3,860	104.49	81.37	22.13
I	3,860	83.93	83.73	0.24
Ave.	1,286	36.89	30.70	15.13

*MPEG-4 configuration

Key Frame: 8 sec/frame, Compression Control: 75

Bit Rate : Variable Bit Rate(Max:3,000kbps)

を注視する特徴があるため、物体領域の動いている領域をより良い画質で得られる事は監視映像の記録には有効的であると考えられる．

MPEG-4 はビットレートを下げることにより更に高い圧縮率の実現可能である．本手法では、可逆圧縮によるデータ保存を行っているが、不可逆圧縮を用いることにより本手法もまた圧縮率を高めることが可能である．今後もさまざまなシーケンスに対しての評価実験を行い、提案手法の改良を行う必要がある．

5 まとめ

本稿では、長時間の監視映像の記録のためにピクセル状態分析によるオブジェクトベース符号化を提案した．評価実験において、提案手法は MPEG-4 より約 15% 圧縮率が高く、かつ注目対象である前景オブジェクト内の

表 2 ピクセル状態分析による 3 状態の割合 [%]

Seq.	動状態 ピクセル	静状態 ピクセル	類似 ピクセル
A	94.3	5.7	76.7
B	93.5	6.5	64.2
C	96.5	3.5	61.0
D	55.9	44.1	71.7
E	55.2	44.8	61.2
F	51.8	48.2	61.0
G	81.3	18.7	47.4
H	87.2	12.8	76.9
I	78.8	22.2	71.4
Ave.	77.2	21.2	65.7

表 3 復元画像の PSNR

Seq.	MPEG[dB]			提案手法 [dB]		
	全体 画像	物体 領域	動状態 領域	全体 画像	物体 領域	動状態 領域
A	34.3	32.8	32.7	33.4	33.5	35.9
B	34.2	32.1	32.0	33.8	32.7	36.3
C	34.3	32.5	32.4	33.7	33.8	37.9
D	34.2	33.0	33.0	33.9	33.7	37.8
E	34.2	32.5	32.5	33.9	33.2	38.0
F	34.2	32.5	32.4	33.7	33.2	37.8
G	33.8	32.1	32.2	32.1	33.6	37.4
H	34.9	33.2	33.2	33.4	33.8	36.8
I	35.8	33.8	33.7	34.7	32.3	37.8
Ave.	34.4	32.7	32.7	33.6	33.3	37.3

動いている領域を MPEG-4 より約 5[dB] 高い 37.3[dB] で復元することが可能である。本提案手法は、侵入物体を高画質で記録することができるため、監視映像の記録において有効であるといえる。

今後は、ピクセル状態分析の出力を用いた長時間映像の要約のための、監視映像のキーフレーム抽出について検討する予定である。

参考文献

- [1] J. Clark, “The Power of ObjectVideo”, <http://www.objectvideo.com/>
- [2] A. Vetro, T. Haga, K. Sumi and H. Sun, “Object-based Coding for Long-term Archive of Surveillance Video”, *IEEE International Conference on Multimedia and Expo*, Jul, 2003.
- [3] C. Stauffer and W.E.L. Grimson, “Adaptive background mixture models for real-time tracking”, *Proc. Computer Vision Pattern Recognition*, Fort Collins, CO, June 1999.
- [4] H. Fujiyoshi and T. Kanade, “Layered Detection for Multiple Overlapping Objects”, *IEEE International Conference on Pattern Recognition*, vol. 4, pp.156-161, 2002.
- [5] A. Lipton, H. Fujiyoshi and R. Patil, “Moving Target Classification and Tracking From Real-time Video” *Proc. of the Workshop on Application of Computer Vision, IEEE*, Oct, 1998, pp. 8-14.
- [6] ISO/IEC JTC1 10918-1 } ITU-T Rec. T81, *Information technology - Digital Compression and coding of continuous-tone still images: Requirements and guidelines*, 1994.
- [7] M. Nelson and J. Gailly, “The Data Compression Book, Second Edition”, M&T Books, 1996.