Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Color-Based Visual Servoing of a Mobile Manipulator with Stereo Vision

H. J. Lee*, M. C. Lee**

**Graduate School of Mechanical and Intelligent Systems Engineering, Pusan National University, Pusan, Korea (e-mail: lhjeong@pusan.ac.kr).*
***School of Mechanical Engineering, Pusan National University, Pusan, Korea (e-mail: mclee@pusan.ac.kr)*

Abstract: In this study, stereo vision system is applied to visual servoing of a mobile manipulator. The robot can recognize a target and compute the 3D position of the target by using a stereo vision system. A stereo vision system enables the robot to find the position of a target without additional information while a monocular vision system needs properties such as geometric shape of a target. Many algorithms have been studied and developed for object recognition. However, most of these approaches have a disadvantage of the complexity of computations and they are inadequate for real-time visual servoing. The other hand color information is useful for simple recognition in real-time visual servoing. In this paper, we refer to object recognition using colors, stereo matching method, recovery of 3D space, and the visual servoing.

1. INTRODUCTION

In these days, robots have been applied to various areas such as industry, service and entertainment. For robots' tasks such as guidance, serving, etc., they have to obtain much information about targets and environments. The vision system is more effective than ultrasonic sensors or laser sensors for recognizing objects and estimating the object's position. Visual servoing is to control the pose of the robot's end-effector using visual information. Visual control of the manipulator has substantial advantages for working with targets whose position is unknown.

A robot generally finds targets by using constructed database. This iterative process requires much computation time due to the complexity of algorithms. In order to simplify this process, color information of targets is used. The robot detects pixels' area of colors considered as a target or a mark. Then the robot can extract the area of the target without respect to a relationship between the camera and the target.

The robot must know position of targets to control the manipulator after detecting targets' area in images. Some information such as the geometric relationship between several feature points on a target is needed to estimate position of the target when monocular vision is used. In this paper, stereo vision is used to determine the 3D coordinate corresponding to an image plane point. The use of a stereo system requires less strict camera calibration while monocular vision is concerned with several assumptions such as geometric properties. The robot can measure the 3D position of objects by stereo matching without any additional geometric properties.

Stereo matching methods divided into two methods: area-based and feature-based. Stereo matching using area-based method is sensitive to noises because it depends on the intensity. Feature-based method includes complex processes such as interpolation but it is effective to reduce noises. In indoor environment with several sources of light, feature-based method is suitable to applications. The position of the target is computed by cameras' geometry and visual information.

In section 2, the system is described. Stereo matching and the coordinate recovery methods in 3D space are stated in section 3. Section 4 expresses about the technique for visual servoing. Results of experiments with the manipulator are shown in section 5. Finally, conclusion is given in section 6.

2. SYSTEM CONFIGURATIONS

2.1 Mobile Robot

Fig. 1 shows configurations of the robot. The system of robot is separated form the mobile robot system, stereo vision system and manipulator system. The driving part of the mobile robot has two servo motors. The sensor part consists of sixteen ultrasonic sensors and two encoders. Encoders generate pulses up to 9,850 per 1 second and ultrasonic sensors can detect the range of 10cm~5m.

2.2 Stereo Vision

CCD cameras which stand in a row on 45cm from the upper plate support resolutions of 640×480. The distance between two cameras' center is 6.5cm and cameras' lenses are set to 6mm focal length. Cameras' specification is as following Table. 1.

2.3 Manipulator

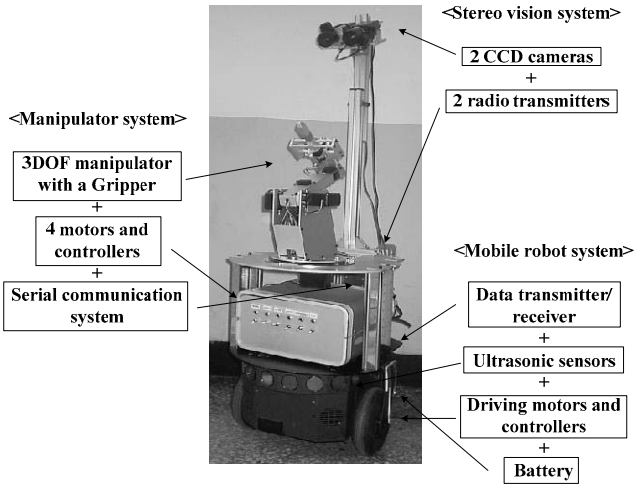


Fig. 1. System configuration of the mobile manipulator

The manipulator has 4 degrees of freedom including one gripper and is controlled by RC servomotors in each joint. The figure of the manipulator is shown in Fig. 2 and Link0, Link1, and Link2 are 15, 20, and 20cm.

3. OBJECT RECOGNITION AND RECOVERY OF A 3D POSE USING A STEREO VISION

3.1 Object Recognition

Visual control of the manipulator has substantial advantages for working with targets whose position is unknown. A robot finds a target with a constructed database. This iterative process takes up much computation time due to the complexity of algorithms. In order to simplify this process, information about color of a target is used. A mark is attached to the target to classify explicitly. The robot analyzes R, G, B values of each pixel and detects pixels corresponding to the mark. In preprocessing, we apply Histogram equalization to input images to reduce effects of darkness. Histogram equalization which is used to redivide the distribution of intensity is useful to an image with a poor intensity's distribution. This process improves the visual appearance of an image by widening peaks, compressing valleys. Books and the manipulator with each mark are shown in Fig. 3(a). Fig. 3(b) and (c) show an original image and the detected marks by color information.

3.2 Feature points extraction

After finding the mark attached to the target, the robot extracts the feature points from the mark image. To find the feature points, 'cornerness' is computed in gray-level. Cornerness is defined as the product of gradient magnitude and the rate of change of gradient direction with gradient magnitude. In order to measure cornerness, Forstner operator is used. If weight, W and cornerness, C in (2) are larger than threshold, we regard these points as candidates of feature points. The threshold value is determined experimentally. Local maxima among candidates are determined as feature points.

Table 1. Camera specification

Camera Specification	
Image Sensor	1/3" Color CCD SONY
Effective Pixel	510(H) × 492(W)
Cell Size	9.6m(H) × 7.5m(W)
TV Type	NTSC
Sync. Type	Internal
Lens (Auto IRIS)	6-12mm Vari focal

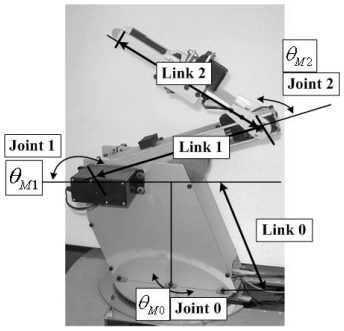
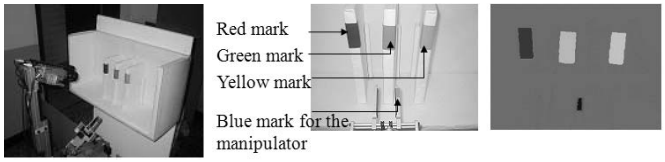


Fig. 2. Figure of the manipulator



(a) (b) (c)

Fig. 3. Mark detection (a) The real experimental environment of the mobile manipulator(b)Original image (c)Detected marks' image

$$A = \begin{bmatrix} < g_r^{\,2} > & < g_r g_c > \\ < g_r g_c > & < g_c^{\,2} > \end{bmatrix} \quad (1)$$

$$W = \frac{Det(A)}{Trace(A)}, \quad C = \frac{4Det(A)}{Trace^2(A)} \quad (2)$$

$< g_r^{\,2} >$, $< g_r g_c >$, $< g_c^{\,2} >$ indicate normalized values of $g_r^{\,2}$, $g_r g_c$, $g_c^{\,2}$ using Gaussian smoothing filter. g_r and g_c are gradient value calculated by Sobel operator. The normalized matrix, A is determined by (1). To get feature points, the left image is used as a reference. The right matching point corresponding to the left feature point is found by using gradient-based matching method. This method depends on gradient values to estimate resemblance between gradient values of the left and right points in gray-level. For searching the most similar point, gradient values of each pixel on the right image are compared with the gradient value of the left feature point within searching window. For efficiency of this operation, a size of a searching window is determined in proportion to the size of the mark in the image. If an object is far from cameras, the disparity which is the difference

between the left and right image is small enough to search the matching point by a smaller window's size. Contrary, if an object is close to cameras a searching window being a large size is needed.

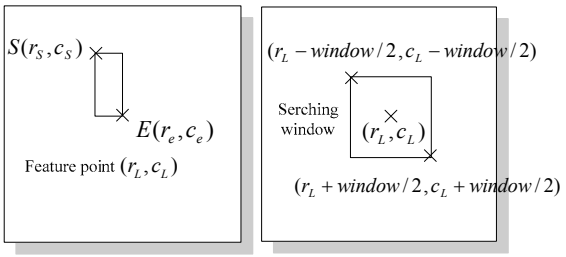


Fig. 4. Searching window in the right image related to a feature point in the left image

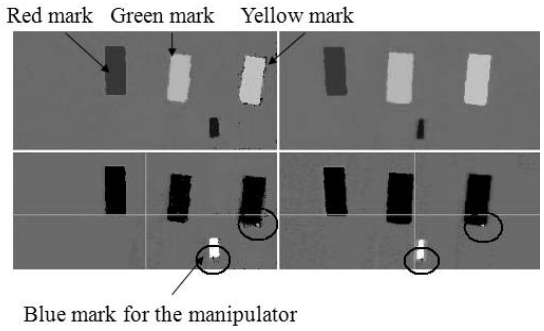
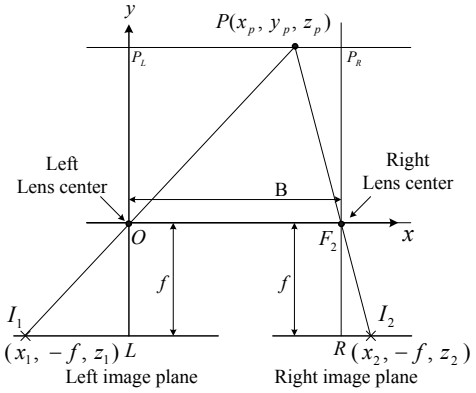
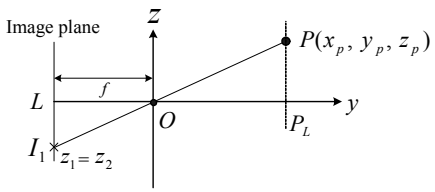


Fig. 5. Matching points of yellow target and the manipulator on the left and the right image



(a)



(b)

Fig. 6. Geometric relationship between cameras and image planes (a)Top view (b)Side view

The size of a window is determined by (3). (r_s, c_s) and (r_e, c_e) indicate the upper left corner and the lower right corner of the mark in the image like Fig. 4. a is a constant determined experimentally. Similarity S is expressed as (4) and the right matching point is (row, column) where S is the smallest. Fig. 5 shows the result of feature points extraction and stereo matching. g_L and g_R are gradient values of the left and right images.

$$window = a(|r_s - r_e| \times |c_s - c_e|) \tag{3}$$

$$S = -\alpha |g_L(left\ matching\ \ point) - g_R(row, column)| \tag{4}$$

3.2 Recovery of a 3D pose

The 3D pose can be measured by the disparity, the camera's focal length, the distance between two cameras and actual size of CCD cell. These parameters were described in section 2. In Fig. 6 describing geometric relationship between cameras and image planes, $P(x_p, y_p, z_p)$ is the position of an object and I_1, I_2 are position of object's phases on the left and right. The position is computed by using similarity of ΔPI_1I_2 and ΔPOF_2, ΔOI_1L and ΔOPP_L as (5).

$$x_p = \frac{x_1 \cdot B}{x_1 - x_2 + B}, \quad y_p = \frac{f \cdot B}{x_2 - x_1 - B}, \quad z_p = \frac{z_1 \cdot B}{x_1 - x_2 + B} \tag{5}$$

4. VISUAL SERVOING

Experimental environment of the mobile manipulator is simply described in Fig. 7. $T(x_T, y_T, z_T)$ and $M(x_M, y_M, z_M)$ indicate the global position of the target and the manipulator. The origin of the global coordinate system is set to the center of the left camera's lens. Control input is determined by (6) and it depends on only visual information. θ_{M0}, θ_{M1}, and θ_{M2} are angles of the manipulator's joints and k_0, k_1, k_2 are gain constants. x_α and z_α shown in Fig. 8 indicate the differential position from the position of target's mark for proper grip. In order to make $\dot{\theta}_{M0}$, $\dot{\theta}_{M1}$ and $\dot{\theta}_{M2}$ zero, meaning the correspondence of the end-in-hand with target's position for gripping, the manipulator is controlled.

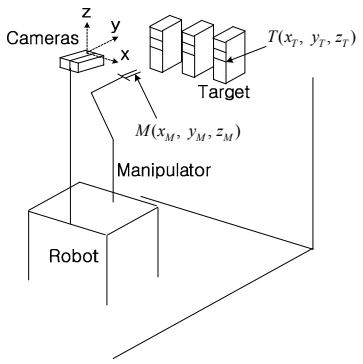


Fig. 7. Experimental environment of the mobile manipulator

$$\begin{aligned}\dot{\theta}_{M0} &= k_0 \times (x_M - (x_T + x_\alpha)) \\ \dot{\theta}_{M1} &= k_1 \times (y_M - y_T) \\ \dot{\theta}_{M2} &= k_2 \times (z_M - (z_T - z_\alpha))\end{aligned} \quad (6)$$

5. EXPERIMENTS

Experiment is achieved to estimate the accuracy of the control of the manipulator and the recovery of targets' 3D pose using visual sensor. Besides we perform localization for the mobile manipulator. A meaningful and still unsolved problem for most applications is to develop a robust and cheap positioning system. Relative position instruments such as encoders, an inertial navigation system, and a vision system are mounted inside of robots for localization and work self alone regardless with external equipment. However a current direction and length to estimate the traveled path from the last measuring data cannot be measured precisely because of the inevitable accumulated errors. Particularly, orientation errors will be larger and larger over elapsed time. On the other hand, a localization system using RFID offers absolute position of robots regardless of elapsed time. we investigate how localization technique can be enhanced by an RFID. Position of a robot and objects is recognized by using an absolute position measurement system based on RFID. A detecting range of a tag affects position error because robot's position is determined by position information of the tag. The magnitude of the error depends on the recognizing range of a reader, a distance of tags attached on the floor, and form of tags' arrangement. Moreover the variation of error is also increased when the robot lies on the boundary of the detection range of tags. To reduce the error and the variation of error, a weighting function based on Gaussian function is used. Moreover localization error is caused by long processing time spent on detecting tags and algorithm operation. The error may be generated up to several tens centimeters according to the robot's speed and seriously may affect localization and path planning of the robot in case of high speed movement. So this error should be compensated to improve the accuracy of localization. In order to measure this error, detecting the moving direction is required. It is not easy to measure the direction by using only one reader system because readers just know if tags are there or not in case of this system. So, Hough transform is used to get the robot's moving direction.

In previous research, the error of localization was measured less than 2.6(cm) in static condition. Localization and orientation experiments are executed on condition that velocity of the robot is 40(cm/sec) and the robot travels straight. The whole space of tags attached to the floor is 40×110(cm). The developed MMI program displays tags' data and position, direction of the robot by using diagram. Fig. 9, Fig. 10, and Table 2 show the experimental results of error correction in two cases, the travel in the direction of y-axis and the travel in the direction of diagonal. In order to verify the accuracy of correction, images which show the state of the robot are captured shortly after that the error is compensated. x_{image} and y_{image} indicate the position estimated by the captured image. We determine the robot's position on images by using the position of the left wheel of the robot and marks attached to the top of the robot. (x_m , y_m) is the position measured by weighted average method and (x_{com} , y_{com}) is the position compensated with the error. The accuracy of x_{com} and y_{com} can be compared with one of x_m and y_m through Fig. 9, Fig. 10, and Table 2.

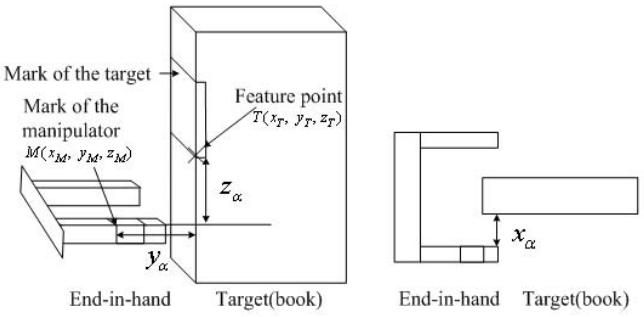


Fig. 8. Position of manipulator gripping targets

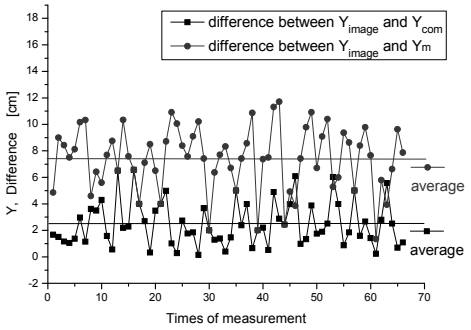


Fig. 9. Error of y_{com} and y_m while the robot moves in the direction of y-axis

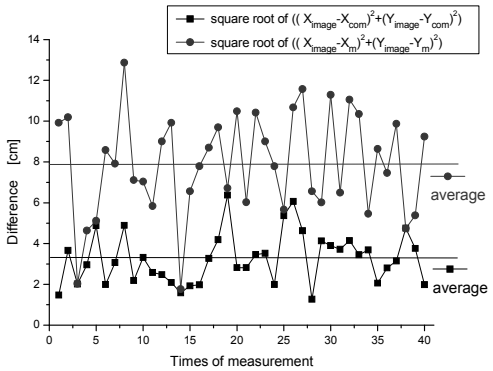
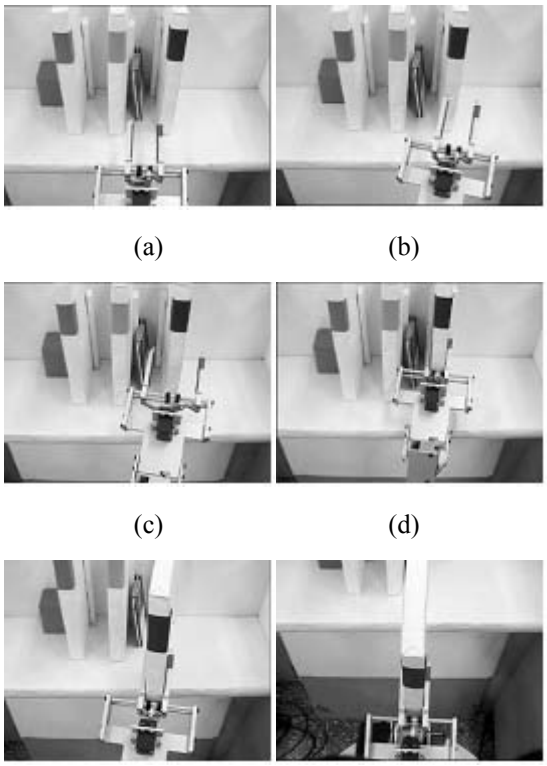


Fig. 10. Error of y_{com} and y_m while the robot moves in the direction of diagonal

Table 2. Error of measurement with and without compensation

Error [cm]		Average	Maximum	Minimum
The straight course	$\lvert y_{image} - y_{com} \rvert$	2.5	6.6	0.2
	$\lvert y_{image} - y_{m} \rvert$	7.5	11.7	1.3
The diagonal course	$\lvert x_{image} - x_{com} \rvert$	1.8	4.5	0.1
	$\lvert y_{image} - y_{com} \rvert$	2.4	5.4	0.1
	$\lvert x_{image} - x_{m} \rvert$	5.2	9.5	0.3
	$\lvert y_{image} - y_{m} \rvert$	5.7	10.4	0.7



(a) (b) (c) (d) (e) (f)

Fig. 11. Procedure of manipulator's task

The average difference between y_{com} and y_{image} is about 2.5(cm) and the deviation is about 1.655(cm) in the straight course. The other side, the average difference between y_m and y_{image} is about 7.5(cm). The deviation is about 2.39(cm). Moreover, in the diagonal course, the average difference between the compensated position and the real position shown by images is less than a half average difference of the position measured by weighted average method. It is confirmed that the accuracy of localization is improved by error correction.

The result of visual servoing is as Table 3. The origin of the global coordinate system is the center of the left camera's lens as Fig. 7. It was confirmed that the average error of measurement was about 6(mm). These errors are small

Table 3. The rate of success performing tasks in each position

Position of the middle target [cm]		35	30	25	20	15	Y / X
Target (L: left) (C: center) (R: right)	L	0	100	100	100	0	-10
	C	0	0	0	0	0	
	R	0	0	0	0	0	
	L	0	100	100	100	100	-5
	C	0	100	0	0	0	
	R	0	0	0	0	0	
	L	0	100	100	100	100	0
	C	0	100	100	100	100	
	R	0	0	0	0	0	
	L	0	100	100	100	0	5
	C	0	100	100	100	100	
	R	0	100	100	100	100	
	L	0	0	0	0	0	10
	C	0	100	100	100	100	
	R	0	100	100	100	100	

enough to manipulate the book. Measurements of the 3D recovery were tested 5times at an interval of five centimetres in the direction of y-axis and x-axis. Table 3 shows the rate of success performing tasks in each position. The position in Table 3 is the distance between the middle target and the origin. The task gripping a book was achieved with a probability of about 100% within the limited end-effector's workspace. Fig. 11 illustrates the robot grasps a red book.

6. CONCLUSIONS

The position of the unknown target was computed by using a stereo vision system and the accuracy of measurement is evaluated. We could confirm that the robot with a manipulator exactly grasps books within the limited end-effector's workspace. Besides the RFID system was used for localization and detection of the moving direction of the mobile robot. In spite of one RFID reader, the position and direction of the robot were measured by weighted average method and Hough transform in this study. And it was confirmed that the accuracy of localization was improved by correcting the error up to 66%.

REFERENCES

Bart Lamiroy, Bernard Espiau, Nicolas Andreff and Radu Horaud(2000). Controlling Robots With Two Cameras: How to Do it Properly, *Proc. of the 2000 IEEE International Conference on Robotics & Automation San Francisco.* pp.2100-2105.

D. Ben-Tzvi, V. F. Leavers and M. B. Sandler(1990). A dynamic combinatorial Hough transform. *Proc. 5th Int. Conf. Image Anal.*, 1990.

D. Hahnel, W. Burgard, D.Fox, K. Fishkin, and M. Philipose(2004). Mapping and localization with RFID Technology. *Proceedings of the 2004 IEEE International Conference on Robotics & Automation*, 2004.

Forstner, W(1987). Quality assessment of object location and point transfer using digital image correction techniques. *Proc. 15th ISPRS Congress.* pp.169-191.

H. Choset and K. Nagatani(2001). Topological SLAM Toward exact localization without explicit localization. *IEEE Trans. on Robotics and Automation.* vol. 17, no. 2, pp. 125-137.

H. J. Lee and M. C. Lee(2005). Localization of Mobile Robot Based on Radio Frequency Identification Devices. *The Society of Instrument and Control Engineers(SICE) Annual Conference 2005*.

Harris, C G(1987). Determination of ego-motion from matched points. *Proc. Alvey Vision Conf. 1987*.

J. Borenstein and L Feng(1996). Measurement and corrections of systematic odometry errors in mobile robots. *IEEE Trans. on Robotics and Automation.* vol. 12, no. 6, pp. 869-880.

J.J. Leonard and H.J.S. Feder(1999). A computationally efficient method for large-scale concurrent mapping and localization. *Proc of the Ninth Int. Sym. on Robotics Research (ISRR)*.

L. Xu, E. Oja and P. Kultannen(1990). A new curve detection method: randomized Hough transform (RHT). *Pattern Recognition Lett. 11, 1990*.

O. Kubitz, M. O. Berger, M. Perlick, and R. Dumoulin(1997) Applicaion of Radio Frequency Identification Devices to Support Navigation of Autonomous Mobile Robots. *Vehicular Technology Conference, 1997 IEEE 47th*.

R. C. Baker and B. Charlie(1989). Nonlinear unstable systems. *International Journal of Control*, vol. 23, no. 4, pp. 123-145.

R. O. Duda and P. E. Hart(1972). Use of the Hough Transformation to Detect Lines and Curves in Pictures", *Communications of Association for Computing Machinery*, vol. 15, no. 1, pp. 11-15, 1972.

Randy Crane(1997). *A Simplified approach to Image Processing*. Prentice-Hall.

Y. Zhang and R. Webber(1996). A Windowing Approach to Detecting Line Segments Using Hough Transform. *Pattern Recognition*, vol. 29, no. 2, pp. 255-265.