

RK-006

# 固定カメラに対する周辺情報を考慮したズーム映像の生成 Generating Magnified View for a Fixed Camera Considering Surrounding Information

渡邊 健司†  
Kenji Watanabe

新田 直子†  
Naoko Nitta

馬場口 登†  
Noboru Babaguchi

## 1. まえがき

近年、セキュリティに対する不安の増大から、安心・安全な生活環境の実現を目的として、多くの場所で監視カメラによる映像サーベイランスが行われている。サーベイランス映像中には、監視者が特に注目すべき領域である出入口や人物といった関心領域が存在するが、一般的に商店街や街中のような広い範囲の状況を高い場所から全体的に撮影している監視カメラにおいては、関心領域が画面上に小さくしか映らないことも多い。そこで、その詳細を知るためにはズーム表示が必要である。

画像・映像に対するズーム表示手法は大別すると、フルズーム表示 [1][2]、fisheye view [3][4][5] に分けられる。前者は、ズーム対象のみを切り出して表示する手法であり、ズーム対象のみに注目することができるが、可視範囲は狭まる。他方、後者は、画面上で局所的に拡大率を変化させ、ズーム対象の周辺領域も変形し表示することにより、ズーム対象とその周辺を同時に表示する。刻々と状況が変化する映像サーベイランスにおいては、関心領域をズーム表示する際、周辺の状況も常に把握できることが望ましい。そのため、後者の手法がより有効であると考えられるが、fisheye view を用いた手法の多くは、ズーム対象からの距離のみに応じて拡大率を設定しているため、周辺領域の内容、つまり周辺情報に対する考慮は十分であるとは言えない。周辺領域の重要な部分に変形するほど、周辺情報は失われると考える。そこで、本研究では、関心領域の拡大のみでなく、周辺の重要な部分についても変形を抑えることにより、周辺情報を保持したズーム映像の生成を目指す。

一方、画像・映像の内容を考慮して情報を保持するという点に関連して、Retargeting に関する研究があげられる [6][7][8]。Retargeting とは、画像サイズと画面サイズが異なる場合に、画像を画面サイズに適応させることを意味する。L. Wolf ら [6] は、映像の内容を考慮し、携帯情報端末などの限られた画面内に、情報量をなるべく減らさず重要な部分を残して映像を表示する手法を提案している。

そこで本稿では、関心領域をズーム表示するという前提の元に、情報量をなるべく減らさないという Retargeting の考え方を取り入れることにより、周辺情報を保持したズーム映像の生成手法を提案する。

## 2. 提案手法

図1に提案手法の全体図を示す。本研究は、監視カメラ映像中の関心領域を、周辺情報を保持しながらズーム表示することを目的としている。そのため、まずズーム対象とする関心領域を抽出する必要がある。次に、同一画面上に、ズーム表示された関心領域とその周辺領域を同時に表示するためには、周辺領域の変形が必要である。

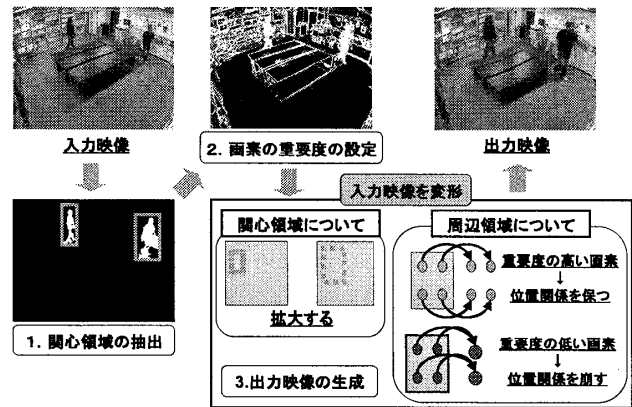


図1: 提案手法の全体図

提案手法では、周辺領域の重要な部分に変形し、周辺情報が失われることを防ぐために、画素に重要度を設定する。設定した重要度をもとに、周辺領域の重要な部分の変形を抑えることで、周辺情報を保持したズーム映像を生成する。そこで、提案手法を以下の3ステップとする。

1. 関心領域の抽出
2. 画素の重要度の設定
3. 出力映像の生成

次節から、各ステップについて詳しく述べる。

### 2.1 関心領域の抽出

監視カメラ映像中の関心領域として、出入口や人物などが考えられるが、提案手法では、固定カメラを仮定し、映像中の動物体を関心領域として抽出する。具体的には、背景の推定、前景領域の抽出、ノイズの除去からなる。

- 背景の推定  
入力映像に対してガウス混合分布による背景モデル構築法 [9][10] を用いて背景を推定する。
- 前景領域の抽出  
入力映像と推定した背景との差を前景領域として抽出する。
- ノイズの除去  
抽出された前景領域には、背景の推定誤差によるノイズが含まれている可能性があるため、2値化し、独立な領域のうち面積が一定値より大きい前景領域のみを関心領域として抽出する。

### 2.2 画素の重要度の設定

2.1節で抽出された関心領域  $R$  以外の領域を周辺領域と定義し、関心領域、周辺領域ともに全ての画素に対して重要度を設定する。図2に示す映像の場合、実線で囲まれた関心領域はズーム表示するとともに、周辺領域の

†大阪大学大学院工学研究科, Graduate School of Engineering, Osaka University

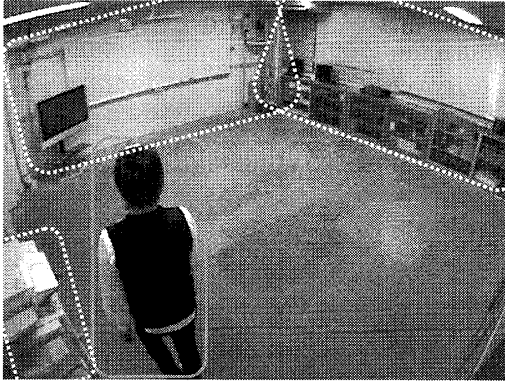


図 2: 画素の重要度の設定

中でも、変形により見づらい映像となると考えられる破線で囲まれた領域の重要度を、床面など他の領域に比べて高く設定し、重要度が高い部分ほど変形を抑えることで、周辺情報を保持した映像を生成する。画素の重要度の設定は、関心領域、周辺領域の画素についてそれぞれ以下のように行う。

- 関心領域の画素

監視カメラ映像中で監視者が特に注目すべき領域である関心領域は、重要度が高い領域であると考えられる。そこで、関心領域の画素の重要度  $K_{i,j,t}$  を、

$$K_{i,j,t} = 1 \quad ((i, j, t) \in R) \quad (1)$$

とする。

- 周辺領域の画素

周辺領域の中には、変形すると映像として見づらくなる領域、つまり重要度が高い領域と、変形しても見え方に影響が少ない領域、つまり重要度が低い領域があると考えられる。中でも特に、画素値の変化が大きい部分であるエッジが変形するほど見づらい映像となることが推察される。そこで、提案手法では、エッジに着目し、画素値の変化量、つまりエッジ量  $E_{i,j,t}$  にもとづいて周辺領域の画素の重要度  $C_{i,j,t}$  を設定する。なお、エッジ量の算出には、Sobel フィルタを使用する。

$$E_{i,j,t} = \left| \sum_{k=0}^2 \sum_{l=0}^2 I_{i,j,t} V_{k,l} \right| + \left| \sum_{k=0}^2 \sum_{l=0}^2 I_{i,j,t} T_{k,l} \right|$$

$$\text{但し } V = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, T = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$$

ここで、 $\max(E_{i,j,t})$  を  $E_{i,j,t}$  の最大値と定義すると、

$$C_{i,j,t} = \frac{E_{i,j,t}}{\max(E_{i,j,t})} \quad ((i, j, t) \notin R) \quad (2)$$

これらをまとめて、画素の重要度  $P_{i,j,t}$  を

$$P_{i,j,t} = \begin{cases} K_{i,j,t} & ((i, j, t) \in R) \\ C_{i,j,t} & ((i, j, t) \notin R) \end{cases} \quad (3)$$

とする。

### 2.3 出力映像の生成

2.1 節、2.2 節の結果をもとに、関心領域をズームし周辺情報を保持した映像を生成する。ある時刻  $t$  における入力フレームの画素を  $(i, j, t) (i = 1, \dots, L, j = 1, \dots, H, t = 1, \dots)$  とし、この画素が出力フレームの画素  $(x_{i,j,t}, y_{i,j,t}, t)$  へと対応付けられるとすると、出力映像を生成するためには、未知数  $x_{i,j,t}, y_{i,j,t}$  を求める必要がある。但し  $L, H$  はフレームサイズである。そこで、 $x_{i,j,t}, y_{i,j,t}$  に関する条件式を全ての画素について立式し、それらを合わせた連立方程式を解くことにより、出力映像を生成する。条件式の立式に関して、[6] を参考にした。 $x_{i,j,t}, y_{i,j,t}$  は独立に考えることができ、同様に計算できるため、まず、 $x_{i,j,t}$  について条件式とその重みを以下のように設定する。

#### フレームの端の画素に関する条件

- 入力フレームの端の画素はそれぞれ、出力フレームの端の画素へと対応付けられる。

$$x_{1,j,t} = 1 \quad (4)$$

$$x_{L,j,t} = L \quad (5)$$

式 (4), (5) により、可視範囲が狭まることを防ぐ。

#### 画素の空間的な連続性に関する条件

- 入力フレームで上下に隣接する画素は、出力フレームでも上下に隣接する。

$$W_s(x_{i,j,t} - x_{i,j+1,t}) = 0 \quad (6)$$

式 (6) により垂直方向の歪みを防ぐ。但し、 $W_s$  は空間的な連続性に関する条件に対する重みである。

#### 画素間の幅に関する条件

- 入力フレームで左右に隣接する画素は、 $x$  軸方向に  $\alpha$ 、または 1 の幅を持って対応付けられる。

$$S_{i,j,t}(x_{i,j,t} - x_{i-1,j,t}) = \begin{cases} \alpha \times S_{i,j,t} & ((i, j, t) \in R) \\ S_{i,j,t} & ((i, j, t) \notin R) \end{cases} \quad (7)$$

式 (7) により、関心領域をズーム表示するとともに、画素間の位置関係を保つ。但し、 $\alpha$  はズーム倍率である。また、 $S_{i,j,t}$  は、画素間の幅に関する条件に対する重みであり、提案手法では、 $S_{i,j,t} = P_{i,j,t}$  として、重要な部分の変形を抑え、周辺情報を保持した映像を生成する。

上記の条件を全てあわせた連立一次方程式は、未知数よりも方程式の数が多い、過剰条件の方程式となるため、その解は一般に存在しない。そのため、最小二乗解を求めることで、入力フレームと出力フレームの画素を対応付け、出力映像を生成する。 $y_{i,j,t}$  に関しても同様に計算する。

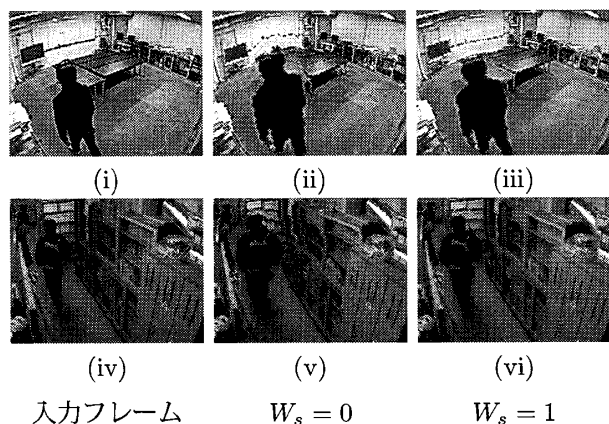


図3: 画素の空間的な連続性に関する条件を考慮した場合, 考慮しない場合の生成映像の比較

表1: 図3に示すフレームにおける関心領域の拡大率  $\eta$

映像	(i)	(ii)	(iii)	(iv)	(v)	(vi)
$\eta$	-	1.47	1.32	-	1.56	1.38

### 3. 実験

#### 3.1 実験概要

提案手法の有用性を確認するために映像生成実験を行った。実験では、室内に設置された固定カメラを使用して予め撮影した映像を入力映像として使用した。また、実験を行ったコンピュータ環境は、OS:Microsoft WindowsXP Professional SP3, CPU:Pentium (R) D 3.20 GHz, メモリ:1.99GBである。過剰条件の連立一次方程式の最小二乗解を求めるために、数値解析ソフトウェアMATLAB R2007bを使用した。出力映像のフレームサイズを  $L = 160$ ,  $H = 120$  とすると、映像の生成速度は約30秒/フレームであった。

生成された映像に対して、式(8)により定義する関心領域の拡大率  $\eta$  をフレームごとに計算し、入力映像中の関心領域が生成映像でどの程度拡大表示されたかを数値的に調べた。

$$\eta = \frac{\text{出力フレームの関心領域を囲む最小矩形面積}}{\text{入力フレームの関心領域を囲む最小矩形面積}} \quad (8)$$

また、生成された映像を比較し特性を調べた。

#### 3.2 画素の空間的な連続性に関する条件の有効性の評価

まず、画素の空間的な連続性に関する条件(式(6))の必要性を調べるために、条件を加えない場合 ( $W_s = 0$ ) と加えた場合 ( $W_s = 1$ ) について映像を生成した。フレーム例を図3に示す。図3は左から順に、入力フレーム,  $W_s = 0$  の結果,  $W_s = 1$  の結果を示す。なお、画素の空間的な連続性に関する条件の有無による影響のみを調べるために、全ての映像で  $\alpha = \sqrt{2}$ ,  $S_{i,j,t} = 1$  とした。また、図3に示すフレームにおける  $\eta$  の計算結果を表1に示す。図3より、画素の空間的な連続性を考慮していない(ii), (v)では、ホワイトボードや本棚に歪みが目立ち、曲線的に変形しているのに対して、考慮している

表2: 図4に示すフレームにおける関心領域の拡大率  $\eta$

映像	(a)	(b)	(c)	(d)
$\eta$	1.33	1.47	1.89	1.59
映像	(e)	(f)	(g)	(h)
$\eta$	1.84	2.22	3.46	2.55

(iii), (vi)では、そのような歪みが抑えられ、違和感が少なくなっていることがわかる。この結果より、画素の空間的な連続性の考慮は、曲線的な変形を抑えた映像を生成するために必要であることが確認された。一方、表1より  $W_s = 1$  とした場合、 $\eta$  が低下することがわかる。これは、条件式の数が増えることにより、画素間の幅に関する条件式が解に与える影響が相対的に小さくなるためと考えられる。

#### 3.3 重要度の考慮の有効性の評価

周辺情報を考慮し、画素の重要度をエッジ量にもとづいて設定することが有効であるかを検証するために、 $W_s = 1$  として、 $\alpha$  と  $C_{i,j,t}$  の値を変化させた映像を生成し、結果を比較した。図4は、上段に、 $\alpha = \sqrt{2}$ , 下段に  $\alpha = 2$  とした映像生成結果を示す。(a)~(c), (e)~(g)は周辺情報を考慮せず  $C_{i,j,t}$  の値を定数とした。それぞれ、(a), (e)は  $C_{i,j,t} = 1.0$ , (b), (f)は  $C_{i,j,t} = 0.5$ , (c), (g)は  $C_{i,j,t} = 0.1$  とした結果である。対して、(d), (h)は、周辺情報を考慮し、 $C_{i,j,t} = \frac{E_{i,j,t}}{\max(E_{i,j,t})}$  とした結果である。また、 $\eta$  の計算結果を表2に示す。

$C_{i,j,t}$  の値を定数とした場合の結果の比較より、 $C_{i,j,t}$  を減少させるほど、関心領域である人物が大きく表示されるが、それと同時に周辺領域に映る物体が収縮、または膨張するなど変形する傾向があることを確認した。これは、画素間の幅の条件の重みに関する  $C_{i,j,t}$  を減少させることが、条件を緩めることにつながるため、結果として周辺領域で隣接する画素間の幅が一定でなくなり、物体が変形していると推察される。この傾向は、 $\alpha$  が大きいほど顕著に見られる。

図5は、図4(f), (h)の一部を拡大したものである。比較のために、上段に入力映像の同じ部分を拡大したものを示した。図5より、周辺情報を考慮していない(f)に比べて、周辺情報を考慮した(h)の方が、(f)よりも関心領域の拡大率が高いにもかかわらず、映像中の時計の変形が抑えられ、入力映像での形に近いまま保たれている様子が確認できる。一方、映像中で重要度が低いと考えられる床面や柱を比較すると、図4(h)の方が、より変形している。これより、周辺情報を考慮することにより、重要でない部分が変形するかわりに、重要な部分については形が保持される傾向があることを確認した。このように、重要度の考慮により、変形を抑えつつ  $\eta$  の値を大きくできることがわかった。なお、図4(b), (d)と(f), (h)の比較により、この傾向は  $\eta$  が大きくなるほど顕著に現れる。重要度の考慮は、生成映像に占める関心領域の面積が大きい場合に特に有効であると推察される。一方、関心領域のアスペクト比がわずかに変化するという問題点が見受けられる。これは、未知数である  $x_{i,j,t}$  と  $y_{i,j,t}$  を独立に求めていることが原因であると考えられる。

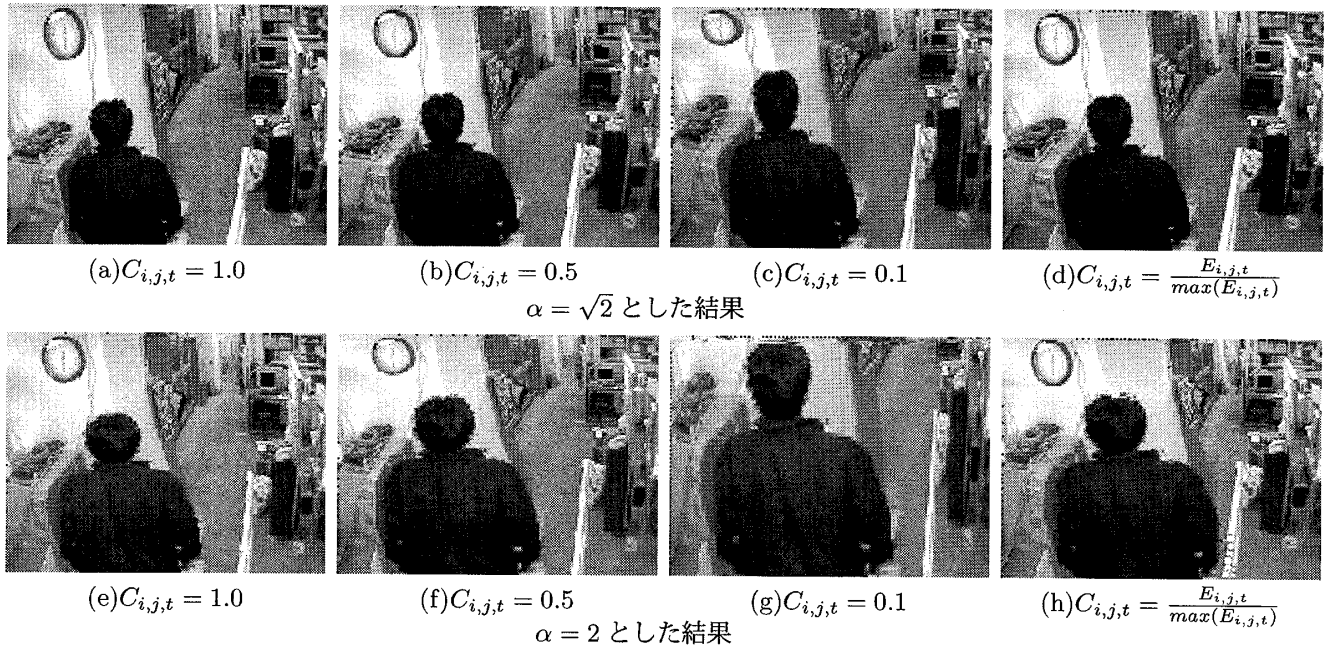
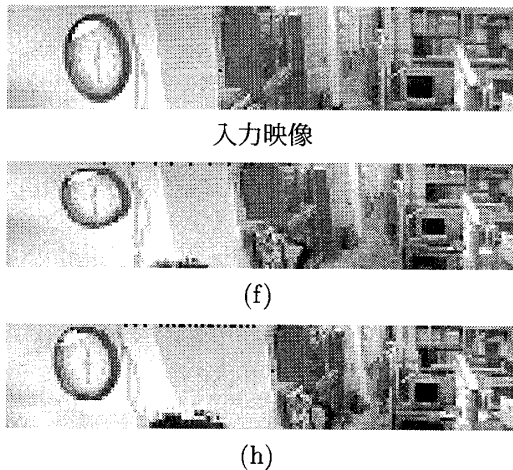
図4:  $\alpha$ ,  $C_{i,j,t}$  の値を変化させた場合の生成映像の比較

図5: 図4(f), (h) の一部

### 3.4 まとめ

本稿では、周辺領域の重要度を考慮することにより、周辺情報を保持したズーム映像を生成する手法を提案した。実験では、提案手法を用いて実際に生成した映像を使用し、条件式の有無やその重みの変化が、生成映像に与える影響を検証した。その結果、エッジ量にもとづいた重要度を重みとして設定することにより、関心領域の拡大率をある程度保つとともに、周辺領域の重要な部分の変形が抑えられる、つまり、周辺情報が保持される傾向があることを確認した。実験には、関心領域が1ヶ所のみ映像を使用した。提案手法を用いて映像中の複数の関心領域を同時にズーム表示することも可能である。今後、関心領域が複数ある映像を用いた評価や、条件式の立式に使用する画素の選択法の改良による処理時間の短縮が必要である。また、変形をより抑えた映像を生成するために、重要度の決定方法や新たな条件式の設定についても検討する必要がある。なお、本研究の一部は、

科研費の補助による。

### 参考文献

- [1] H. El-Alfy, D. Jacobs, and L. Davis, "Multi-scale Video Cropping," *Proc. 15th ACM International Conference on Multimedia*, pp. 97–106, 2007.
- [2] B. B. Bederson and J. D. Hollan, "Pad++: A Zooming Graphical Interface for Exploring Alternate Interface Physics," *Proc. 7th Annual ACM Symposium on User Interface Software and Technology*, pp. 17–26, 1994.
- [3] G. Wang, T. Wong, and P. Heng, "Real-Time Surveillance Video Display with Saliency," *Proc. 3rd ACM International Workshop on Video Surveillance & Sensor Networks*, pp. 37–44, 2005.
- [4] 塩澤 秀和, 奈須 庄健, 井前 吾郎, 重野 寛, 岡田 謙一, "魚眼ビューにおける曲線フォーカスの提案", 情報処理学会研究報告, Vol.2002, No. 38, pp. 41–48, 2002.
- [5] 岡崎 哲夫, 川野 弘道, 畠山 裕爾, "マルチフィッシュアイ・ネットワーク表示法", 電子情報通信学会技術研究報告, 情報ネットワーク (IN), Vol.94, No.368, pp. 59–66, 1994.
- [6] L. Wolf, M. Guttman, and D. Cohen-Or, "Non-homogeneous Content-driven Video-retargeting," *Proc. 11th IEEE International Conference on Computer Vision*, pp. 1–6, 2007.
- [7] V. Setlur, S. Takagi, R. Raskar, M. Gleicher, and B. Gooch, "Automatic Image Retargeting," *Proc. 4th ACM International Conference on Mobile and Ubiquitous Multimedia*, pp. 59–68, 2005.
- [8] P. Chiu, K. Fujii, and Q. Liu, "Content Based Automatic Zooming: Viewing Documents on Small Displays," *Proc. 16th ACM International Conference on Multimedia*, pp. 817–820, 2008.
- [9] C. Stauffer and W. E. L. Grimson, "Adaptive Background Mixture Models for Real-time Tracking," *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 246–252, 1999.
- [10] P. KaewTraKulPong and R. Bowden, "An Improved Adaptive Background Mixture Model for Real-time Tracking with Shadow Detection," *Proc. 2nd European Workshop on Advanced Video-Based Surveillance Systems*, pp. 1–5, 2001.