

*Full Length Research Paper*

# Reduction of data size for transmission in localization of mobile robots

T. Matsumoto<sup>1</sup>, T. Takahashi<sup>1</sup>, M. Iwahashi<sup>1</sup>, T. Kimura<sup>1</sup>, S. Salbiah<sup>2</sup> and N. Mokhtar<sup>2\*</sup>

<sup>1</sup>Faculty of Engineering, Nagaoka University of Technology, Nagaoka, Niigata, Japan.

<sup>2</sup>Department of Electrical Engineering, Faculty of Engineering, University of Malaya, Kuala Lumpur, Malaysia.

Accepted 22 November, 2010

In the SLAM (Simultaneous Localization and Mapping) technology, an environmental map is generated by a mobile robot. When it results in failure, it is necessary to inspect the scene. This localization and browsing require transmission of video signal to a remote place. In the system in this paper, an indoor mobile robot has two cameras. One is the "upward" which captures scenery of ceiling. The other is "forward" for scenery in front of the robot. Video signals from the cameras are encoded and transmitted from the robot to a remote server. It causes a problem that data size is too huge to be transmitted. To cope with this problem, the Functionally Layered Coding (FLC) was reported. In the existing FLC, visual motions are estimated by using the rotation invariant phase only correlation (RI-POC) technique. It can estimate two kinds of motions - translation and rotation. However, it requires doubled computational complexity and many components to be transmitted. In this paper, we analyze relation between kinetic movements of a robot and visual motions observed in videos, and propose to replace RI-POC by a simple POC. It was confirmed that the proposed method reduced data size for transmission to 61.6%.

**Key words:** Data size, video compression, localization, mapping.

## INTRODUCTION

So far, various kinds of robot vision technologies such as SLAM (Simultaneous Localization and Mapping) have been reported (Desouza et al., 2002; Nister et al., 2004; Munguia et al., 2007). In general, an environmental map is successfully generated by these techniques. When it results in failure, it is necessary to inspect the scene visually to confirm the current situation. Therefore, it is required to transmit video signal to a remote place for automatic localization constantly, and for manual browsing of the scenery when the necessity arises.

In the system we deal with in this paper, an indoor mobile robot has two cameras. One is "upward" which captures scenery of ceiling of the room. The other is "forward" for scenery in front of the robot. Visual motions observed in the video of the "upward" camera are utilized to generate a mosaic image of the ceiling (a ceiling map). Similarly, visual motions in another robot are utilized to

automatically localize it on the ceiling map (Wilson et al., 2006; Papanikolopoulos et al., 1993).

If there is no debris on the floor, location of the robot is precisely estimated. However if there is debris, estimated location contains error. The error was reduced by utilizing visual motions in video of the "forward" camera (Matsumoto et al., 2010). It requires two cameras for the error correction.

In this system, video signals from the cameras are encoded and transmitted from the robot to a remote server, so that computational complexity for video processing in the robot can be reduced for long time battery operation in a tiny electrical circuit. Therefore the problem to be discussed in this paper is data size to be transmitted.

In a huge system such as in Konolige et al. (2002), we have multiple robots which also multiply data size to be transmitted via internet from the robots to a server. Huge data brings about congestion, packet loss and delay in a digital communication network.

In Udomsiri et al. (2009), a solution to this problem – Functionally Layered Coding (FLC) - was proposed. In

\*Corresponding author. E-mail: [norrimamokhtar@um.edu.my](mailto:norrimamokhtar@um.edu.my).  
Tel. +6012-2285060.

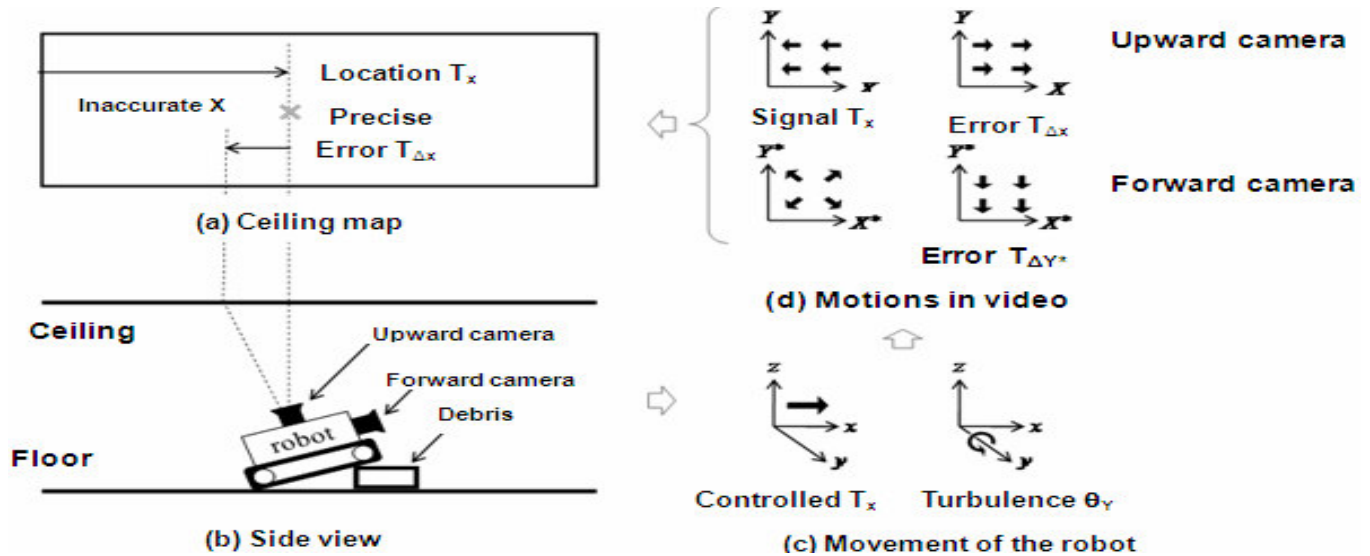


Figure 1. A mobile robot and a ceiling map.

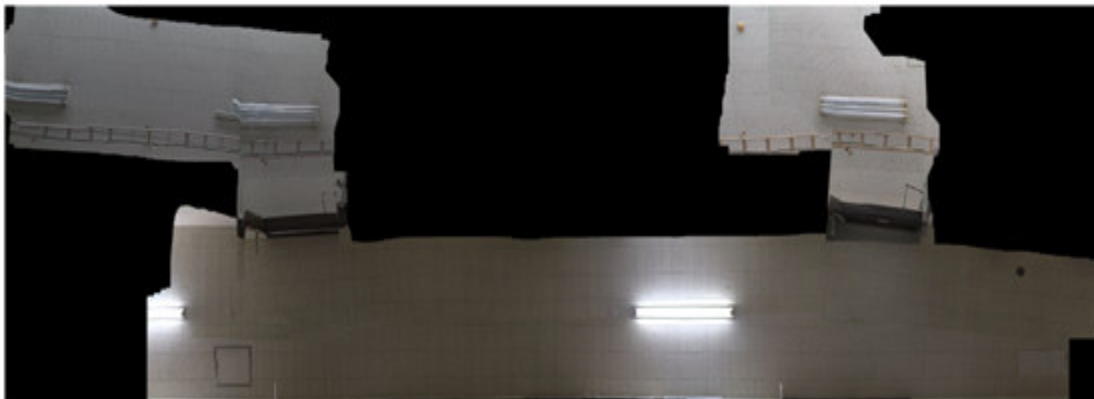


Figure 2. A ceiling map generated by a mobile robot.

FLC, only the minimum components are transmitted to a server for the automatic localization of a robot. Other components are transmitted only when a people browse the scenery or the ceiling map is generated.

In the existing FLC, visual motions are estimated by using the Rotation Invariant Phase Only Correlation (RI-POC) technique. It can estimate two kinds of motions - translation and rotation (Sasaki et al., 1998). However, it requires doubled computational complexity and many components to be transmitted.

In this paper, we analyze relation between kinetic movements of a robot and visual motions observed in videos, and propose to replace RI-POC by a simple POC reported in Ito et al. (2004). The proposed method not only reduces computational complexity in motion estimation of video, but also reduces data size to be transmitted for robot to robot communications.

## SYSTEM OVERVIEW AND PROBLEM TO BE DISCUSSED

### Localization and mapping of mobile robots

Figure 1 illustrates an overview of the system. A mobile robot has two cameras - upward and forward - as illustrated in Figure 1(b). Video from the "upward" camera captures scenery of the ceiling. Visual motions (motion vectors) between frames in the video are utilized to generate a mosaic image of the ceiling (a ceiling map). Figure 2 illustrates an example. Similarly, a motion vector between a current frame and the ceiling map is utilized to estimate location of a robot on the ceiling map.

When there is no debris on the floor, location of the robot  $T_x$  is precisely estimated as illustrated in Figure 1(a). However, if there is debris, the estimated location

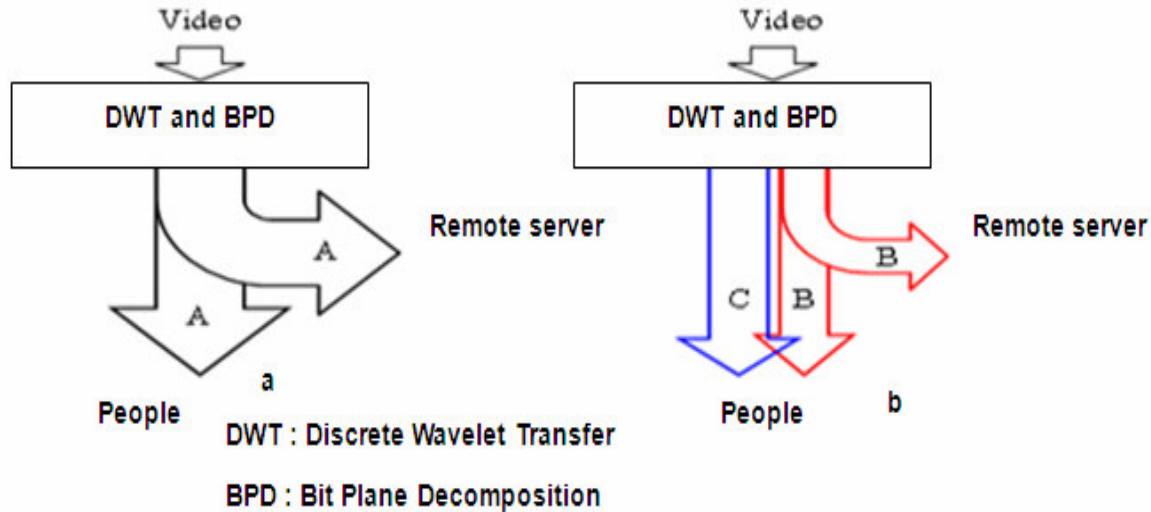


Figure 3. Video encoding: (a) motion JPEG 2000, (b) functionally layered coding (FLC).

Table 1. Movement, motion and motion estimation.

Movement of robot			Controlled		Turbulence	
			Translation	Rotation	Translation	Rotation
			$T_x$	$\theta_z$	$\theta_y$	$\theta_x$
Motion in video		Upward	$T_x$	$\theta_z$	$T_{\Delta x}$	$T_{\Delta y}$
		Forward	---	$T_y^*$	$T_{\Delta x}^*$	$\theta_{\Delta z}^*$
	Existing	Upward	RI-POC		---	
		Forward	---		RI-POC	
Motion estimation	Proposed	Upward	POC	---		
		Forward	---	RI-POC		

contains error  $T_{\Delta x}$ . The error can be compensated by utilizing feedback of the visual motion  $T_{\Delta y}^*$  in video of the "forward" camera. This procedure is detailed in Figure 1(c), (d) and explained in Matsumoto et al. (2010). It requires two cameras for this error correction.

**Data size to be transmitted**

In the system, video signals from the cameras are encoded and transmitted from the robot to a remote server, so that computational complexity for video processing in the robot can be reduced for long time battery operation on a tiny electrical circuit.

**Existing FL**

In Figure 3(a), the set "A" generally includes all the components produced by the discrete wavelet transform (DWT) and the bit plane decomposition (BPD) in the

JPEG 2000. Denoting data size of A, B and C as  $D_A$ ,  $D_B$  and  $D_C$ , the relation " $D_A = D_B + D_C$ " is basically satisfied. Therefore " $D_A > D_B$ " holds and data size for localization of a mobile robot is reduced by the FLC.

Figure 3(b) illustrates a solution to the problem. It is the functionally layered coding (FLC) proposed in Udomsiri et al. (2009). In the system, only a set "B" is transmitted to a server for localization of a robot on the ceiling map. Another set "C" is transmitted only when a people browse the scenery or the ceiling map is generated or renewed.

Table 1 summarizes relation between kinetic movements of a robot and visual motions (motion vectors) observed in videos. In the existing FLC, the two motions in video of the upward camera - translation  $T_x$  and rotation  $\theta_z$  - are estimated with RI-POC. When there exist turbulences, the two motions in the forward camera - translation  $T_{\Delta x}^*$  and rotation  $\theta_{\Delta z}^*$  - are estimated with RI-POC to reduce errors in localization (Matsumoto et al., 2010).

As explained later, RI-POC contains two POCs. It requires doubled computational complexity. It also

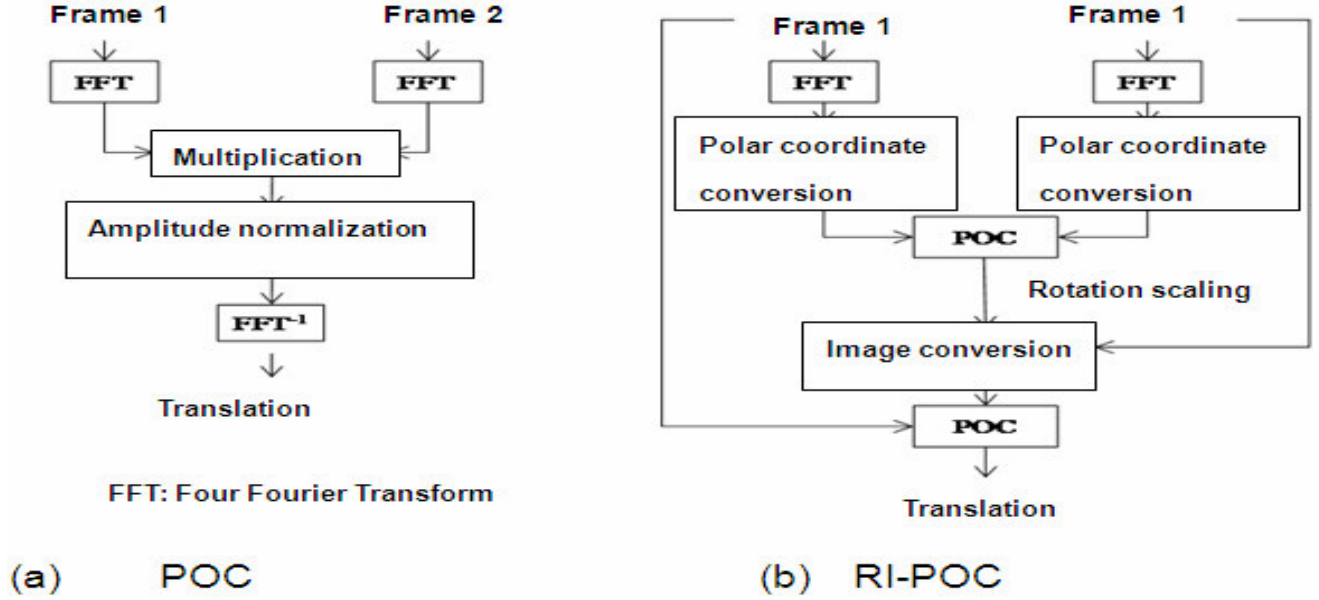


Figure 4. Motion estimation in the remote server: (a) POC, (b) RI-POC.

requires many components to be transmitted as the set "B". In this paper, we replace RI-POC by a simple POC reported in Ito et al. (2004) to reduce data size for transmission by analyzing relation between kinetic movements of a robot and visual motions in videos.

## PROPOSED METHOD FOR DATA SIZE REDUCTION

### Motion estimation

Figure 4(a) illustrates procedure of the POC. It calculates a motion vector  $mv$  - translation - between two frames as

$$mv = \underset{\mathbf{m}}{\operatorname{argmax}} E_{POC} \quad \mathbf{m} = (m_1, m_2) \quad (1)$$

where  $mv$  is given as a two dimensional displacement  $\mathbf{m} = (m_1, m_2)$  in the number of pixels which minimizes the criteria  $E_{POC}$ . It is defined as

$$E_{POC} = F^{-1} \left[ \frac{X_1(\mathbf{w}) \overline{X_2(\mathbf{w})}}{|X_1(\mathbf{w}) \overline{X_2(\mathbf{w})}|} \right] \quad (2)$$

and

$$\begin{cases} X_q(\mathbf{w}) = F \left[ x_q(\mathbf{n}) \right] = \sum_{\mathbf{n}} x_q(\mathbf{n}) W_N^{-\mathbf{w} \cdot \mathbf{n}} \\ \hat{x}_q(\mathbf{m}) = F^{-1} \left[ X_q(\mathbf{w}) \right] = \frac{1}{N} \sum_{\mathbf{w}} X_q(\mathbf{w}) W_N^{\mathbf{w} \cdot \mathbf{m}} \end{cases} \quad (3)$$

$$W_N = e^{j2\pi/N}, \quad q \in \{1, 2\}, \quad \mathbf{n} = (n_1, n_2)$$

where  $x_1(n)$  and  $x_2(n)$  denote pixel values at location  $\mathbf{n} = (n_1, n_2)$  in

different two frames.  $F[x]$  is the discrete Fourier transform of  $x$ , and  $F^{-1}[X]$  is its inverse,  $|X|$  and  $\overline{X}$  denote amplitude and conjugate of a complex number  $X$  respectively. In Equation (2), amplitude of the forward transform is normalized so that the estimation becomes robust against lighting condition (Ito et al., 2004).

The POC estimates "translation" only. It is enough as far as a robot goes straight. When a robot rotates, it is necessary to estimate "rotation" angle. For this case, the RI-POC is applied.

Figure 4(b) illustrates RI-POC. In the first stage, it outputs "rotation" angle and "scaling" factor as visual motions between the two frames. In the next stage, it generates "translation" as the third visual motion information. Namely, RI-POC requires doubled POCs as well as computational complexity.

### Proposed FLC

As summarized in Table 1, the existing FLC estimates the visual motions  $(T_x, \theta_z)$  in the upward camera and  $(T_{\Delta x}, \theta_{\Delta z})$  in the forward camera. Since each of them includes a "rotation", RI-POC is used in both of the upward and the forward.

In this paper, we propose to estimate the rotation movement  $\theta_z$  from the translation motion  $T_y$  in the forward, not from the rotation motion  $\theta_z$  in the upward. We apply relation between movements and motions reported in Matsumoto et al. (2010). Due to this replacement, it becomes possible to apply POC - not RI-POC - to the video from the upward camera. As a result, the movements  $(T_x, \theta_z)$  are derived and location of a robot is precisely estimated.

It is obvious that this replacement reduces computational complexity in motion estimation in video from the upward camera. It is also expected to contribute to reduce data size to be transmitted. In the next section, we experimentally examine this point.

## EXPERIMENTAL RESULTS

Firstly, we investigate the minimum components to be included into the set B in Figure 3(b). This set is utilized

**Table 2.** Estimation error  $T_{\Delta x}$  in [pixel].

		<b>Motion estimation</b>			<b>RI-POC</b>		
		<b>Bit plane</b>	9	8	7	6	5
<b>Existing</b>	Band	all	0.00	0.00	0.25	13.07	29.83
		1LL	0.24	0.23	7.04	19.93	30.79
		2LL	15.77	15.36	20.96	25.13	35.73
		3LL	25.68	29.04	34.58	32.43	-
		<b>Motion estimation</b>			<b>POC</b>		
<b>Proposed</b>	Band	all	0.00	0.00	0.00	0.00	3.81
		1LL	0.73	0.73	0.73	0.73	15.17
		2LL	1.16	1.15	1.14	4.69	28.92
		3LL	4.96	8.24	14.72	26.02	-

**Table 3.** Data size in [kB] to be transmitted from the upward camera.

<b>Bit plane</b>		<b>9</b>	<b>8</b>	<b>7</b>	<b>6</b>	<b>5</b>
<b>Band</b>	all	243.61	163.78	86.88	31.34	6.10
	1LL	66.74	46.73 (Existing)	26.81	10.82 (Proposed)	2.71
	2LL	17.61	12.4	7.12	2.96	1.09
	3LL	4.97	3.59	2.22	1.24	0.83

to estimate visual motions in video of the upward camera. It is carried out with RI-POC in the existing method and POC in the proposed method. The set B is also utilized to estimate motions in the forward. In this case, RI-POC is used in both of the existing and the proposed.

Secondly, we evaluate data size to be transmitted as the set B. We show that the data size can be reduced by the propose method.

We used video with  $640 \times 480$  pixels and 30 fps, DWT with 9/7 filter and EBCOT. Motion estimation is applied 100 times to a  $256 \times 256$  region at random location in the video.

### Components to be transmitted as the set B

In Figure 3(b), input video is decomposed into four frequency bands {1LL, 1LH, 1HL, 1HH} by applying DWT once. Only 1LL is decomposed again to produce {2LL, 2LH, 2HL, 2HH}. Repeating this procedure generates low frequency bands at different stages {1LL, 2LL, 3LL, ...} (JTC1/ SC29, 2004). The BPD decomposes a low frequency band into several bit planes as explained in Udomsiri et al. (2009). When the value of input video has 8 bit depth, the frequency bands has 9 bit planes.

As a result, the set B contains only a few bit planes including MSB of one of the low frequency bands {1LL, 2LL, 3LL, ...}. In the experiments below, we investigate the number of bit planes and the stages of the DWT in

the set B.

### Motion estimation error and data size to be transmitted

Table 2 summarizes the error  $T_{\Delta x}$  for various combinations of the components. When all the bands and all the bit planes are included into the set B, there is no error. In case of RI-POC, the bit plane can be reduced from 9 to 8 without generating any error. It should be noted that it can be furthermore reduced to 6 in case of POC.

When a tolerance of the motion estimation is given as 1 pixel, combination in the bold line satisfies this specification. In this case, RI-POC requires 7 bit planes (7 BP) of all the bands or 8 bit planes (8BP) of 1LL and so on. Data sizes of those combinations are 86.88 and 46.73 kB, respectively as summarized in Table 3. Namely, the existing method requires 46.73 kB at minimum for transmitting (1LL, 8BP). Similarly, the proposed method requires 10.82 kB at minimum for transmitting (1LL, 6BP). Table 4 summarizes the findings in this paper. The existing method uses RI-POC to calculate visual motions in video of the upward camera. It requires 46.73 kB to transmit the components (1LL, 8BP). On the contrary, the proposed method uses POC in the upward. It requires 10.82 kB to transmit the components (1LL, 6BP). In total, data size of the existing method is 93.46 kB. It is reduced

**Table 4.** Differences between the two methods.

		Motion estimation	Components in 1 <sup>st</sup> layer	Data size	Total data size
<b>Existing</b>	Upward	RI-POC	(1LL, 8BP)	46.73	93.46
	Forward	RI-POC	(1LL, 8BP)	46.73	(100%)
	Upward	POC	(1LL, 6BP)	10.82	57.55
<b>Proposed</b>	Forward	RI-POC	(1LL, 8BP)	46.73	(61.6%)

to 61.6% by the proposed method.

## Conclusions

In this paper, we proposed to replace a complex RI-POC by a simple POC for motion estimation of video in localization of a mobile robot by utilizing relation between kinetic movements of a robot and visual motions observed in videos. It was confirmed that the proposed method reduced data size for transmission to 61.6%. It contributes to develop a robot vision network with large number of mobile robots under a limited capacity of a digital communication network.

## REFERENCES

- Desouza GN, Kak AC (2002). Vision for Mobile Robot Navigation: a Survey. *IEEE Trans. Pattern Analysis Machine Intelligence*, 24(2): 237-267.
- Ito K, Nakajima H, Kobayashi K, Aoki T, Higuchi T (2004). A Fingerprint Matching Algorithm Using Phase-Only Correlation. *IEICE Trans. Fundamentals*, E87-A(3): 682-691.
- JTC1/ SC29 (2004). Information technology - JPEG 2000 image coding system: Core coding system. ISO/IEC15444-1.
- Konolige K, Ortiz C, Vincent R, Agno A, Eriksen M, Limketkai B, Lewis M, Briesemeister L, Ruspini E (2002). CENTIBOTS: Large Scale Robot Teams. DARPA Software for Distributed Robotics: Technical Report.
- Matsumoto T, Takahashi T, Iwahashi M, Kimura T, Mokhtar N (2010). Visual compensation in localization of a robot on a ceiling map. *Sci. Res. Essays*, (submitted).
- Munguia R, Grau A (2007). Monocular SLAM for Visual Odometry. *IEEE International Symposium Intelligent Signal Processing (WISP)*, pp. 1-6.
- Nister D, Naroditsky O, Bergen J (2004). Visual Odometry. *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 1(27): 1-652 - 1-659.
- Papanikolopoulos NP, Khosla PK, Kanade T (1993). Visual tracking of a moving target by a camera mounted on a robot: a combination of control and vision. *IEEE Trans. Robotics Automation*, 9(1): 14-35.
- Sasaki H, Kobayashi K, Aoki T, Kawamata M, Higuchi T (1998). Rotation measurements using rotation invariant phase-only correlation. *ITE Tech Rep.*, 45: 55-60.
- Udomsiri S, Taguchi H, Takahashi T, Iwahashi M, Kimura T (2009). Functionally Layered Video Coding Based on JP2K for Robot Vision Network. *J. Robotics Mechatronics*, 21(6).
- Wilson CA, Theriot JA (2006). A Correlation-Based Approach to Calculate Rotation and Translation of Moving Cells. *IEEE Trans. Image Process.*, 15(7): 1939-1951.