

# ロボットビジョン・ネットワークのための機能的階層符号化

## FUNCTIONALLY LAYERED CODING FOR ROBOT VISION NETWORK

田口 秀樹\* 高橋 知将\* 岩橋 政宏\* 木村 哲也\*\*

長岡技術科学大学工学部 \*電気系 \*\*システム安全系

Hideki TAGUCHI Tomoyuki TAKAHASHI Masahiro IWAHASHI Tetsuya KIMURA

Nagaoka University of Technology

### 概要

本報告では、複数のクローラ型ロボットと人々が、通信ネットワークを介して連携探索を行う状況(ロボットビジョン・ネットワーク)において、通信路の限られた伝送容量を有効活用できる、機能的階層符号化を提案する。提案法では、複数の探査ロボットから得られた映像が、常時、リモートロボットに伝送され、各探査ロボットの自己位置がモニタされる。この際、この自己位置推定という「機能」に、必要最小限の情報のみを「基本階層」として伝送することで、全体のデータ伝送量を削減する。また、必要に応じて人が映像を閲覧する際は、追加データである「拡張階層」のみを通信することで、現場映像を視認できる。本報告では、JPEG 2000 の要素技術を成分分解に用い、基本階層に含めるべき成分を実験的に決定し、データ伝送量の低減効果を確認する。

### 1 はじめに

近年、ロボットによる遠隔探査や自己位置推定に関する様々な研究が行われており、全方位カメラやステレオビジョンを用いた距離深度や障害物検出など、各種の技術が開発されている[1,2,3]。

自己位置推定は、探査ロボットに固定されたカメラの映像から、それ自身の動きを推定する技術であり、SLAM (simulation location and mapping) に活用されている[2,3]。この際、映像処理により、フレーム間のオプティカルフローや動きベクトルの推定(ME: motion estimation)が必要となる[4,5,6]。

複数のクローラ型探査ロボットと人々が、通信ネットワークを介して連携探索を行うロボットビジョン・ネットワークでは、高性能なリモートロボットが、複数の探査ロボットの映像を集約して、各探査ロボットの自己位置を推定する。このロボット間通信にとっては、ME という機能を遂

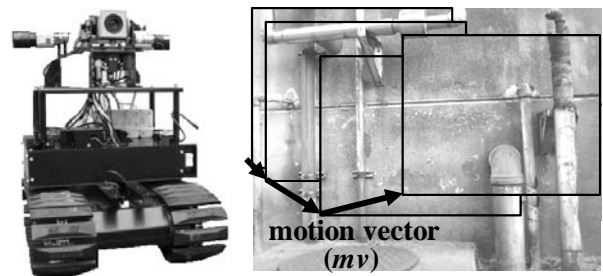
行する上で支障のない、必要最小限の情報が通信されればよい。

そこで本報告では、映像信号を複数の成分に分解し、ME 機能の実現上、必要最小限の成分のみを抽出することで、ロボット間通信におけるデータ伝送量を削減する。この成分を「基本階層」として伝送することで、各探査ロボットの位置が、リモートロボットにより常時モニタされる[8,9]。

また、必要に応じて人が現場映像を視認できるよう、視覚にとって必要な成分を、「拡張階層」として追加的に伝送する。これにより、災害現場における通信路の、限られた伝送容量の有効活用が可能となる。

提案法では、デジタルシネマの国際標準として採用された、JPEG 2000 の IP コア[7]に基づき映像符号化回路を構成する。ここでは、映像信号は、離散ウェーブレット変換(DWT)により、幾つかの周波数帯域に分解され、更に、ビットプレーン分解(BPD)により、複数のビットプレーンに分解される。これらの成分は、EBCOT によりエン트로ピ符号化される。

本報告では、これらの成分の中から、ME に必要最小限の成分のみを抽出し、これを基本階層として伝送することで、ロボット間通信時のデータ伝送量を削減できることを確認する。



(a) 探査ロボット (b) 動きベクトル (mv)

図1 自己位置推定のための動きベクトルの推定

## 2 ロボットビジョン・ネットワーク

以下、図 1(a)の探査ロボットに固定された天井カメラの映像が、リモートロボットに伝送され、動きベクトル ( $mv$ ) が ME 処理により推定され、図 1(b)の天井マップを作成する場合を考察する。

図 2(a)のように、DWT と BPD により得られた成分の中から、人にとって必要最小限の成分を抽出し、この集合  $S_H$  をリモートロボットへ伝送する場合を、本報告における従来法と呼ぶ。この集合の中から、リモートロボットにとって必要最小限の成分のみを抽出することで、ロボット間の通信量を削減することが、本報告の課題となる。

## 3 提案法

### 3.1 機能的階層符号化の考え方

図 2(b)の提案法では、ME に必要最小限の成分からなる集合  $S_R$  を、リモートロボットに伝送する。これにより  $S_R$  のデータサイズ  $DS[S_R]$  が、 $S_H$  のデータサイズ  $DS[S_H]$  より小さく、データ削減率  $\eta$  ;

$$\eta = \frac{DS[S_R]}{DS[S_H]} \quad (1)$$

が 1 より小さくなることが期待される。ここで、図 3 のように、集合をそれぞれ、

$$S_{HR} = S_H \cap S_R, S_{HO} = S_H \setminus S_{HR}, S_{RO} = S_R \setminus S_{HR} \quad (2)$$

と表記すると、探査ロボットは、 $DS[S_{HR}] + DS[S_{RO}]$  を常時リモートロボットに伝送し、 $DS[S_{HR}] + DS[S_{RO}]$  を人に必要時のみ伝送する。今回は  $S_{RO} = \phi$  の場合を考え、 $S_{HR}$  を基本(1st)階層、 $S_{HO}$  を拡張(2nd)階層として、図 4 の処理により符号化する。

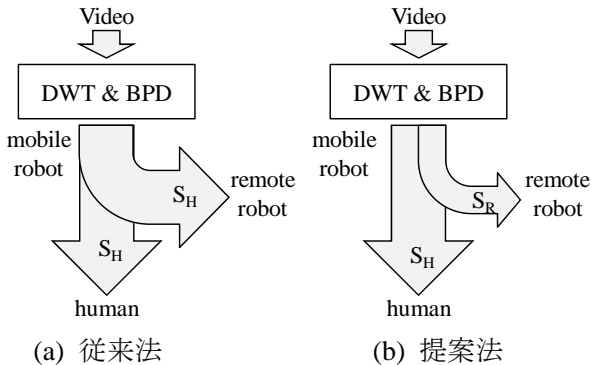


図 2 ロボット間通信への特化によるデータ圧縮

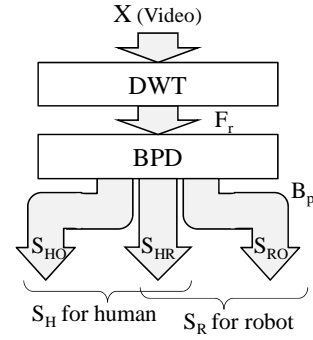


図 3 伝送データの機能的な階層符号化

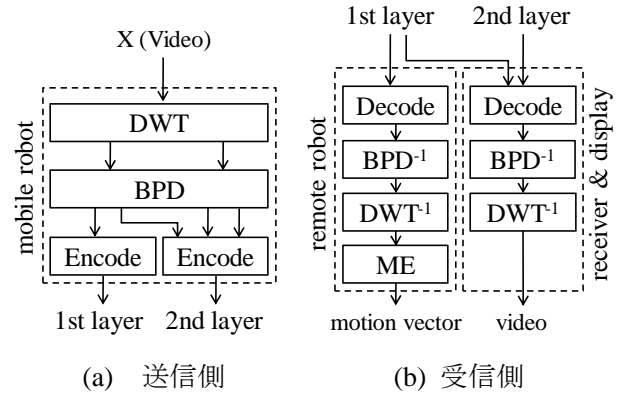


図 4 提案法のアルゴリズム

### 3.2 DWT による帯域成分への分解

映像信号  $X$  に 2D-DWT を適用すると、4 つの帯域成分  $\{LL1, HL1, LH1, HH1\}$  が生成される。同様の処理を  $LL1$  に適用すると、更に 4 つの帯域成分  $\{LL2, HL2, LH2, HH2\}$  が生成される。 $n$  回のオクターブ分割の結果、 $4+3(n-1)$  個の帯域成分が生成される。この処理を、

$$DWT[X] = F_r, F_r = \{HH1, HL1, LH1, HH2, HL2, LH2, \dots, LLn\} \quad (3)$$

と表す。ここで、 $F_r$  は帯域成分の全集合である。このとき、以下の集合族が定義される。

$$f_r \in \{sLL \mid sLL = HHm + HLm + LHm + \dots + HHn + HLn + LHn + LLn, s = m-1, m \in \{2, 3, \dots, n-1\}\} \quad (4)$$

### 3.3 BPD によるビットプレーン成分への分解

DWT の帯域成分は、BPD により、更にビットプレーン成分に分解される。これを、

$$BPD[F_r] = F_r \times B_p, B_p = \{8th, 7th, \dots, 2nd, 1st\} \quad (5)$$

と表記する。ここで、"1st"は MSB を意味する。このとき、集合族を以下のように定義する。

$$b_p \in \left\{ q \mid q = \sum_{Q=1}^q Q_{th}, q \in \{8,7,\dots,2,1\} \right\} \quad (6)$$

本報告では、 $S_{HR}$  に含めるべき集合族の直積集合  $f_r \times b_p$  を、5.において実験的に決定する。

### 3.4 必要最小成分を抽出するための尺度

人への通信に用いる成分の集合  $S_H$  は、閲覧時に要求される画質  $\theta$  のもとで、データサイズ  $DS[f_r \times b_p]$  が最小となる直積集合  $f_r \times b_p$  として、

$$S_H = \arg \min_{f_r \times b_p} DS[f_r \times b_p] \text{ subject to } SM[f_r \times b_p] > \theta \quad (7)$$

と定義される。画質評価の尺度である  $SM$  ]には、誤差分散に基づく PSNR [dB]が用いられる。

一方、ロボット間通信のための集合  $S_R$  は、ME において許容される誤差  $\sigma$  のもとで、データサイズ  $DS[f_r \times b_p]$  が最小となる直積集合  $f_r \times b_p$  として、

$$S_R = \arg \min_{f_r \times b_p} DS[f_r \times b_p] \text{ subject to } MV[f_r \times b_p] < \sigma \quad (8)$$

と定義する。ここで、 $MV[f_r \times b_p]$  は、 $f_r \times b_p$  から再構成された映像を用いて ME を行ったときの、 $mv$  の誤差量を表す。今回は、標準偏差を用いる。

### 4 動き推定 (ME) の方法

本報告では、ME として CC[4], SSD[5], POC[6] を用いて評価する。動きベクトル  $mv$  は、

$$mv_i = \arg \max_{\mathbf{m}} E_i, \quad (9)$$

$$i \in \{CC, SSD, POC\}, \quad \mathbf{m} = (m_1, m_2)$$

により推定する。但し、

(a) CC(cross correlation)

$$E_{CC} = F^{-1} \left[ X_1(\mathbf{w}) \overline{X_2(\mathbf{w})} \right] \quad (10)$$

(b) SSD(sum of squared difference)

$$E_{SSD} = -\sum_{\mathbf{n}} \{x_1(\mathbf{n}) - x_2(\mathbf{n} - \mathbf{m})\}^2 \quad (11)$$

(c) POC(phase only correlation)

$$E_{POC} = F^{-1} \left[ \frac{X_1(\mathbf{w}) \overline{X_2(\mathbf{w})}}{|X_1(\mathbf{w}) \overline{X_2(\mathbf{w})}|} \right] \quad (12)$$

とする。ここで、 $x_1(\mathbf{n})$  と  $x_2(\mathbf{n})$  は位置  $\mathbf{n}=(n_1, n_2)$  における信号値を表し、離散フーリエ変換は以下で定義される。

$$\begin{cases} X_q(\mathbf{w}) = F [ x_q(\mathbf{n}) ] = \sum_{\mathbf{n}} x_q(\mathbf{n}) W_N^{-\mathbf{w} \cdot \mathbf{n}} \\ \hat{x}_q(\mathbf{m}) = F^{-1} [ X_q(\mathbf{w}) ] = \frac{1}{N} \sum_{\mathbf{w}} X_q(\mathbf{w}) W_N^{\mathbf{w} \cdot \mathbf{m}} \end{cases} \quad (13)$$

$$W_N = e^{j2\pi/N}, q \in \{1,2\}$$

## 5 実験結果

本報告では、480×360[pixel]の天井画像から、256×256[pixel]の画像領域を2枚、ランダムな位置から切り出し、式(10),(11),(12)により  $mv$  を推定した。これを300回行い、 $mv$  の誤差量を測定した。

### 5.1 動き推定の誤差の評価

表1には、各種の  $f_r \times b_p$  において観測された誤差量の標準偏差  $\sigma$  を示す。CC の場合、 $f_r \in \{all, 1LL\}$  において、全ての  $b_p$  に対し  $\sigma=0$  となった。これは、成分を 1LL と MSB のみに削減しても、ME の誤差が生じないことを意味する。更にもっと中で、 $f_r \times b_p = 1LL \times 1$  のときにデータサイズが最小となった。つまり、この組み合わせが、リモートロボットにとって必要最小限の成分の集合  $S_R$  となる。なお、今回の実験では、SSD や CC に比べ、POC の耐性が弱い。POC は全帯域を必要とするためと考えられる。

### 5.2 データ削減率

図5には、各種のパラメータ値におけるデータ削減率の、10種類の画像に対する平均を示す。POC は0.7~18.3 [%]、CC と SSD は0.7~3.2 [%] 程度となっている。人は図6(a),(c)の画像を閲覧するが、リモートロボットにとって、ME を行う上では、図6(b),(d)の情報で十分といえる。

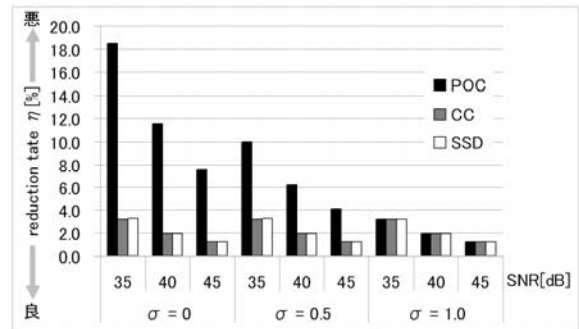


図5 データ削減率の複数画像に対する平均値

表1 MEの推定誤差 $\sigma$  [pixel]

CC		$f_r$ (Frequency band)				
		<i>all</i>	<i>1LL</i>	<i>2LL</i>	<i>3LL</i>	<i>4LL</i>
$b_p$ (bit plane)	8	0	0	0.07	0.48	3.71
	7	0	0	0.07	0.48	3.77
	6	0	0	0.07	0.48	3.72
	5	0	0	0.08	0.48	3.68
	4	0	0	0.09	0.49	3.62
	3	0	0	0.12	0.54	4.11
	2	0	0	0.16	0.62	4.11
	1	0	0	0.25	0.67	3.90
SSD		$f_r$ (Frequency band)				
		<i>all</i>	<i>1LL</i>	<i>2LL</i>	<i>3LL</i>	<i>4LL</i>
$b_p$ (bit plane)	8	0	0	0	0.34	1.20
	7	0	0	0	0.34	1.19
	6	0	0	0	0.35	1.22
	5	0	0	0	0.37	1.25
	4	0	0	0	0.34	1.15
	3	0	0	0	0.49	1.66
	2	0	0	0.08	0.63	2.70
	1	0	0	0.27	0.74	2.89
POC		$f_r$ (Frequency band)				
		<i>all</i>	<i>1LL</i>	<i>2LL</i>	<i>3LL</i>	<i>4LL</i>
$b_p$ (bit plane)	8	0	0.47	1.44	5.92	19.06
	7	0	0.47	1.44	5.91	19.06
	6	0	0.47	1.44	5.91	18.99
	5	0	0.47	1.44	5.92	19.20
	4	0	0.47	1.44	5.92	19.74
	3	0	0.47	1.44	6.07	20.85
	2	0	0.47	1.44	5.98	61.62
	1	0	0.47	1.44	6.04	110.50

表2 データ削減率 $\eta$  [%]

CC		SNR $\theta$ [dB]		
		45.9	40.1	34.3
error $\sigma$ [pixel]	0	0.712	0.896	1.243
	0.5	0.697	0.879	1.219
	1	0.696	0.877	1.217
SSD		SNR $\theta$ [dB]		
		45.9	40.1	34.3
error $\sigma$ [pixel]	0	0.712	0.896	1.243
	0.5	0.697	0.879	1.219
	1	0.696	0.877	1.217
POC		SNR $\theta$ [dB]		
		45.9	40.1	34.3
error $\sigma$ [pixel]	0	0.726	0.914	1.268
	0.5	0.712	0.896	1.243
	1	0.712	0.896	1.243

## 6 まとめ

本報告では、人とロボットが映像情報を共有する、ロボットビジョン・ネットワークにおいて有効な、機能的階層符号化を提案した。ロボット間通信に必要なとなる最小限の成分を特定することで、伝送すべきデータ量を、従来法の0.7~18.3 [%]

程度に低減できることが確認された。

謝辞：本研究の一部は、文科省・科研費・基盤(C) 20560351により実施された。

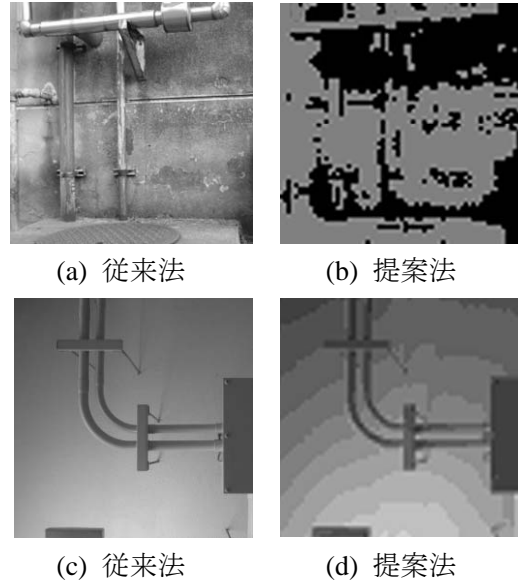


図6 MEのための再生画像の比較

## 7 参考文献

- [1] G. N. Desouza, A. C. Kak, "Vision for Mobile Robot Navigation: a Survey", IEEE Trans. Pattern Analysis Machine Intelligence, vol.24, Issue 2, pp.237-267, Feb. 2002.
- [2] D. Nister, O. Naroditsky, J. Bergen, " Visual Odometry ", Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Vol. 1, 27, pp. I-652 - I-659, July 2004.
- [3] R. Munguia, A. Grau, "Monocular SLAM for Visual Odometry", IEEE International Symposium on Intelligent Signal Processing (WISP), pp.1 - 6, Oct. 2007.
- [4] C. A. Wilson, J. A. Theriot, "A Correlation-Based Approach to Calculate Rotation and Translation of Moving Cells", IEEE Trans. Image Processing, Vol.15, Issue 7, pp.1939-1951, July 2006.
- [5] N. P. Papanikolopoulos, P. K. Khosla, T. Kanade, "Visual tracking of a moving target by a camera mounted on a robot: a combination of control and vision", IEEE Trans. Robotics Automation, Vol.9, Issue1, pp.14-35, Feb. 1993.
- [6] K. Ito, H. Nakajima, K. Kobayashi, T. Aoki, T. Higuchi, "A Fingerprint Matching Algorithm Using Phase-Only Correlation", IEICE Trans. Fundamentals, Vol.E87-A No.3 pp.682-691, March 2004.
- [7] JTC1/ SC29, "Information technology - JPEG 2000 image coding system: Core coding system", ISO/IEC15444-1, 2004.
- [8] タンスリヤボン, 鈴木貴之, 岩橋政宏, "JPEG2000 を活用したアウェアネス映像通信システム", 画像電子学会論文誌, vol.36, no.5, pp.807-813, Sept. 2007.
- [9] S. Udomsiri, M. Iwahashi, S. Muramatsu, "Functionally layered video coding for water level monitoring", IEICE Trans. Fundamentals of Electronics, Commun. Computer Sciences, Vol.E91-A, No.4, pp.1006-1014, Apr. 2008.