Scientific Research

Table of Contents

Volume 1 Number 1 **December 2009**

Intelligent Control and Automation (ICA)

Journal Information

SUBSCRIPTIONS

The *Intelligent Control and Automation* (Online at Scientific Research Publishing, www.SciRP.org) is published quarterly by Scientific Research Publishing, Inc.,USA.

E-mail: service@scirp.org

Subscription rates: Volume 1 2009

Print: $50 per copy.

Electronic: free, available on www.SciRP.org.

To subscribe, please contact Journals Subscriptions Department, E-mail: service@scirp.org

Sample copies: If you are interested in subscribing, you may obtain a free sample copy by contacting Scientific Research Publishing, Inc at the above address.

SERVICES

Advertisements

Advertisement Sales Department, E-mail: service@scirp.org

Reprints (minimum quantity 100 copies)

Reprints Co-ordinator, Scientific Research Publishing, Inc., USA.

E-mail: service@scirp.org

COPYRIGHT

PRODUCTION INFORMATION

For manuscripts that have been accepted for publication, please contact:

E-mail: ica@scirp.org

Intelligent Control and Automation, 2009, 1, 1-11

Evaluating Employee Responses to the Lean Enterprise System at a Manufacturing Company in Cape Town, South Africa

Bingwen Yan
Cape Peninsula University of Technology

Keith Jacobs
Cape Peninsula University of Technology

Abstract

There is usually much reaction amongst employees when a new system is introduced in an organization. These changes are intended to improve performance but sometimes cause considerable controversy amongst the employees and management. This study examines the implementation of Lean Enterprise (LE) system and it attempts to analyze the reactions of employees at a manufacturing company (SMC.CO) in Cape Town, South Africa. Some of the questions that were asked in the research include the following: What benefits did employees perceive by the introduction of LE? How did employees respond to the implementation of the LE at SMC (in other words, did they welcome it or not)? A semi-structured questionnaire was utilized to determine the responses of employees with respect to the benefits of these innovative approaches of production with specific reference to LE. The finding of this study indicates that LE plays a significant role in a company.

Keywords: *Lean Enterprise System, Employee Responses*

1. Introduction

Companies routinely introduce new systems to enhance efficiency. In most cases employees, especially those at the lower levels of the organisation are not consulted about such changes. They are seen and treated as mere receptacles that have to implement what is put before them. In some cases innovations are well received; but often they are not, and rebellion follows. Those employees who embrace the innovation can reap great benefits, including reduced cost, raised productivity, and short lead times. Those who rebel can sometimes cause great damage, which often results in work stoppages and lost time. In other words, the human factor—the people involved, those who will drive the activity—is often ignored.

Industry must realize that the people who lead major enterprises have to be considered when any change, especially drastic change, considered. In these times, when technology has seemingly begun to overshadow human beings, it is especially important to remember that people are still the developers of machines.

2. Background of the Study

With the implementation of new operating systems within the company, employees might feel unappreciated and marginalized if not consulted about the implementation. Too often both management and consultants hurry to get the job done and may undermine the

importance of understanding employee feelings and attitudes [13]. The feelings and attitudes of employees may influence the course of the LE implementation. While employees' high zeal can assist the implementation of LE, employees with lower morale may interfere with the process of LE implementation. The employees sometimes react indifferently, or do not give full cooperation when they are not properly informed about pending innovations in companies. In this study, the LE has been implemented at SMC since 1999. The LE model (figure 1) was developed in 2003 by SMC as a structured way to improve company performance. Sustaining and expanding lean benefits requires a supportive system, a framework to "focus" the lean principles to be followed. The support is required until LE has been internalized by the organization and become self-sustaining. The LE model focuses the company's vision, production excellence, business process excellence, people excellence, and business excellence.

Figure 1. Lean Enterprise Model
(Source: Author based on SMC's profile)

However, employee responses and reflections to the implementation of Lean Enterprise (LE) had not been previously analysed at SMC.With the above thoughts in mind, the researcher endeavored to elucidate what the employees' responses were in a particular company that introduced a new management system. The researcher went to SMC.CO and spoke to the key persons, the manufacturing manager and financial director, in this company and they agreed to assist in his research. They had not thought about this issue before and were therefore keen to discover what their employees' responses were, and whether or on what level the employees accepted the new system.

3. Problem Statement and Research Questions

This study is driven by the following research questions:

- What benefits did the employees perceive that they received through the introduction of LE?

- How did employees respond to the implementation of LE at SMC?

4. Literature Review

4.1. Human Factors

Human factors can play a significant role in an organization. Various perspectives emerge from the literature review on hoe human factors influence the Lean process. Sawhney and Chason (2005:76-79) concur with the postulate of several authors, inter alia [16]:

Lean is a knowledge-intensive process and as such relies heavily on the skills of the people and how they respond to changes [4]. The dependability and reliability of the workforce become more important because Lean introduces fragility into the system by stretching it and removing contingencies [19]. Further, Lean calls for a feeling of ownership of the process, and Lean implementation is based on the implicit belief that the workforce "naturally wants to work" [7]. Moreover, in the context of the Lean philosophy of minimizing waste of any kind, it is important not only to eliminate material waste, but also waste caused by human behavior. Behavioral productivity is as important as manufacturing productivity [5]. Lean also calls for flexibility and involvement of the workforce since it introduces more interdependencies between all "actors involved in the production process" [2].

4.2. Lean Enterprise (LE)

The term "Lean" was first coined by Womack et al in The Machine that Changed the World [19]. It was also introduced as a manufacturing approach: ". . .compared to mass production it uses less of everything-half the human effort in the factory, half the manufacturing space, half the investment in tools, half the engineering hours to develop a new product in half the time. Also it requires keeping far less than half the needed inventory on site, results in many fewer defects, and produces a greater and ever growing variety of product."

Lean principles may be applied to any organizational type and can be applied to all areas within the business [14]. Lean is a three-pronged approach incorporating a belief in quality, waste elimination and employee involvement, supported by a structured management system (Figure 2) [14].

Figure 2. Structured Management System [14].

From these initial concepts mentioned above, an array of researchers, academics, companies, and industries have developed an expanded vision of the values, behaviours and practices within enterprises that constitute a new and emerging expression of what it means to be an “LE” [20]. A commonly held definition of LE was described as: “a group of individuals, functions, and sometimes legally separate but operationally synchronized organisations” [14].

4.3. The Benefits and Problems of the Lean Implementation

According to Emiliani et al (2005:371), senior managers become interested in adopting Lean principles and practices because they result in many benefits, such as higher quality products and services, higher productivity, better customer focus, faster responses, and higher asset efficiency [6]. Heumans (2002:31) summarized the benefits and the immediate results are: reduced cycle time, fewer material handling errors, and improved labor productivity (Table 1) [10].

Table 1. Lean improvements lead to strategic benefits

Lean Improvement	Immediate Result	Business Benefit
One-piece flow work cells	Reduced production cycle time	On-time delivery
Improved flow between operations through internal JIT	Reduced work-in process	Less space required
Kanban systems	Fewer material handing errors	On-time delivery
External supplier JIT	Reliable material sourcing	On-time delivery to manufacturing
Setup & changeover reduction	Shorter production runs possible	Reduced raw materials quantities
Total productive maintenance	Less production downtime	Greater production flexibility
First-time quality	Inspection eliminated	Quality to customers assured
Employee involvement	Improved labor productivity	Improved quality of product

Source: Author based on Heumans (2002:31). Leading the Lean Enterprise [10].

Smeds (1994:66-82) concluded that the positive attitude towards development and innovation has been preserved in the plant, and "Lean" ideas are spreading further in the company, which amplifies the transition to a Lean enterprise [17]. Many companies that have adopted Lean manufacturing principles modeled after the Toyota Production System (TPS) have been able to enhance their competitive position [1]. However, not all the perceptions of Lean production are positive. Lean production systems could be viewed through a Marxist lens as being exploitative and inducing high pressure on the shop floor workers [11]. Lean production is de-humanizing and exploitative [18] and it maintains that JIT could lead to higher work intensity and stress levels among line operators [7]. Lean production practices can underline work intensity and increase stress [12].

4.4. The Employees’ Responses to the JIT System

The JIT production system is a highly integrated production, sales and distribution system leading to continuous flow through the whole supply chain, and it reduces waste and improves quality in all business operations [3]. The implementation of any new program in an organisation requires support from most departments in the company [9]. Employees like the JIT environment better than the batch-processing environment, and management can successfully make organisational changes necessary to implement JIT without negatively affecting employee attitudes [8]. For example, in a batch-processing environment, an employee's primary responsibility is to achieve a high output on a single task, employees have the security of knowing what their job is each day and seeing all the work-in-process sitting around indicates there is work to be done; under JIT, not only is work-in-process greatly reduced, but also employees do not know what they will be doing each day [8]. The effects of a two-phase introduction of JIT manufacturing practices on job characteristics and psychological wellbeing [15], this shows that the employees saw themselves as having greater control related to the timing or pacing of their tasks and the methods used to carry them out. Employees should be encouraged to view JIT as an opportunity to improve the company's competitive position as well as an opportunity to secure greater job security for themselves [9].

By reviewing the literature regarding the benefits of Lean implementation and the employees' response to the JIT system, the researcher learned that the human factor certainly plays a significant role in any organization, specifically in a manufacturing company. The understanding of the researcher with regard to JIT and Lean manufacturing is as follows: Simply put, JIT is a comprehensive management system placing emphasis on eliminating waste, reducing cost, and enhancing a firm's competitiveness; Lean manufacturing focuses on reducing inventories and using the exact amount of resources, such as space, inventory and employees required to achieve high performance.

5. Research Method

This case study utilized a quantitative research approach. A case study is defined as an empirical phenomenon within its real-life context, especially when the boundaries between phenomenon and context are clearly evident [21].

SMC.CO is located a few kilometers from the university campus and therefore easily accessible. All 82 employees at SMC were chosen as the sample for this research.

The questionnaire was designed by following the Likert scale style. It consists of two major parts: a personal profile of the respondent and questions relating to their decision-making mechanisms. The questionnaire contained questions that identified what employees thought about LE, and it used the item-total correlation-formulation to calculate each index, such as mean, range, and standard deviation.

The data have been analyzed using the Statistical Package for Social Sciences (SPSS), version 16.0. The data analysis through SPSS generated the results of descriptive statistics such as frequency, mean, standard deviation, etc. These distributions showed the frequencies of employees' responses and percentages for each of the items in the questionnaire with regard to the LE implementation at SMC. In addition, Kruskal-Wallis Tests and Chi-Square were used to test for significant differences (Alpha level = 0.05). The full results of the study are reported in the next section.

6. Results and Discussion

The researcher handed out 82 questionnaires and received back 54 completed questionnaires (66% response rate). It took almost three weeks to collect the questionnaires. The final number of respondents was 54. Two of these were unusable because they were totally spoilt. Thus only 52 questionnaires were analysed in this research. The response results are given below:

6.1. Descriptive statistics for sample

The biographical characteristics of the respondents are presented in graphical format below.

Results depicted in Figure 3 indicates that 67% (n = 35) of the sample was male, while only 33% (n = 17) was female.

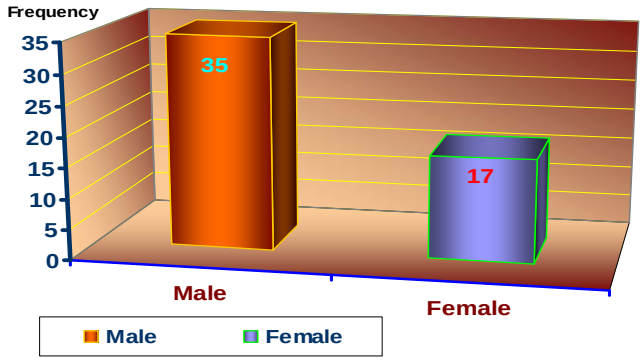


Figure 3. Gender

In Figure 4, respondents with qualifications higher than Grade 9 were in the majority (n = 34, that is 60%), while respondents with lower than Grade 9 qualifications comprised 35% (n = 18) of the sample.

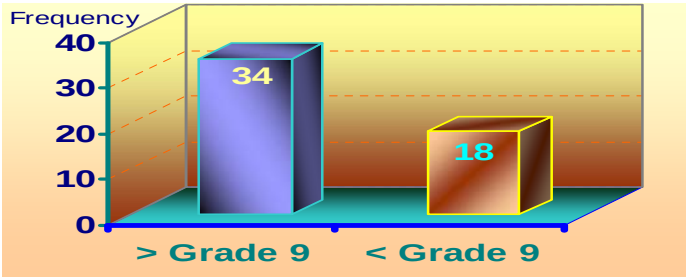


Figure 4. Educational level

From Figure 5, it can be inferred that the majority of the respondents, that is 32 are younger than 40 years of age, while a further 20 respondents are older than 40 years of age.

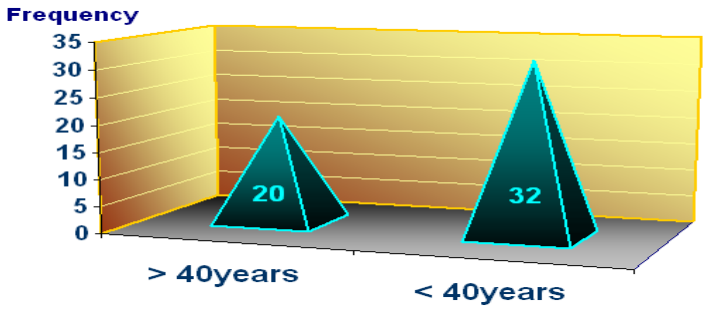


Figure 5. Age

Results in Figure 6 indicates that the majority of the respondents, that is 81% (n = 44) were shop-floor employees, while management comprised 15% (n= 8) of the respondents. Two respondents, that is 4%, did not indicate their job title.

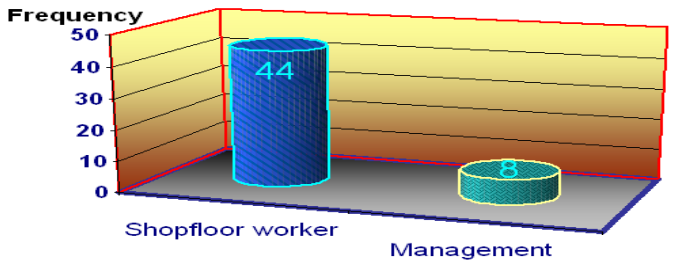


Figure 6. Job title

Figure 7 provides an overview of the race of the sample. It is evident that the majority of the respondents, that is 87% (n = 40) were Colored, while 9% (n = 4) were Black and only 2 respondents, that is, 4%, were White.

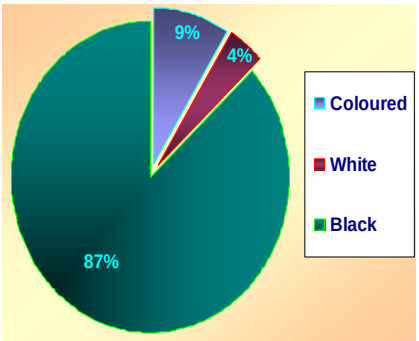


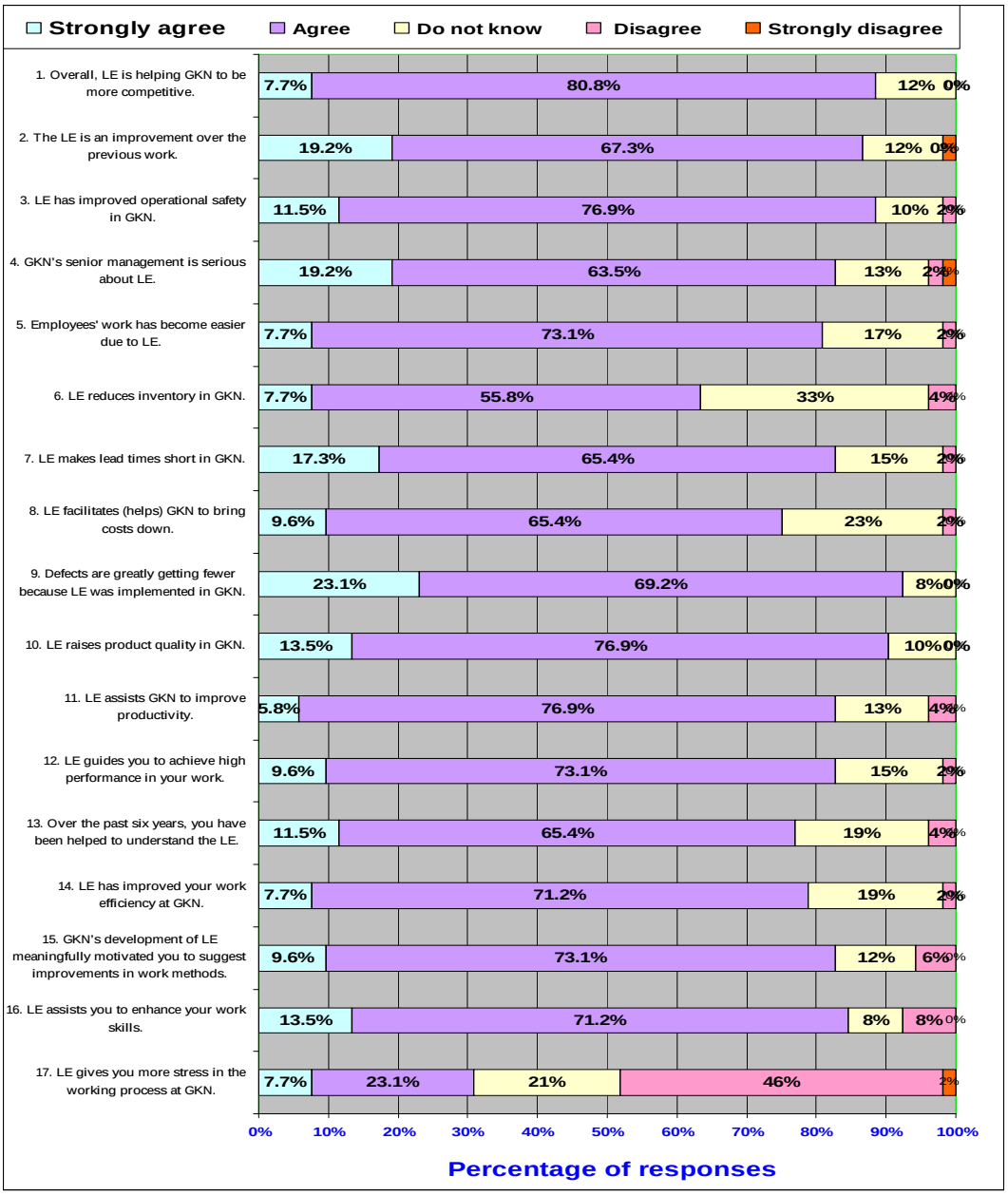
Figure 7. Racial classification

6.2. Descriptive statistics for the self-administered questionnaire

An important part of responses in the questionnaire was where the employees were required to explain why they had chosen a particular score from the numerical ranking 1—5.

All the comments were analysed through the Statistical Package for Social Sciences (SPSS) and shown in Table 2.

Table 2. Percentage of Responses



According to Table 2, responses to questions 1 to 16 indicated that most of the employees agree that the LE method as a system which reaps significant benefits for the company. These benefits were affirmed by the employees at SMC. This demonstrates that, if LE method is used correctly to address production problems, operational performance will improve.

Overall, total responses have shown positively from question 1 to 16 besides 17. However, it indicated that LE did not affect employees negatively in terms of work intensity

and stress. According to the total counts of responses, 70.1% (401) agreed, 12.9% (74) strongly agreed, 15% (86) did not know, 1.6% (9) disagreed, and 0.3% (2) strongly disagreed.

The results above indicate that the majority of the respondents were positive about the contribution of the introduction of LE into their enterprise, with the majority of them responding in the affirmative with respect to improvements in productivity, quality and operational safety.

6.3. Perceptions Regarding the Implementation of LE at SMC

In general, the findings show that most employees' responses to and reflections on, LE are positive. Many of the benefits of the implementation of LE were addressed above, and all of these benefits were described in the literature review. These included: raising competitiveness, a shorter lead time, improving productivity, raising quality, cost cutting and saving, enhancing operational safety, achieving high performance, improving work skills, raising work efficiency, and motivating employees' initiatives.

For example, in Table 2, in response to question 1, 80.8% employees agreed and 7.7% strongly agreed that LE is assisting SMC to be more competitive. According to the comments made in response to the questionnaire, a high number of employees believe that LE makes SMC's product quality better than that of their competitors. One of the shop floor workers responded: "LE makes the quality of the production number one, because LE helps SMC to reduce cost of scrap, improves operational efficiency". Some of the employees believe that if everyone follows LE completely, SMC will be an excellent company. In response to question 2, 67.3% employees agreed the LE is an improvement over the previous work, 19.2% strongly agree, 12% do not know.

Some employees believe that a lot of improvements due to the implementation of LE, such as time and cost cutting. Other employees agree that PPM (the rate of scraps) has improved a lot; staff and managers comment that LE really makes management easier. In response to question 3, 76.9% employees agreed and 11.5% strongly agreed that LE has improved operational safety at SMC. Ten percent did not know, and 2% disagreed. Many employees believe that accidents and injuries have reduced tremendously as LE established a safe environment for the employees.

Despite these positive responses, a few comments were made negatively, such as insufficient training, and LE is not fully implemented in all cells. For instance, there are employees who felt that there is no improvement because some employees are still making simple quality mistakes.

Responses to question 17 focused on whether LE resulted in greater work intensity and increased stress. Surprisingly, the comments indicate that 23.1% employees agree, 7.7% strongly agree, 21% do not know, 46% disagree, and 2% strongly disagree. Obviously, the numbers who disagree were much higher than the numbers who agree. However, the literature review gave the opposite viewpoints to the reality expressed above.

Several authors contended that Lean can be an important factor of work intensity and stress. In order to clarify this issue at SMC, the researcher later utilized the quantitative method to demonstrate the work stress that exists at SMC. Therefore, both positive and negative responses can reflect the employees' original perspectives in different ways due to the LE. It also derives the quantitative components needed to identify and test the results in this research.

7. Conclusion and Recommendations

The employees' responses showed that the LE implementation had a generally positive. The finding of this study indicates that LE plays a significant role in company's performance.

The overall benefits from the implementation of the LE included the following: enhanced company competitiveness, reduction of costs, a shorter lead time, elimination of waste, and improved product quality. Essentially, the researcher found that increased work intensity and stress to which employees referred were not necessarily reflected in their responses to other questions. The employees' work became more regular due to the implementation of the LE, and employees believe that the LE is assisting their work in the correct way at SMC.

The organization should consider establishing an internal monitoring body to evaluate the efficacy of LE. Management support is crucial in this regard, and corporate strategy and written policies underpinning LE play a significant role as well. It should be noted that the findings pertain specifically to the organization at which this research was undertaken. This small sample is a consequence of the size of the organization as well as of the exploratory nature of the study and the restrictions on its nature.

8. Recommendations for future research

Although employees were overwhelmingly positive about the benefits of the introduction of LE at SMC, the stress induced by its introduction warrants further attention, since coping with organizational restructuring, business process re-engineering, and change are important considerations confronting a multitude of organizations. Organizations with larger workforces are generally perceived to be more progressive, which could possibly account for some of the positive responses in the present study. A similar study should be conducted comparing similar industries with each other, involving a larger sample.

References

[1] Beachum D. 2005. Lean manufacturing beefs up margins: *pull systems, takt time, and one-piece flow benefit the operation of a powder coating system***.** Metal Finishing. [Online]: http://www.sciencedirect.com. 27 October 2005.

[2] Biazzo, S., Panizzolo, R. 2000. The Assessment of Work Organisation in the Worker's Perspective. *Intergrated Manufacturing Systems*, 11(1), 6-15.

[3] Chandra. S., Kodali. R. 1998. Justification of just-in-time manufacturing systems for Indian industries. *Integrated Manufacturing Systems*. MCB UP Ltd, 9(5):314-323. [Online]: http://www.emeraldinsight.com/10.1108/09576069810230428. 5 March 2006.

[4] Drew, J., McCallum, B., Roggenhofer, S. 2004. Journey to Lean: Making operational Change Stick. New York: Palgrave Macmillan.

[5] Emiliani, M.L. 1998. Continuous personal improvement. *Journal of Workplace Learning*, 10(1):29-38.

[6] Emiliani, M.L., Stec, D.J. 2005. Leaders lost in transformation. *Emerald Group Publishing Limited*. pp370-387. [Online]: http://www.emeraldinsight.com/10.1108/01437730510607862. 26 October 2005.

[7] Forza, C. 1996. Work Organisation in Lean Production and Traditional Plants: What are the differences? *International Journal of Operations & Production Management*, 16(2), 42-62. MCB UP Ltd. [Online]: http://www.emeraldinsight.com/10.1108/01443579610109839. 28 December 2005

[8] Groebner, D. F., Merz C. M. 1994. The Impact of Implementing JIT on Employees' Job Attitudes. *International Journal of Operations & Production Management*. MCB UP Ltd, 14 (1): 26 – 37. [Online]: http://www.emeraldinsight.com/10.1108/01443579410049289. 26 October 2005.

[9] Gupta. Surendra. M., Al-Turki. Yousef A.Y., Perry. Ronald F. 1999. Flexible Kanban system. *International Journal of Operations & Production Management*. MCB UP Ltd. 19 (10): 1065-1093. [Online]: http://www.emeraldinsight.com/10.1108/01443579910271700. 6 March 2006.

[10] Heumans B. 2002. Leading the lean enterprise. *Industrial Management;* Sep/Oct 2002; 44, 5; ABI/INFORM Global. [Online]:
http://proquest.umi.com/pqdweb?did=222373981&sid=2&Fmt=4&clientId=48290&RQT=309&VName=PQD. 6 March 2006.
[11] Hines, P., Rich, N. 2004. "Learning to evolve: a review of contemporary lean thinking". *International Journal of Operations & Production Management*, 24 (10): 998. Emerald Group Publishing Limited. [Online]: http://www.emeraldinsight.com/10.1108/01443570410558049. 9 December 2005.
[12] Klein, J. 1989. The human cost of manufacturing reform. *Harvard Business Review*, March-April, pp.60-66.
[13] Koo, H., Koo, LC., Tao, F K. C. 1998. Analysing employee attitudes towards ISO certification. *Managing Service Quality*. 8(5): 312-319. MCB UP Ltd published. [Online]:
http://www.emeraldinsight.com/10.1108/09604529810235772. 6 march 2006.
[14] Lucansky P., and Burke R. 2003. What is Lean Enterprise? *Supply Chain Planet International Limited published*, London. [Online]: http://www.supplychainplanet.com. 24 October 2005.
[15] Mullarkey. S., Jackson. P.R., Parker. S.K. 1995. Employee reactions to JIT manufacturing practices: a two-phase investigation. *International Journal of Operations & Production Management*, 15(11): 62-79. MCB University Press. [Online]: http://www.emeraldinsight.com/10.1108/01443579510102909. 26 March 2006.
[16] Sawhney R., Chason S. 2005. Human Behavior Based Exploratory Model for Successful Implementation of Lean Enterprise in Industry. *Performance Improvement Quarterly*, 18(2):76-96. [Online]: http://cpi.utk.edu/publications/piq.pdf. 12 March 2006.
[17] Smeds Riitta. 1994. Managing Change towards Lean Enterprises. *International Journal of Operations & Production Management*, MCB University Press.14 (3): 66-82. [Online]:
http://proquest.umi.com/pqdweb?did=878211&sid=2&Fmt=3&clientId=48290&RQT=309&VName=PQD. 31 Oct. 2005.
[18] Williams, K., Harlam, C., Williams, J., Cutler, T., Adcroft, A., and Johal, S. 1992. "Against lean production", *Economy and Society*, 21 (3):321-54.
[19] Womack, J. P., Jones, D. T., and Roos, D. 1990. The machine that changed the world. New York, NY: Macmillan Publishing Company.
[20] Womack, J. P., and Jones, D.T. 1996. Lean thinking. New York, NY: Simon & Schuster.
[21] Yin, R. K. 2003. “CASE STUDY RESEARCH: Design and Methods”. 3rd Edition. SAGE Publications Ltd.

Intelligent Control and Automation, 2009, 1, 12-21

TDoA based UGV Localization using Adaptive Kalman Filter Algorithm

W.J. Sung, S.O. Choi, K.H. You
Sungkyunkwan University, Suwon, 440-746, Korea

Abstract

The measurement with a signal of time difference of arrival (TDoA) is a widely used technique in source localization. However, this method involves much nonlinear calculation. In this paper, we propose a method that needs less computation for UGV location tracking using extended Kalman filtering based on non linear TDoA measurements. To overcome the inaccurate results due to limited linear approximation, this paper suggests a position estimation algorithm based upon an adaptive fading Kalman filter. The adaptive fading factor enables the estimator to change the error covariance according to the real situation. Through the comparison with other analytical methods, simulation results show that the proposed localization method achieves an improved accuracy even with reduced computational efforts.

Keywords*: Adaptive Kalman Filter Algorithm, TDoA, UGV*

1. Introduction

The confirmation of a present position and the estimation of a future path are one of the most important techniques for unmanned ground vehicle (UGV) realization [1]. For many cases, the global positioning system (GPS), which is famous for the good performance in position estimation, has been widely used [2-3], [9]. However, there have been some problems to be applied for UGV. For example, GPS system needs separate receivers, and constantly it has to obtain signals from more than 3 satellites for position tracking. It is also vulnerable to the indoor case or the reflected signal fading.

There have been many efforts to solve the localization problem in a closed analytical method [4-6]. In case of analytical method, it demands many computations and most of them do not consider the external noises which happen in real process. Therefore there exists a limit to be used in a real time process directly such as UGV position estimation.

As a position tracking method, we apply the time difference of arrival (TDoA) signal which needs no special equipment. To reduce the computational efforts and to estimate more precisely even for noise added real process, the Kalman filter can be used effectively as an estimator. However, to measure the position of UGV based on TDoA signal, it needs to use an extended Kalman filter (EKF) [7] through linear approximation of localization and TDoA measurement process at each estimation iteration.

In this paper, we use the modified Kalman filter algorithm (adaptive fading Kalman filter: AFKF) [8] for precise position tracking. To do so, we do the system modeling for UGV localization process. Using the adaptive fading factors, the error covariance can be modified to follow the real system model. Finally some simulation results show the effectiveness of the proposed position estimation algorithm through the comparison with EKF.

2. System modeling for UGV localization

2.1. Analytical methods

The signals sent to each base station (BS: the known position) by a mobile station (MS: the unknown position) have a time difference because of the BS's scattered location. The basic concept of position estimation is to use the hyperbolic curves from the definition of TDoA as shown in figure 1. The analytic method to find the target position can be summarized as follows.

Let $\mathbf{p}_i = col\{x_i, y_i\}$, $i = 1, 2, 3, \cdots, m$ be the known locations of m receivers and let $\mathbf{u} = col\{x, y\}$ be the unknown location of the UGV. Using the norm of a difference between the receiver position and UGV, we define as

$$\begin{aligned} r_i^o &= \|\mathbf{u} - \mathbf{p}_i\| \\ r_{i1}^o &= ct_{i1} \end{aligned} \tag{1}$$

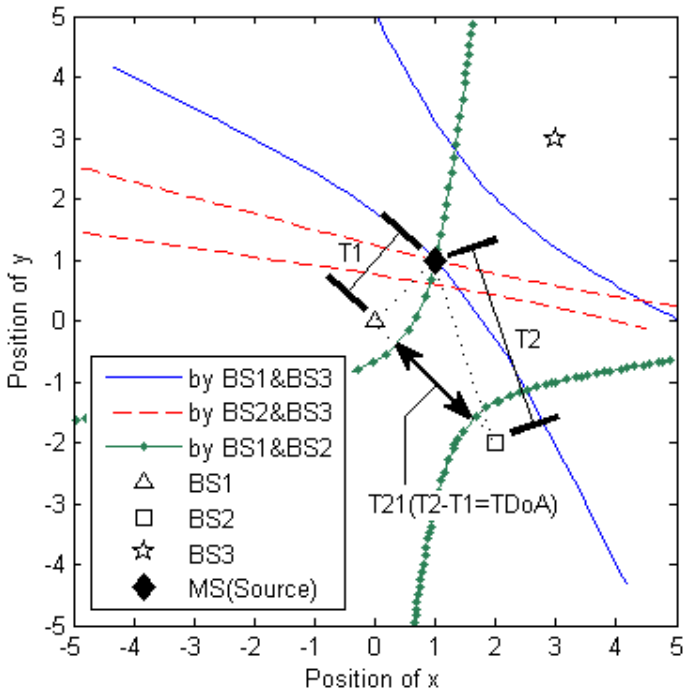


Figure 1. Geometric method using hyperbola.

where t_{i1} is the TDoA between the i-th receiver and the first one, c is the signal propagation velocity. Using the relation in (1), it is easily derived as

$$r_i^o = r_1^o + r_{i1}^o, \ i = 2, 3, \cdots, m. \tag{2}$$

After squaring Eq. (2) and using $(r_i^o)^2 = \langle \mathbf{u}-\mathbf{p}_i, \mathbf{u}-\mathbf{p}_i \rangle$, the equation regarding the UGV location can be expressed in the following nonlinear form.

$$\begin{aligned} \|\mathbf{u}\|^2 - 2\mathbf{p}_i^T\mathbf{u} + \|\mathbf{p}_i\|^2 &= \|\mathbf{u}\|^2 - 2\mathbf{p}_1^T\mathbf{u} + \|\mathbf{p}_1\|^2 \\ &+ 2r_1^o r_{i1}^o + (r_{i1}^o)^2, \quad i = 2, \cdots, m \end{aligned} \tag{3}$$

Eq. (3) can be rewritten as

$$\begin{aligned} \|\mathbf{p}_1\|^2 - \|\mathbf{p}_i\|^2 + (r_{i1}^o)^2 &= 2\langle \mathbf{p}_1 - \mathbf{p}_i, \mathbf{u} \rangle \\ &- 2r_1^o r_{i1}^o, \ i = 2, \cdots, m \end{aligned} \tag{4}$$

Since r_1^o is the distance from the UGV to receiver #1, the solution of $\mathbf{u}(x, y)$ must satisfy the following additional constraint .

$$(r_1^o)^2 = (x - x_1)^2 + (y - y_1)^2 + (z - z_1)^2 \tag{5}$$

The above equations of (4) and (5) can be simplified by placing the position of receiver #1 as the origin of the coordinate system, i.e., $\mathbf{p}_1 = 0$.

$$\begin{aligned} \frac{1}{2}(-\|\mathbf{p}_i\|^2 + (r_{i1}^o)^2) &= -\langle \mathbf{p}_i, \mathbf{u} \rangle - r_1^o r_{i1}^o, \ i = 2, 3, \cdots, m \\ x^2 + y^2 + z^2 - (r_1^o)^2 &= 0 \end{aligned} \tag{6}$$

The nonlinear localization equation can be rewritten in the form of linear matrix equation.

$$\begin{aligned} \boldsymbol{\rho} &= -\begin{bmatrix} r_{21}^o \\ \vdots \\ r_{m1}^o \end{bmatrix}, \mathbf{d} = \begin{bmatrix} \langle \mathbf{p}_2, \mathbf{p}_2 \rangle \\ \vdots \\ \langle \mathbf{p}_m, \mathbf{p}_m \rangle \end{bmatrix}, \mathbf{G} = \begin{bmatrix} \mathbf{p}_2^T \\ \vdots \\ \mathbf{p}_m^T \end{bmatrix}, \\ \mathbf{h} &= \frac{1}{2}\begin{bmatrix} \|\mathbf{p}_2\|^2 - (r_{21}^o)^2 \\ \vdots \\ \|\mathbf{p}_m\|^2 - (r_{m1}^o)^2 \end{bmatrix} = \frac{1}{2}(\mathbf{d} - \boldsymbol{\rho} \bullet \boldsymbol{\rho}) \end{aligned} \tag{7}$$

where $\bullet$ denotes the Hadamar elementwise vector multiplication, $\boldsymbol{\rho}, \mathbf{h}$ and $\mathbf{d}$ are (m-1)-vectors, and $\mathbf{G}$ is an $(m-1)\times 3$ matrix, respectively.

Now the vector $\mathbf{u}$ is then a solution of the following UGV localization problem.

$$\begin{aligned} \mathbf{G}\mathbf{u} &= \mathbf{h} + \boldsymbol{\rho} r_1^0 \\ (r_1^o)^2 &= \langle \mathbf{u}, \mathbf{u} \rangle \end{aligned} \tag{8}$$

2.2. System modeling

To implement the TDoA based localization method using hyperbolic curve under the noise added real situation, it is required to reduce the computational effort to solve in real time process. In real situation, the TDoA is affected by the external noises.

$$t = t_o + \Delta t \tag{9}$$

where t_o is the ideal TDoA and Δt is the added noise. As a robust solution, Kalman filter approach can be more efficient than hyperbolic method if we use TDoA as measurement data.

To apply Kalman filter in a localization problem, the state-space equation needs to be formulated. Therefore the UGV translational problem can be modeled in a discrete form as following.

$$\begin{aligned} s(k+1) &= As(k) + Bu(k) + w(k) \\ A &= \begin{bmatrix} 1 & 0 & \Delta & 0 \\ 0 & 1 & 0 & \Delta \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}, \end{aligned} \tag{10}$$

where $s(k) = [x \ y \ \dot{x} \ \dot{y}]^T$, $u(k)$ is the known velocity of a moving UGV, Δ is the sampling time, and $w(k)$ is the process noise in AWGN. The output equation can be formulated using the measurement of TDoA value.

$$\begin{aligned} z(k) &= h(s(k), v(k)) \\ &= \frac{1}{c}(\| Ms(k) - p_i \| - \| Ms(k) - p_j \|) + v(k) \\ M &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \end{aligned} \tag{11}$$

where p_i is the position of *i-th receiver, and c* is the propagation speed, and $v(k)$ is the measurement noise in AWGN. The output *z(k)* is the TDoA value obtained from the definition. In this model, however, because of the nonlinear terms in the output equation, the modified Kalman filter (extended Kalman filter: EKF) should be used [10]. The EKF algorithm can be used through the linearization of partial differentiation .

$$H_k \approx \frac{\partial h_k}{\partial s}, \quad V_k \approx \frac{\partial h_k}{\partial v} \tag{12}$$

3. Localization using adaptive fading Kalman filter

The extended Kalman filter is a well-known method for position tracking system, but the divergence of estimated result from a modeling error is considered to be a crucial weaknesses. Generally, the dynamic properties and errors are considered together in Kalman filter. However the divergence could happen in any circumstance if the result of real system doesn't match to the ideal value of the estimated model. Especially, for the nonlinear system model based on TDoA data the divergence from the real value can

cause a serious problem. As a solution to prevent the divergence of Kalman filter is the adaptive fading Kalman filter (AFKF), which uses the fading factor in updating the Kalman gain. The basic theory is that the estimated result is regulated by the degree of divergence (DoD).
The EKF algorithm consists of two parts. The summary for EKF is given as

1) Time update (prediction part)
(a) The state projection:
$$\hat{s}_k^- = f(\hat{s}_k, u_k, w_k)$$
(b) The error covariance projection:
$$P_k^- = A_k P_{k-1} A_k^T + W_k Q_k W_k^T$$

2) Measurement update (correcting part)
(a) Kalman gain update:
$$K_k = P_k^- H_k^T [H_k P_k^- H_k^T + V_k R_k V_k^T]^{-1}$$
(b) The error covariance update:
$$P_k = [I - K_k H_k] P_k^-$$
(c) The estimate update with measurement z_k :
$$\hat{s}_k = \hat{s}_k^- + K_k [z_k - h(\hat{s}_k^-, u_k, v_k)]$$
$$\hat{z}_k = h(\hat{s}_k^-, u_k, v_k)$$
$$\phi_k = z_k - \hat{z}_k$$

As stated in (1) time update process, we evaluate the estimate of system state by time flow, and we calibrate the state estimate comparing the differences between a real measured value and an estimate through an estimation modeling in (2) measurement update process.

The EKF itself cannot guarantee a stable accuracy, since it has the possibility of divergence under real circumstances. This happens from the linearization of nonlinear system. The more accurate position tracking can be possible with a precision estimation algorithm. The estimation accuracy can be increased through AFKF by adding an adaptive fading factor. The suboptimal fading factor λ_k adjusts the variance of the predicted state vector.

$$P_k^- = \lambda_k A_k P_{k-1} A_k^T + w_k Q_k w_k^T \tag{13}$$

where $\lambda_k = diag(\lambda_1, \lambda_2, \cdots, \lambda_m)$. If the estimated values get close to the real value, the fading factor becomes $\lambda_k \le 1$. That is, it enters into a steady state process. In other cases, the fading factor λ_k is updated as follows:

$$\begin{aligned} \lambda_{k+1} &= \max\left\{1, \frac{\alpha \cdot tr[N_k]}{tr[M_k]}\right\}, \\ N_k &= C_0 - R_k - H_k Q_k H_k^T \\ M_k &= H_k A_k P_k A_k^T H_k^T \end{aligned} \tag{14}$$

$$C_0 = \begin{cases} \dfrac{\phi_0 \phi_0^T}{2}, & k = 0 \\ \dfrac{\lambda_k \phi_k \phi_k^T}{1+\lambda_k}, & k \geq 1 \end{cases}$$

where α is a positive constant, and $tr[\cdot]$ means a trace of a matrix.

The process of (14) is called as an adaptive fading loop. Through the adaptive fading loop, the divergence degree can be determined and the error covariance P_k can be changed adaptively with the fading factor λ_k. Finally the estimate of the TDoA output ($\hat{z}_k$) can be updated. To confirm the divergence due to the modeling errors, we can define the degree of stable (DoS) as following. If $\| z_k - \hat{z}_k \| \leq \varepsilon$ and $\hat{z}_k - \hat{z}_{k-1} \leq 0$, then $\lambda_{k+1} \leq 1$. In other cases with $\lambda_k > 0$, then $\lambda_{k+1} = \lambda_k - 0.1$. The fading factor λ_k can be adjusted adaptively using DoS.

4. Simulation results

We suppose that the UGV is moving at a constant speed but it changes a direction every 2 sec. Figure 2 shows the UGV circumstance with two receivers. In Fig. 2, the dotted line is the expected moving trajectory of UGV, and the solid line is the real trajectory of UGV. The real trajectory is different from the ideal one because of the measurement noise $w(k)$ and the process noise $v(k)$.

Figure 3 shows the performance of the proposed localization algorithm through the comparison with EKF. The performance is measured in terms of the norm of positioning error, i.e., $\|s(k) - \hat{s}(k)\|$. As shown in Fig. 3, the positioning error is much smaller than that of EKF. It means that the position estimation with AFKF is tracking more precisely to the real value $s(k)$. Figure 4 shows the change of the fading factor λ_k. It changes very steeply to correct the position error from the beginning and the estimate $\hat{s}(k)$ gets close to the real value within ε-neighborhood after 1 sec since the fading factor becomes small.

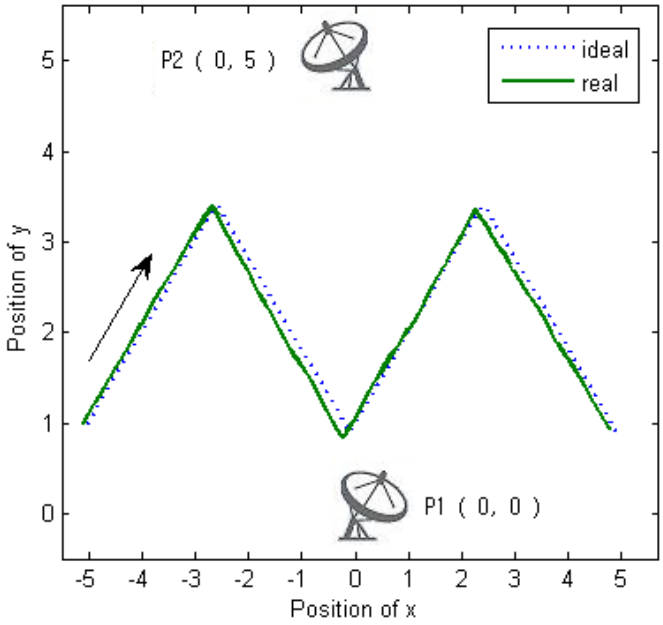


Figure 2. Simulation circumstance for UGV

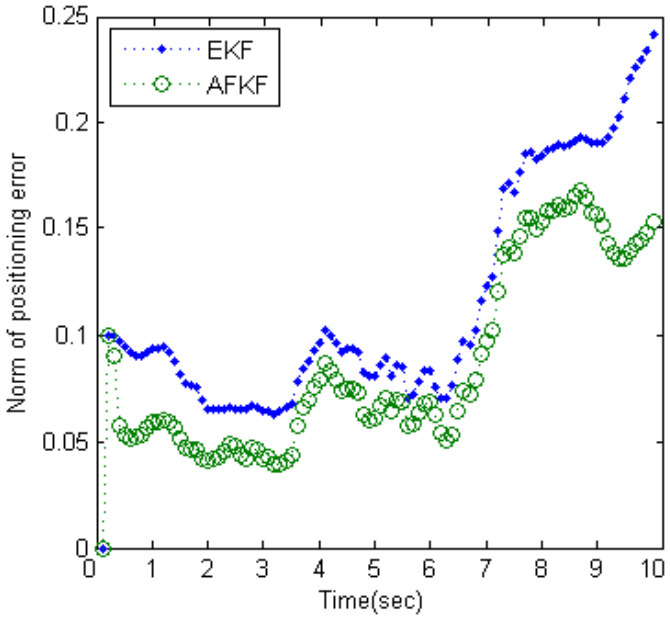


Figure 3. Performance comparison for EKF and AFKF

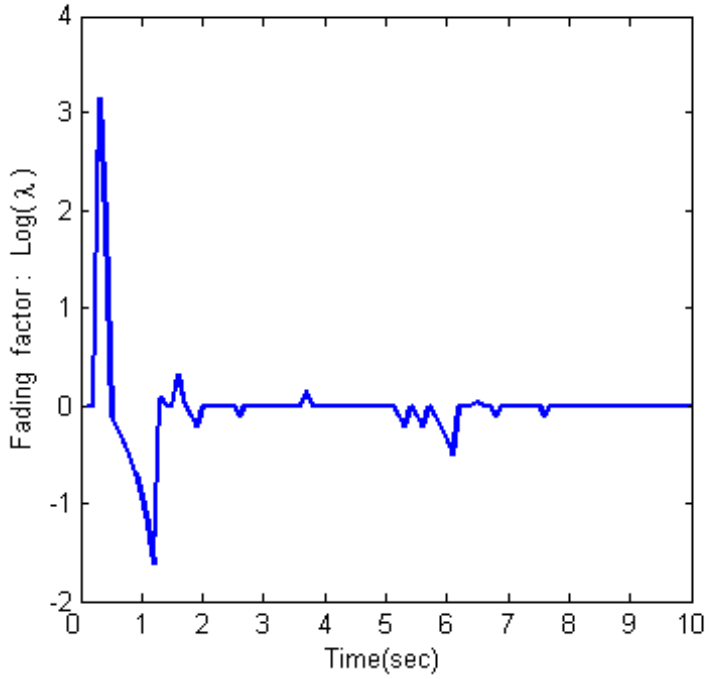


Figure 4. Change of fading factor, λ_k.

Figure 5 shows the results of path estimation for 2 different methods. In this figure, the solid line is a real UGV trajectory, the dotted line is a trajectory estimated by EKF, and the circle-dotted line is the one by AFKF. With the adaptive fading factor, the error covariance has been changed each estimation step. As shown in Fig. 5, the trajectory estimation using AFKF is close to the real value under noise added real circumstance.

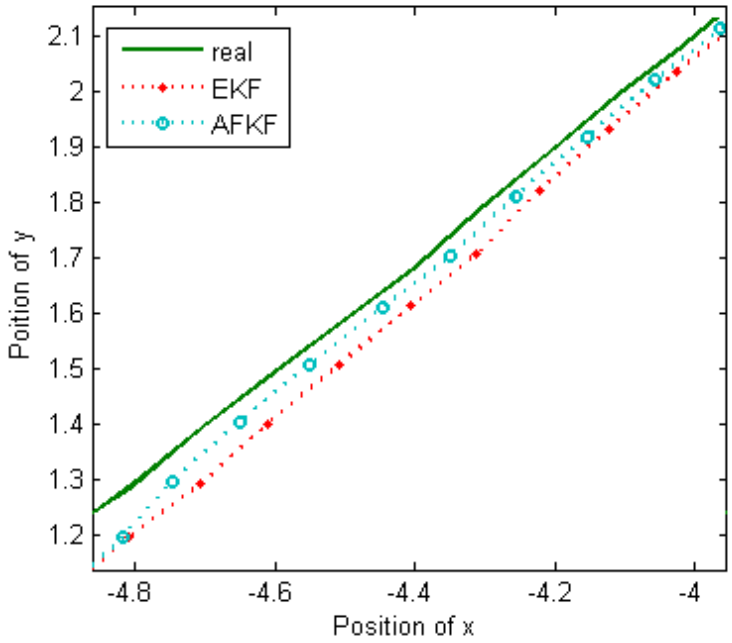


Figure 5. Comparison of trajectory estimation

5. Conclusion

In case of TDoA based position tracking system, it requires much calculation efforts to solve the nonlinear

process. In this paper, we applied EKF to estimate the precise position under real situation. The EKF gets the estimated value through linearization of a nonlinear process. However, there still remains the restriction of divergence if the result of real system is much different from the ideal estimated model which is approximated by partial differentiation.

To overcome the divergence problem in EKF, we applied AFKF algorithm which changes the error covariance using an adaptive fading factor. Through the simulation results, it is confirmed that the trajectory estimation using AFKF follows the real one more precisely. The positioning error from AFKF is less than that performed by EKF.

References

[1] R. Madhavan and C. Schlenoff, "The effect of process models on short-term prediction of moving objects for unmanned ground vehicles," *International IEEE Conf. Intelligent Transportation Systems*, pp. 471-476, 2004.

[2] D.J. Torrieri, "Statistical theory of passive location systems,".*IEEE Tr. on Aerospace and Electronic Systems*, Vol. AES-20, No. 2, pp. 183-197, 1984.

[3] H.K. Kwang, J.G. Lee, C.G. Park, "Adaptive two-stage EKF for INS-GPS loosely coupled system with unknown fault bias", *Jour. of Global Positioning System*, Vol. 5, pp. 62-69, 2006.

[4] K.C. Ho and Y.T. Chan, "Solution and performance analysis of geolocation by TDoA," *IEEE Tr. Aerospace & Electronic Systems*, Vol. 29, No. 4, pp. 1311-1322, 1993.

[5] M. Najar and J. Vidal, "Kalman tracking based on TDOA for UMTS mobile location," *IEEE International Symp. Personal, Indoor and Mobile Radio Communications*, Vol. 1, pp. B45-B49, 2001.

[6] H.C. Schau and A.Z. Robinson, “Passive source localization employing intersecting spherical surfaces from Time-of-Arrival differences,” *IEEE Tr. Acoustics, Speech, & Signal Processing*, Vol. ASSP-35, No. 8, pp. 1223-1225, 1987.
[7] L.J. Levy, “The Kalman filter: navigation's integration workhorse”, *Annual report in Applied Physics Laboratory*, Johns Hopkins University, 1997
[8] Q. Xia, M. Rao, Y. Ying, and X. Shen “Adaptive fading Kalman filter with an application,” *Automatica*, Vol. 30, No. 8, pp. 1333-1338, 1994.
[9] C.P. Gleason, *Tracking human movement in an indoor environment using mobility profiles*, Master’s thesis, University of Nebraska-Lincoln, August, 2006.
[10] S.J. Julier and J.K. Uhlmann, “Unscented filtering and nonlinear estimation,” *IEEE Review*, Vol. 92, No. 3, pp. 401-422, 2004.

Intelligent Control and Automation, 2009, 1, 22-30

Adaptive TLS Approach for Nonlinearity Compensation in Laser Interferometer

S.C. Lee, G.H. Heo, K.H. You
Sungkyunkwan University, Suwon, 440-746, Korea

Abstract

The heterodyne laser interferometer has been widely used in linear displacement and precise measurement field. However the periodic nonlinearity that arises from incomplete laser sources and non-ideal optical components restricts the precise measurement at the nanometer level. In this paper, the total least squares (TLS) algorithm which can obtain optimal compensation parameters of nonlinearity is introduced. Using the TLS algorithm, the nonlinearity error is reduced and the measurement data can be more stabilized. The effectiveness of TLS approach is verified through the comparison of the experimental results with those obtained by a capacitance displacement sensor.

Keywords: *Laser Interferometer, TLS approach, Nonlinearity Compensation*

1. Introduction

Recently the laser interferometry has been widely used in the length-related measurement. The application includes photo lithography in semiconductor manufacture, metrology, and some velocity sensors as well. There are two main kinds of interferometry. The first one is a heterodyne laser interferometer which uses two orthogonal frequencies in a laser head. The other is a homodyne interferometer which uses a single-frequency laser [1]. Between two kinds of interferometry especially the heterodyne laser interferometer is well known for its easy alignment with the optical device, fast response, and high signal-to-noise ratio, etc. However, the measurement accuracy of heterodyne interferometry is restricted by a periodic nonlinearity due to the nonideal laser sources and imperfect optical devices [2-3]. The principal nonlinearity errors happen from frequency mixing, polarization mixing, polarization-frequency mixing, and ghost reflections.

Many researches have been carried out for nonlinearity compensation of a laser interferometer. Theoretical analysis studied by Guo [4] shows that the system nonlinearity arisen from the nonorthogonality and the ellipticity of two-frequency input lights can be compensated using a single phase compensator. Eom [5] proposed a simple digital control system for the frequency stabilization of an internal mirror He-Ne laser. Freitas [6] shows that the second harmonic errors are caused by the two orthogonal linearly polarized inputs of a heterodyne interferometer.

In this paper, we compensate the nonlinearity errors using total least squares (TLS) algorithm with the analytical nonlinearity modeling in [7]. Under the ideal environment the typical configuration for heterodyne interferometry is shown in figure 1. The two frequencies (ω_1, ω_2) are separated through PBS. One of them ($A\omega_1$) proceeds to a fixed-mirror while the

other signal ($B\ \omega_2$) proceeds to a moving mirror. The reflected two frequencies are recombined again through PBS and are detected by a photo detector (B).

In Fig. 1 if there does not exist nonlinearity each electric field can be expressed as follows. (See [8] for detail.)

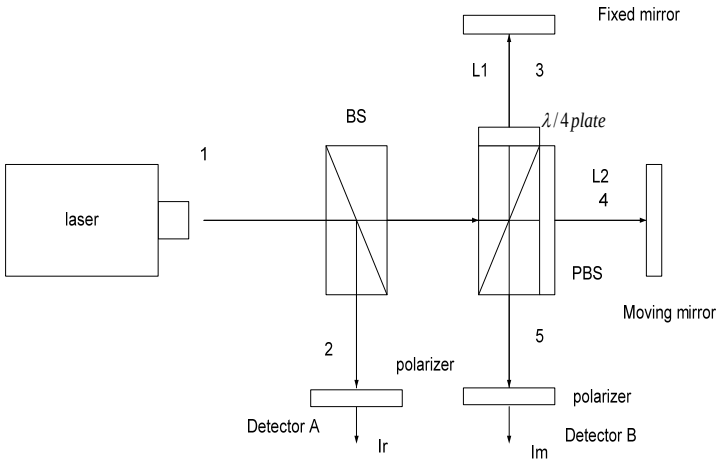


Figure 1. General configuration of heterodyne interferometry: (a) BS (beam splitter),
(b) PBS (polarization beam splitter). (1) $2\pi f_1 t,\ 2\pi f_2 t,$
(2) $A2\pi f_1 t + B2\pi f_2 t,$ (3) $A2\pi f_1 t,$ (4) $B2\pi f_2' t,$ (5) $A2\pi f_1 t + B2\pi f_2' t.$

$$
\begin{aligned}
E_{A1} &= \frac{1}{\sqrt{2}} A e^{j(2\pi f_1 t)} + \frac{1}{\sqrt{2}} A e^{j(\phi_A)} \\
E_{A2} &= \frac{1}{\sqrt{2}} B e^{j(2\pi f_2 t)} + \frac{1}{\sqrt{2}} B e^{j(\phi_B)} \\
E_{B1} &= \frac{1}{\sqrt{2}} A e^{j(2\pi f_1 t)} + \frac{1}{\sqrt{2}} A e^{j(\phi_A)} \\
E_{B2} &= \frac{1}{\sqrt{2}} B e^{j(2\pi f_2 t)} + \frac{1}{\sqrt{2}} B e^{j(\phi_B + \Delta\phi)}
\end{aligned}
\tag{1}
$$

Each intensity of I_r and I_m detected from photo detector *A* and *B* is expressed respectively as follows.

$$
\begin{aligned}
I_r &\propto (E_{A1} + E_{A2})(E_{A1} + E_{A2})^* \\
&= \frac{1}{2}(A^2 + B^2) + AB\cos[2\pi\Delta f t + (\phi_B - \phi_A)] \\
I_m &\propto (E_{B1} + E_{B2})(E_{B1} + E_{B2})^* \\
&= \frac{1}{2}(A^2 + B^2) + AB\cos[2\pi\Delta f t + (\phi_B - \phi_A) + \Delta\phi]
\end{aligned}
\tag{2}
$$

Here A and B are the amplitudes, ϕ_A and ϕ_B are the initial phase values, and Δf is the frequency difference of $f_2 - f_1$. The phase difference between the measurement signal and the reference signal is $\Delta\phi$ which can be transformed to a displacement as

$$\Delta\phi \approx \frac{4\pi n(L_2 - L_1)}{\lambda_m} = \frac{4\pi n\Delta L}{\lambda_m}$$
$$\Delta L = \frac{\Delta\phi\lambda_m}{4\pi n} \tag{3}$$

where λ_m is a mean wavelength, n is a refractive index, and L stands for the displacement measurement.

2. Nonlinearity analysis in heterodyne laser interferometer

The measurement resolution of heterodyne laser interferometer is limited by the unwanted nonlinearity caused by frequency mixing, polarization mixing, frequency-polarization mixing, and ghost reflections. From the detector *B*, the intensities of the electric fields which include every error component can be expressed as following forms [7].

$$\begin{aligned} E_{B1} &= \bar{A}e^{j(2\pi f_1 t)} + \beta_f e^{j(2\pi f_2 t)} + \alpha_p e^{j(2\pi f_1' t)} \\ &\quad + \beta_{pf} e^{j(2\pi f_2 t + \pi/2)} \\ E_{B2} &= \bar{B}e^{j(2\pi f_2' t)} + \alpha_f e^{j(2\pi f_1' t)} + \beta_p e^{j(2\pi f_2 t)} \\ &\quad + \alpha_{pf} e^{j(2\pi f_1' t + \pi/2)} \end{aligned} \tag{4}$$

where $\bar{A}$ and $\bar{B}$ represent $A\cos\phi_1$ and $B\cos\phi_2$, respectively, ϕ_1 and ϕ_2 are defined as the angles between the real polarization and the axes indicated by the PBS. α and β are the cross-reflected beams of Af_1 and Bf_2, respectively. The symbol of a prime means a Doppler-shifted frequency and the small subscripts of *f, p, pf* indicate the nonlinearity from frequency-mixing, polarization-mixing, and frequency-polarization mixing, respectively. To separate DC component we apply a high pass filter. The initial phase value can be disregarded under nonlinearity circumstance. The measurement intensity I_m with no DC components and the disregarded initial phase value are expressed as follows.

$$\begin{aligned} I_m &\propto (E_{B1} + E_{B2})(E_{B1} + E_{B2})^* \\ &= \bar{A}\bar{B}\cos(2\pi\Delta ft + \phi) + (\bar{A}\beta + \bar{B}\alpha)\cos(2\pi\Delta ft) \\ &\quad + (\alpha\beta + \beta_{pf}\alpha_{pf})\cos(2\pi\Delta ft - \phi) \\ &\quad + (\bar{A}\beta_{pf} - \bar{B}\alpha_{pf})\sin(2\pi\Delta ft) + (\alpha\beta_{pf} - \beta\alpha_{pf}) \\ &\quad \times \sin(2\pi\Delta ft - \phi) \end{aligned} \tag{5}$$

where $\alpha = \alpha_p + \alpha_f$ and $\beta = \beta_p + \beta_f$.

The nonlinearity is included in the second and the third cosine terms and the two sine terms. In Eq. (5) the sine terms indicate errors caused by imperfect alignment of PBS. As the

nonlinearity from the optical components is generally smaller than that caused by laser sources [7], we can ignore the partial nonlinearity errors from mirrors and PBS. Therefore if the orientation of PBS is arranged carefully, the last two sine terms in (5) can be eliminated. We use a lock-in-amplifier to obtain a simplified expression with phase information.

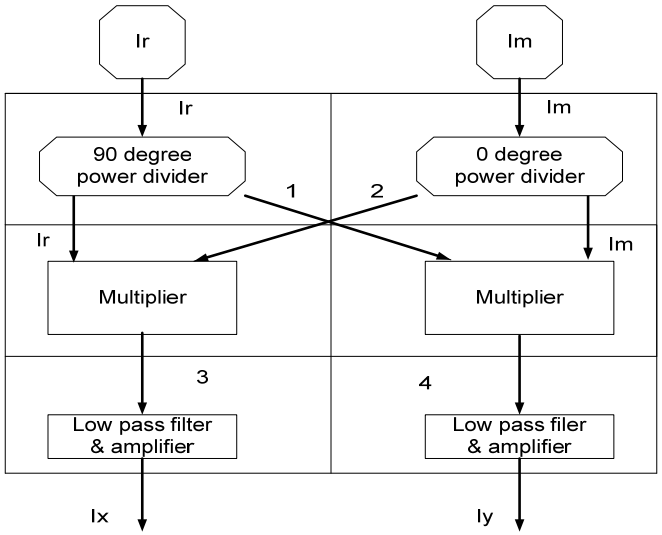


Figure 2. Block diagram of a lock-in-amplifier:
(1) $I_r e^{j\pi/2}$, (2) I_m, (3) $I_m \times I_r$, (4) $I_m \times I_r e^{j\pi/2}$.

Figure 2 shows the configuration of a lock-in-amplifier. Each I_r and I_m from photo detector *A* and *B* is connected to the lock-in-amplifier as a signal input. In this process, to obtain the intensities of $I_x$and I_y, the 90 degree shifted signal of I_r and the original I_r is multiplied by another input signal I_m in (5).

$$\begin{aligned} I_m I_r \approx & \cos(2\pi\Delta ft)[\bar{A}\bar{B}\cos(2\pi\Delta ft+\phi)+(\bar{A}\beta+\bar{B}\alpha) \\ & \times\cos(2\pi\Delta ft)+(\alpha\beta+\beta_{pf}\alpha_{pf})\cos(2\pi\Delta ft-\phi)+ \\ & (\bar{A}\beta_{pf}-\bar{B}\alpha_{pf})\sin(2\pi\Delta ft)+(\alpha\beta_{pf}-\beta\alpha_{pf}) \\ & \times\sin(2\pi\Delta ft-\phi)] \\ I_m I_r e^{j\pi/2} \approx & \sin(2\pi\Delta ft)[\bar{A}\bar{B}\cos(2\pi\Delta ft+\phi)+(\bar{A}\beta+\bar{B}\alpha) \\ & \times\cos(2\pi\Delta ft)+(\alpha\beta+\beta_{pf}\alpha_{pf})\cos(2\pi\Delta ft-\phi) \\ & +(\bar{A}\beta_{pf}-\bar{B}\alpha_{pf})\sin(2\pi\Delta ft)+(\alpha\beta_{pf}-\beta\alpha_{pf}) \\ & \times\sin(2\pi\Delta ft-\phi)] \end{aligned} \tag{6}$$

After passing through a low pass filter the magnitudes of I_x and I_y are expressed as follows.

$$
\begin{aligned}
I_x &= \left(\frac{\bar{A}\bar{B} + \alpha\beta + \beta_{pf}\alpha_{pf}}{2} \right) \cos\phi - \left(\frac{\alpha\beta_{pf} - \beta\alpha_{pf}}{2} \right) \sin\phi \\
&\quad + \frac{\bar{A}\beta + \bar{B}\alpha}{2} \\
&\quad @\delta_1 \cos\phi - \delta_2 \sin\phi + \delta_3 \\
I_y &= \left(\frac{\alpha\beta + \beta_{pf}\alpha_{pf} - \bar{A}\bar{B}}{2} \right) \sin\phi - \left(\frac{\alpha\beta_{pf} - \beta\alpha_{pf}}{2} \right) \cos\phi \\
&\quad - \frac{\bar{A}\beta_{pf} - \bar{B}\alpha_{pf}}{2} \\
&\quad @\mu_1 \sin\phi - \delta_2 \cos\phi - \mu_2
\end{aligned}
\tag{7}
$$

where the parameters of $\delta_1, \delta_2, \delta_3, \mu_1$ and μ_2 represent the constants of nonlinearity compensation considering the frequency mixing, polarization mixing, and frequency-polarization mixing. The sine term in I_x and the cosine term in I_y can be eliminated by adjusting PBS carefully.

3. Adaptive nonlinearity compensation using TLS algorithm

As an estimation method, the TLS algorithm finds a solution of a linear system influenced by errors. TLS algorithm has an excellent performance based on least squares (LS) methods. In the linear equations of $Ax \approx b$ the general LS method finds an optimal solution x^* which minimizes $\left\| b - \hat{b} \right\|$ subject to $\hat{A}x = \hat{b}$. Here $\hat{A}$ and $\hat{b}$ mean the matrix without errors. While the TLS algorithm finds a solution x^* which minimizes $\left\| [A;b] - [\hat{A};\hat{b}] \right\|_F$ subject to $\hat{A}x = \hat{b}$. Here $\| \cdot \|_F$ indicates the Frobenius norm. The singular value decomposition (SVD) [9] is used through TLS method.

In this section we deal with nonlinearity compensation of heterodyne laser interferometer using TLS algorithm. To apply TLS algorithm, we utilize the displacement measurement from a capacitance displacement sensor (CDS) as a reference signal. The measured lengths from CDS are transformed to phase information according to Eq. (3). With the reference phase from CDS, the intensities of I_x and I_y with no nonlinearity errors are represented as follows.

$$
\begin{aligned}
I_x &= \frac{AB}{2} \cos\phi \\
I_y &= \frac{AB}{2} \sin\phi
\end{aligned}
\tag{8}
$$

The intensities ($\hat{I}_x$, $\hat{I}_y$) of a laser interferometer under the nonlinearity errors can be expressed in the same way.

$$\begin{aligned}\hat{I}_x &= \frac{AB}{2}\cos\hat{\phi} \\ \hat{I}_y &= \frac{AB}{2}\sin\hat{\phi}\end{aligned} \tag{9}$$

Since the error terms from imperfect PBS arrangement can be ignored in Eq. (7), we rewrite $\hat{I}_x$ and $\hat{I}_y$ using the measured phase from CDS to remove the unwanted nonlinearity.

$$\begin{aligned}\hat{I}_x &= \left(\frac{\bar{A}\bar{B}+\alpha\beta+\beta_{pf}\alpha_{pf}}{2}\right)\cos\phi+\frac{\bar{A}\beta+\bar{B}\alpha}{2} \\ &@\delta_1\cos\phi+\delta_3 \\ \hat{I}_y &= \left(\frac{\alpha\beta+\beta_{pf}\alpha_{pf}-\bar{A}\bar{B}}{2}\right)\sin\phi \\ &@\mu_1\sin\phi\end{aligned} \tag{10}$$

Here the compensation parameters of δ_1, δ_3, and μ_1 with unwanted nonlinearity are expressed as $\delta_1=(\bar{A}\bar{B}+\alpha\beta+\beta_{pf}\alpha_{pf})/2$, $\delta_3=(\bar{A}\beta+\bar{B}\alpha)/2$ and $\mu_1=(\alpha\beta+\beta_{pf}\alpha_{pf}-\bar{A}\bar{B})/2,$ respectively. The constants of δ_1, δ_3, and μ_1 terms are used to compensate the shifted elliptical Lissajous to a circular trajectory centered at the origin. To obtain the compensation parameters of δ_1, δ_3, and μ_1, TLS algorithm can be used. To use TLS algorithm $\hat{I}_x$ and $\hat{I}_y$ should be expressed in a linear matrix form as follows.

$$\begin{aligned}&Kx=m \\ &K=\begin{bmatrix}\cos\phi & 1 & 0 \\ 0 & 0 & \sin\phi\end{bmatrix},\ m=\begin{bmatrix}\hat{I}_x \\ \hat{I}_y\end{bmatrix}\end{aligned} \tag{11}$$

Where 0 entries in Eq. (11) mean the eliminated $\cos\phi$ term and $\sin\phi$ term under ideal PBS arrangement, and the compensation parameters of x is $x^T=[\delta_1,\delta_3,\mu_1]$. To take a SVD, the above matrix in (11) can be transformed to an augmented one as

$$\bar{K}=\begin{bmatrix}\cos\phi & 1 & 0 & \vdots\hat{I}x \\ 0 & 0 & \sin\phi & \vdots\hat{I}y\end{bmatrix} \tag{12}$$

Using the result of SVD, we can obtain the optimal compensation parameters using TLS algorithm from right singular vector [10].

$$x^*=-\frac{1}{(v_{n+1,n+1})}[v_{1,n+1,}\cdots v_{n,n+1}]^T \tag{13}$$

The compensated intensities of $\hat{I}_x$ and $\hat{I}_y$ can be transformed to optimal one using the compensation values of δ_1^*, δ_3^* and μ_1^*

$$I_x^* = \frac{\hat{I}_x - \delta_3^*}{2\delta_1^*} AB, \quad I_y^* = \frac{\hat{I}_y}{2\mu_1^*} AB \tag{14}$$

The compensated intensities in (14) form approximately an ideal circle with a radius of *AB/2*. Finally the compensated phase (ϕ^*) and the measurement length using arc-tangent can be obtained as

$$\begin{aligned} \phi^* &= \tan^{-1}\left(\frac{I_y^*}{I_x^*}\right) \\ L^* &= \frac{\phi^* \lambda_m}{4\pi n} \end{aligned} \tag{15}$$

4. Experimental results

In this section we demonstrate the effectiveness of TLS algorithm with some experimental results and computer simulation results. The laser head used in experiment is WT307b from Agilent Technologies with 0.63299112 μm wavelength (λ_m). The amplitude of *A* and *B* is set to 1 volt under the experimental circumstance and refractive index (*n*) is 1.00000002665. As a translational stage to be measured in nanometer-scale, the linear piezo-electric transducer (PI: p-621.1CL) is used. We utilize the capacitance displacement sensor (PI: D-100) as a reference signal.

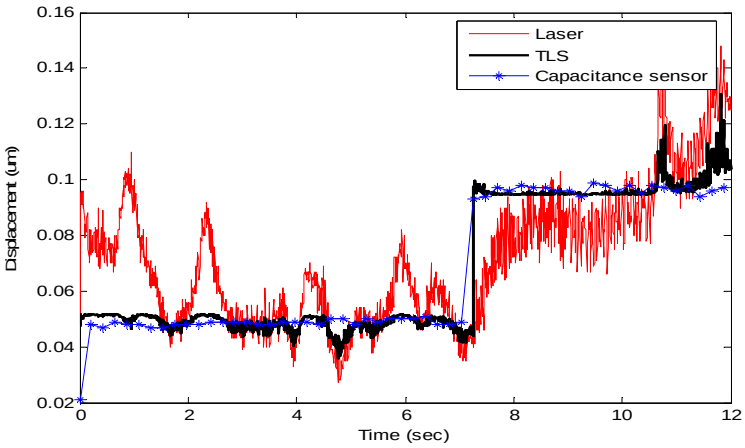


Figure 3. Comparison of fixed positions (*y* = 50, 100 *n*m)

Figure 3 shows the performance of the TLS method compared with CDS and laser measurement according to each fixed position, such as 50 *n*m and 100 *n*m, respectively. The solid line is a displacement measured from laser interferometer with no compensation and the starred line is a reference signal from CDS. The thick line is a compensated displacement

measurement using TLS algorithm. As shown in Fig. 3 we can obtain more stable measurement values which have less chattering on each fixed position after applying the adaptive compensation algorithm.

Figure 4 shows the performance of the TLS method through the computer simulation for a moving stage within the range of $y = 0 \sim 150$ *n*m. Here, the sold line is a displacement measurement including nonlinearity errors, and the thick line is the compensated displacement measurement using TLS algorithm. The reference signal which indicates an ideal displacement measurement is represented as a starred line. As shown in Fig. 4 we can obtain the compensated measurement values with TLS method.

Figure 5 shows the difference between laser displacement measurement and compensated measurement using TLS method for a moving stage with experimental results. In the same way the reference signal from CDS is used as a reference. As shown in Fig. 5, the displacement measurement after TLS compensation is much similar to CDS displacement values.

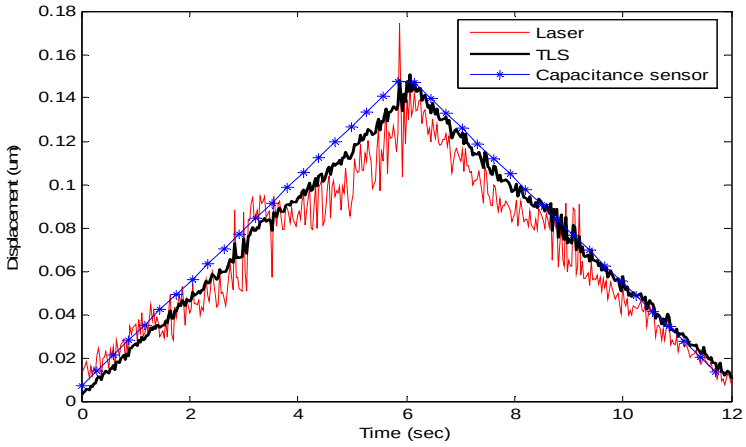


Figure 4. Nonlinearity compensation simulation for a moving stage.

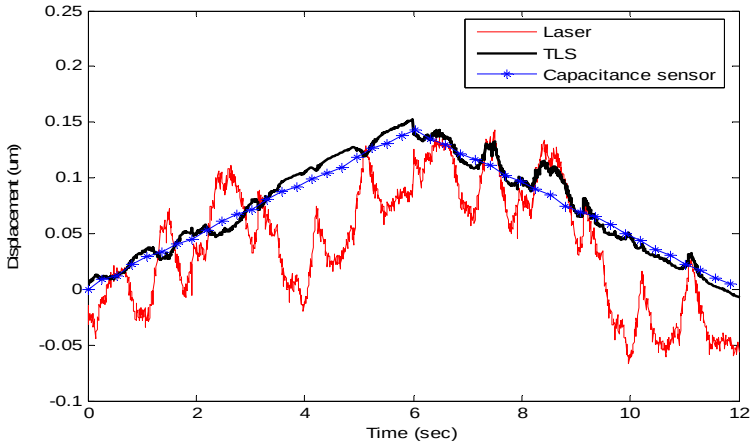


Figure 5. Comparison of a moving stage ($y = 0 \sim 150$ *n*m).

5. Conclusion

As a device of a nano-scale precise displacement measurement, the heterodyne laser interferometer is affected by unwanted nonlinearity errors which include frequency mixing, polarization mixing, frequency-polarization mixing, and ghost reflections. In this paper, an approach using the TLS algorithm for nonlinearity compensation is suggested. To make the problem solving more simple, the error caused by PBS arrangement is ignored. It is shown that the nonlinearity error can be compensated by optimal parameters from TLS method. Also it is represented that TLS method can reduce the chattering through some experimental results. Therefore, the TLS compensation algorithm enables the interferometer robust to nonideal environments.

Acknowledgement

This work was supported by the Korea Research Foundation Grant funded by the Korean Government (KRF-2008-313-D00327).

References

[1] T.B. Eom, T.Y. Choi, and K.H. Lee, "A simple method for the compensation of the nonlinearity in heterodyne interferometer," *Precision Engineering*, Vol. 13, 2002, pp. 222-225.
[2] J. Lawall, E. Kessler, "Michelson interferometry with 10 pm accuracy," *Review of Scientific Instruments,* Vol. 71, 2000, pp.2669-2676.
[3] W. Hou G. Wilkening, "Investigation and compensation of the nonlinearity of heterodyne interferometers," *Measurement Science & Technology*, Vol.7, 1992, pp.520-524.
[4] J. Guo, Y. Zhang, and S. Shen, "Compensation of nonlinearity in a new optical heterodyne interferometer with doubled measurement resolution," *Optics Communications*, Vol. 184, 2000, pp. 49-55.
[5] T.B. Eom, H.S. Choi, and S.K. Lee, "Frequency stabilization of an internal mirror He-Ne laser by digital comtrol," *Review of Scientific Instruments*, Vol. 73, 2002, pp. 221-224.
[6] J.M. Freitas and M.A. Player, "Importance of rotational beam alignment in the generation of second harmonic errors in laser heterodyne interferometry," *Measurement Science & Technology*, Vol. 4, 1993, pp. 1173-1176.
[7] C.M. Wu and R.D. Deslattes, "Analytical modeling of the periodic nonlinearity in heterodyne interferometry," *Applied Optics,* Vol. 37, No. 28, 1998, pp. 6696-6700.
[8] C.M. Wu and C.S. Su, "Nonliearity in measurement of length by optical interferometer," *Measurement. Science & Technology*, Vol. 7, 1996, pp. 62-68.
[9] M. Xia, E. Saber, G. Sharma, and M. Tekalp, "End-to-end color printer calibration by total least squares regression," *IEEE Tr. Image Processing*, Vol. 8, No. 5, 1999, pp. 700-716.
[10] S.V. Huffel and J. Vandewalle, *The total least squares problem computational aspects and analysis*, SIAM, Philadelphia,1991.

Intelligent Control and Automation, 2009, 1, 31-39

The Study of Improving the Accuracy In the 3d Data Acquisition of Motion Capture System

Changho Han, Soonchul Kim, and Choonsuk Oh,
Department of Information and Communications, SunMoon University

Abstract

We introduced the motion capture system using two CCD cameras recently, but could not show any better accuracy than a system using PSD camera. In this paper, we propose two kinds of method to improve the accuracy of the 3D acquisition of the motion capture system using CCD cameras. The applied methods are a distortion removal and z-axis adjustment. We show the result how much the accuracy on the 3D acquisition system improved through comparing with the previous system

Keywords: *3d Data Acquisition, PSD camera, Motion Capture System*

1. Introduction

There are many motion capture systems to obtain 3D data [1] from performances or something moving and the technology of them is growing higher yearly. Specially, the motion capture system of VICON is famous of the precision motion tracking systems, serving customers and CG animation applications in film, visual effects, computer games, and broadcast television. It assisted Image works with installation, provided some custom mo-cap processing tools and provided support on set, which featured over 200 VICON MX40+ cameras in a volume capable of capturing a performer's full facial, body and finger movement, along with various marker props, which included everything from silverware to swords and set pieces. The F40 boasts a resolution of 4 megapixels, captures 10-bit grayscale using 2352 x 1728 pixels, can capture speeds of up to 2,000 frames per second, and can get great data at speeds up to 136,000 markers per second. The typical accuracy of a VICON MX is in the 0.1mm range. However, this equipment's price is very high, so many cheap products is developing now.

In this paper, we show the motion capture system using by active markers and CCD cameras at section 2, and describe the proposed methods to improve accuracy at section 3. The experiments and the result of the efficient on the applied methods will be uncovered at section 4. In the last section, we will discuss about the performance of the proposed methods.

2. Overview of the motion capture system

The existing motion capture system consists of two CCD cameras and the LED markers which has no any cable or radiation control board. The first step in this system is preprocessing to obtain 2D data from the real-time captured stereo images. In the next step, it will be reconstructed 3D data[10] by using disparity algorithm. When 3D

data has extracted successfully from the original images, they will be used in the motion recognition system.

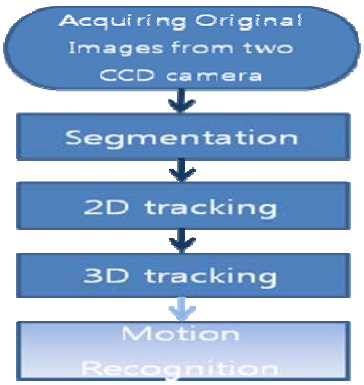


Figure 1. Operating procedure of the motion capture system

2.1 Preprocessing

To distinguish markers point in two obtained stereo images by CCD cameras, we used segmentation algorithm simply. At this time, it should be careful of illumination when the system captures performance's motion. Because the scan range of CCD camera is available in visible radiation area. Unfortunately, under the natural light, it cannot recognize any positions of motions. As you see the figure 2, the original input images are dark. The four points of each image is extracted by segmentation algorithm. Four points will be used the motion recognition system that all work in lab. The system is for boxing so it requires at least four of the data points.

Figure 2. Two original stereo images

If you want to get data points for full body animation, you will need more than four points and have to find correspondence points of markers in the other side to reconstruct 3d data points. In the real situation, it may lose the correspondence points because of some kind of noise. Then, we fail to get 3d points from the two images. When we experiment the acquiring 3d data, we disturbed by the around reflective stuff, natural light, etc.

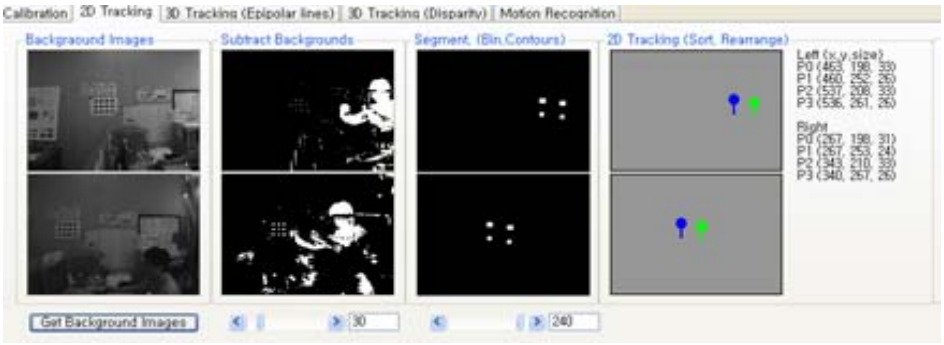


Figure 3. Preprocessing and acquiring 2d data points

As shown in figure 4, the marker consists of LEDs, batteries, and diffuser. It does not have any control hardware system. The diffuser has used to avoid veiling reflections. If a 3d point is not appeared in two images of CCD camera at the same time, it will be ignored, because all points have to have correspondent points in the opposite images. The visible range of marker is within about 125 degree at horizontally.

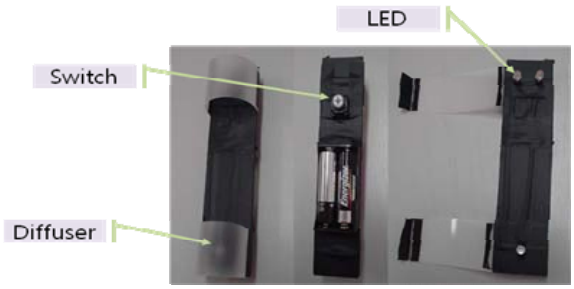


Figure 4. Marker's images

2.2 Calibration

Two cameras are mounted on a bar. There are some essentials conditions. Two cameras point parallel, perpendicular to the bar. Each camera can slide sideways and be clamped. It needs for the facility to make fine pointing adjustments of one camera to correct for minor misalignments.

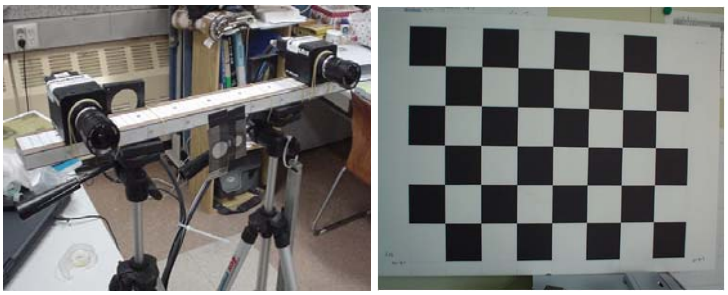


Figure 5. Stereo camera and calibration pattern

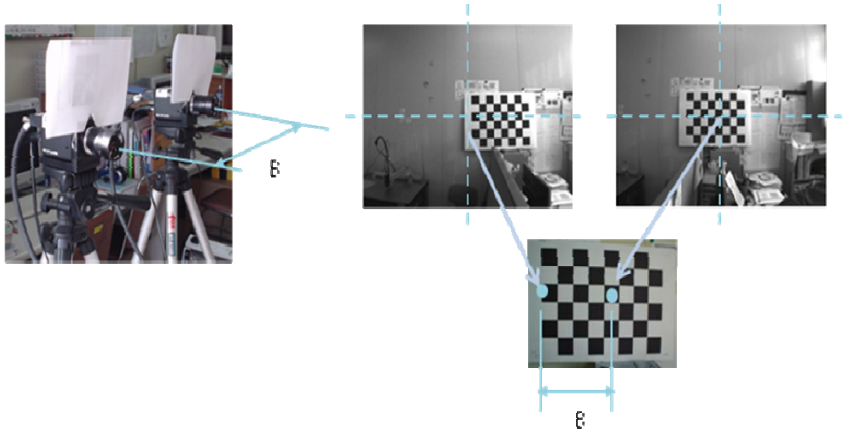


Figure 6. Calibration for parallel setting

To prove two cameras are parallel, we adjust the camera position according to the two points which are the each center point of captured stereo images from chess board. If the distance of two points we found are as same as it between cameras possibly, it will be satisfied the parallel position condition.

2.3 Acquiring 3d data points

After getting 2d data points through the preprocess-ing, it can be calculated 3d data points by using disparity algorithm[2,3,4,5,6,7]. For test, we get 2d data points from the chess board. The number of data points is 35, 7x5 totally. Since a chessboard is used to calibrate a camera[8], this OpenCV function is useful to find where the chessboard is. Finding a chessboard is often a difficult procedure as the contrast between the black and white squares has to be clear and even a little obstruction on the edges of the chess board is enough for OpenCV to fail to find the chess board.

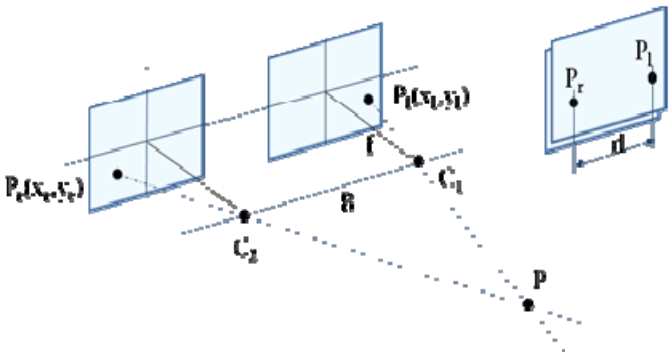


Figure 7. Disparity diagram

Stereo vision can produce a dense disparity map. Lots of researchers have proposed about disparity maps, ex. Area-based stereo methods, window based stereo methods, Bayesian model of stereo matching, cooperative stereo algorithm, etc. However, we utilized a disparity method to 3d disparity space with basic theory as shown in figure 8.

At P(x,y,z) in the real world system, x and y are calculated by average of $P_l(x_1,y_l)$ and $P_r(x_r,y_r)$. The z value can be calculated by disparity algorithm. Of course is not exactly getting

the accuracy depth from that equation. It added some experiment parameters to the equation to get more closed depth data.

3. Proposed algorithms for improving accuracy

There are two kind methods to reduce error of the existing motion capture system. One is to remove distortion of lens, the other is to adjust depth at z-axis.

3.1. Removing distortion

Since lens distortion[11] affects where world points are imaged, the direction of the ray along which a pixel is projected, estimation of lens distortion in a device can significantly improve calibration results, especially for wide-angle lenses. The most important type of distortion is radial, which increases with distance from the center of distortion in the image. The center of distortion is usually located at or near the principal point. In general, the amount of radial distortion is inversely proportional to the focal length of the lens. Most usually assume distortion models that contain radial and tangential distortions. The latter effect is also called decentering distortion [14]. If (x_d, y_d) is a point of a distorted image and (x_u, y_u) is the corresponding point of the undistorted image,

$$x_u = (1 + \sum_{i=1}^{\infty} K_i r^{2i}) x_d + (2P_1 \bar{x}\bar{y} + P_2(r^2 + 2\bar{x}^2))(1 + \sum_{i=1}^{\infty} P_{i+2} r^{2i}) \quad (1)$$

$$y_u = (1 + \sum_{i=1}^{\infty} K_i r^{2i}) y_d + (P_1(r^2 + 2\bar{y}^2) + 2P_2 \bar{x}\bar{y})(1 + \sum_{i=1}^{\infty} P_{i+2} r^{2i})$$

$$x_u = (1 + K_1 r^2 + k_2 r^4) x_d + (2P_1 \bar{x}\bar{y} + P_2(r^2 + 2\bar{x}^2)) \quad (2)$$

$$y_u = (1 + K_1 r^2 + k_2 r^4) y_d + (P_1(r^2 + 2\bar{y}^2) + 2P_2 \bar{x}\bar{y})$$

where $\bar{x} = xd - Cx$, $\bar{y} = yd - Cy$, $r2 = \bar{y}2 + \bar{x}2$. Cx and C_y are the optical center. K_i and P_i are the parameters of the radial and tangential distortions, respectively. Actually, the higher order terms on the right-hand of (1) are ignored. The constants that obtained by OpenCV calibration function in this system are k_1=-0.2457, k_2=0.1113, p_1=0.000987, p_2=-0.0000735.

3.2. Adjust depth

As second method, we adjusted the distance with a scale and ellipse equation that looks like fisheye lens. As you see the table 1, it needs change the distance of measured data at z-axis. The equation 3 shows relation between z' and z in 2d, 3d geometry[9].

$$z' = \sqrt{(z^2 - x^2)}$$
$$z' = \sqrt{(z^2 - x^2 - y^2)} \quad (3)$$

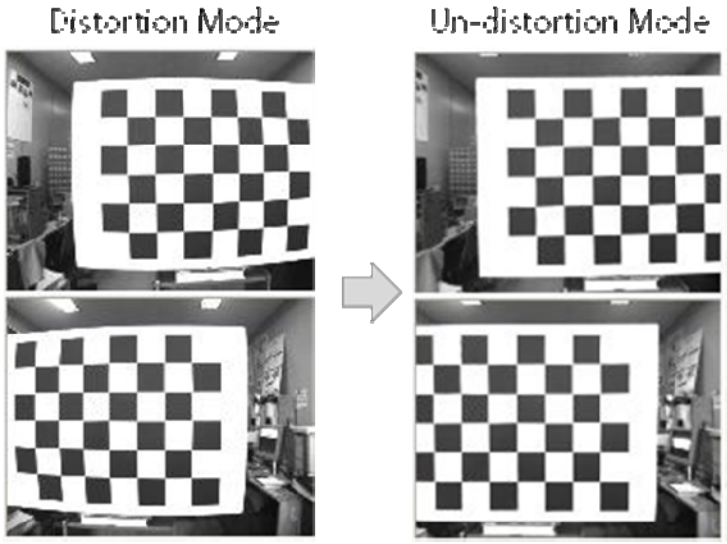


Figure 8. The result of un-distortion image

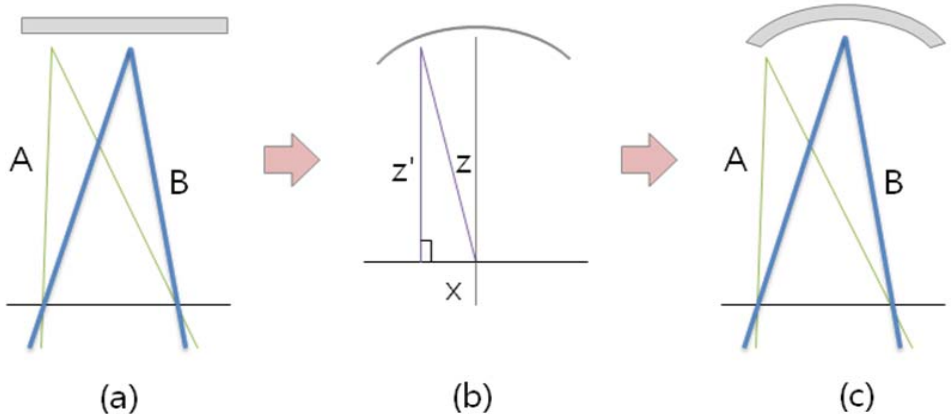


Figure 9. Adjust depth at z-axis

Table 1. The measured and world distance

	Distortion		Un-Distortion	
World Z	Z	Scale	Z	Scale
750	772.6	0.971	732.0	1.025
1000	1003.6	0.996	972.2	1.028
1250	1234.8	1.012	1209.3	1.034
1500	1467.7	1.022	1444.0	1.039
1750	1702.7	1.028	1684.4	1.038
2000	1952.8	1.024	1934.1	1.034
2250	2170.1	1.037	2155.1	1.044

4. Experiments & Result

We tested our method on several different situations, three of which are showed below.

Table 2. The errors of the accuracy of motion capture.

	Removing Distortion			Adjusting Distance			Removing and Adjusting		
Dist(mm)	x	y	z	x	y	z	x	y	z
750	9.4	5.4	26.0	18.2	14.4	30.7	5.9	3.9	87.0
1000	4.1	5.6	40.4	8.2	10.7	26.2	4.5	4.6	64.4
1250	5.7	5.0	57.5	8.1	7.5	28.2	5.5	5.2	66.1
1500	5.2	4.7	64.8	8.7	6.3	38.6	4.2	8.1	59.8
1750	5.4	5.9	78.0	6.2	6.8	45.3	6.6	8.0	50.2
2000	9.7	4.2	92.0	10.8	5.7	49.8	5.9	7.1	72.3
2250	9.7	7.0	116.0	11.0	7.2	72.3	5.5	7.4	66.0
max	9.7	7.0	116.0	18.2	14.4	72.3	6.6	8.1	87.0

According to the our experiments, we can show that the accuracy is exceed the previous system three times when comparing the result of adjusting distance to it of previous system in the Table 3. It does not take into account the calibration error when we hold the chess board by hand. However, If you suppose to use this system for game or animation[12][13] which is not need a large scale area, about 750~1250mm, you can get the accuracy is 20~30mm.

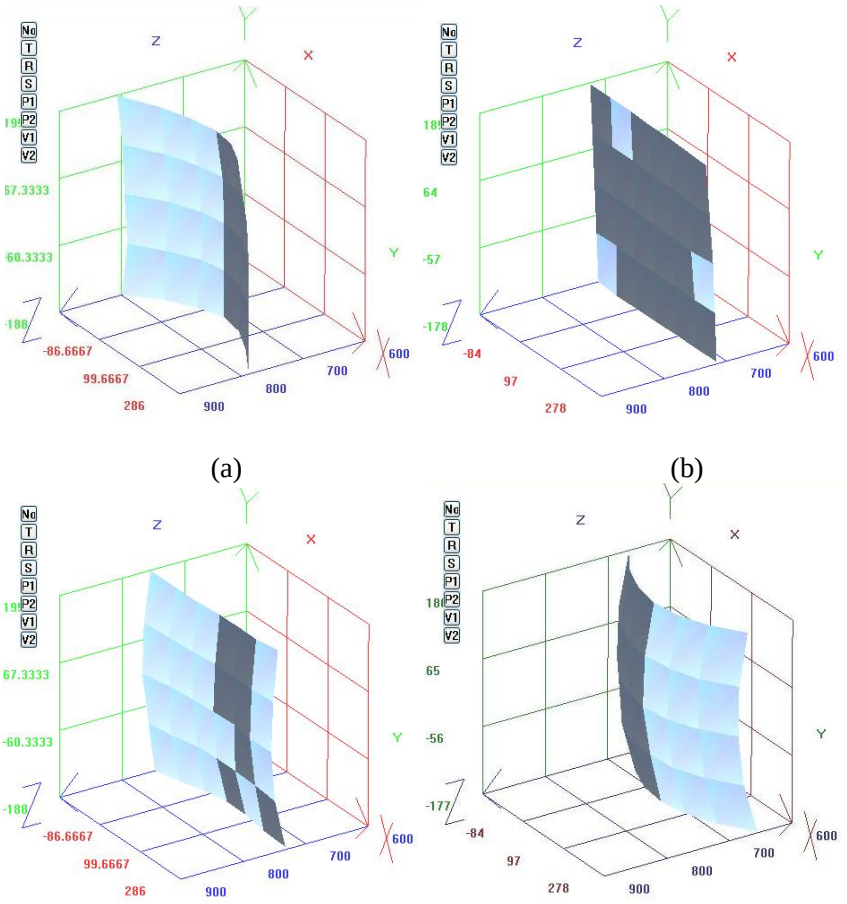
Table 3. The maximum errors of the accuracy of motion capture system.

Measurement Distance(mm)	Previous system	Removing Distortion	Adjusting Distance	Removing +Adjusting
750	96	26	30	87
1000	79	40	26	64
1250	96	58	28	66
1500	123	65	38	60
1750	147	78	45	50
2000	164	92	49	72
2250	179	116	72	66

5. Conclusion and Future work

In this work, the two main methods by which the accuracy for the previous motion capture system are proposed and tested. The method of removing distortion consumes lots of time, because of computing of whole images. It may reduce the estimating time later. The other method of adjusting distance is more effective than it of removing distortion.

One drawback of this system is that it does not robust when capturing motion data, because of natural light. In the next research, we will change the working range of illumination such as infrared ray so that the capturing data is robust.



(a) (b)

(c) (d)

Figure 10. The result of 3d data from chess board; (a) the previous motion capture system, (b) after removing distortion, (c) after adjusting distnace, (d) after removing and adjusting.

Acknowledgement

This work was supported by a Grant-in-Aid for Strategy Technology Development Programs from the Korea Ministry of Knowledge Economy (Robot Component Technical Program; No.10030817).

References

[1] Trucco E., Verri A. 1998. Introductory Techniques for 3-D Computer Vision, Prentice-Hall (eds).
[2] Lucas B., Kanade T. “An Iterative Image Registration Technique with an Application to Stereo Vision”, in *Proc. of 7th International Joint Conference on Artificial Intelligence (IJCAI)*, 1981, pp. 674-679.
[3] Gilbert S´ebastien, Lagani`ere Robert, “Registration of a Moving Rigid Object Using a Stereoscopic Vision Setup”, in *Proc. of the 3rd IEEE International Workshop on Haptic, Audio and Visual Environments and their Applications - HAVE 2004*, 2004, pp. 171-175.
[4] K.Kim, W.Woo, "3D Camera Tracking from Disparity Images," *VCIP*, pp. 1381-1388, 2005.
[5] Y.K.Yu, K.H.Wong, S.H.Or and M.M.Y.Chang, “Recursive recovery of position and orientation from stereo image sequences without three-dimensional structures”, *in Proc. IEEE CVPR,* New York, Jun. 2006.

[6] Kanade, T. 1994. “Development of a video-rate stereo machine”, In Proc. of ARPA Image Understanding Workshop. Monterey, CA, 549-558.
[7] C. L. Zitnick and T. Kanade, "A volumetric iterative approach to stereo matchingand occlusion detection," CMU Technical Report CMU-RI-98-30, 1998.
[8] Chang Shu, Alan Brunton, and Mark Fiala. Automatic grid finding in calibration patterns using Delaunay triangulation. Technical Report NRC-46497/ERB-1104, National Research Council of Canada, Institute for Information Technology, 2003.
[9] Hartley, R. and Zisserman, A. 2000. Multiple view geometry in computer vision. Cambridge University Press, Cambridge, UK.
[10] DiFranco, D., Cham, T., and Rehg, J. 2001. Reconstruction of 3-D figure motion from 2-D correspondences, In Proc. of Computer Vision and Pattern Recognition, Kauai, HI, vol. 1: 307-314.
[11] D. Brown, “Decentering distortion of lenses,” Photogrammetric Eng., vol. 32, no. 3, pp. 444–462, 1966.
[12] W. Frey, M. Zyda, R. McGhee and W. Cockayne, “Off-The-Shelf, Real-Time, Human Body Motion Capture for Synthetic Environments,” Technical Report NPSCS-96-003, Computer Science Dept., Naval Postgraduate School, Monterey, California 93943-5118, USA.
[13] Chai, J., Hodgins, and J. K., "Performance Animation from Low-dimensional Control Signals, " *ACM Transactions on Graphics* (SIGGRAPH 2005).

Intelligent Control and Automation, 2009, 1, 40-49

Hierarchical Role Graph Model for UNIX Access Control

Abderrahim Ghadi[1, 2], Driss Mammass[1], Maurice Mignotte[2], and Alain Sartout[2]

[1] *Irf-Sic–Fsa, Ibn Zohr University, Morocco*
[2] *Irma, University of Strasbourg, France*

Abstract

The access control system is a very important step in the implementation of the security policy of an information system. Access control checks what a user can do directly, as well as what programs executing on behalf of the users are allowed to do. In this way the access control seeks to prevent the activities which will be able to endanger the safety of the system.

The aim of this paper is to try to model the access control system in the operating systems of the type UNIX. The modeling will be based on a combining of the UNIX access control system, namely Super-User model, and RBAC[1] model [1, 11, 12]. In order to get a model nearest possible at reality, we will use the notion of roles and the privileges graph to build our graph. The properties of graph theory well are used in order to evaluate the stability and the robustness of our model.

Keywords: *Privilege, Role, Graph, Hierarchy, Access Control, SuperUser, DAC, MAC, RBAC*

1. Introduction

Safety, and more particularly access control [8, 9], are current problems in data processing [3, 4, 7]. Indeed, it becomes today important to be able to control the floods of information in the networks and the information systems. It is advisable to develop within the computing systems of the mechanisms making it possible to filter the accesses in order to let pass only those authorized. It is a question for that of laying down a security policy, i.e. the characterization of the allowed accesses. Access control is the center of gravity of computer security. Its function is to control which subjects (users, processes, machines, etc.) have access to which resources in the system, which files they can read, which programs they can execute, how they share data with other subjects, and so on.

The aim of this paper is to combine the access control model of UNIX (model of the type DAC[2] [13], based on the access modes and the concept of user-group-other) and RBAC model based on the roles [2]. A frequently asked question is what is the difference between roles and groups? A major difference between most implementations of groups and the concept of roles is that groups are typically treated as a collection of users and not as a collection of permissions. A role is both a collection of users on one side and a collection of permissions on the other one. The role serves as an intermediary to bring these two collections together. The resulting model will be presented in the form of graph [5, 6] which one will release from the mathematical results. The modeling of access control system give a

[1] RBAC: Role Based Access Control
[2] DAC: Discretionary Access Control

clearer vision of security system and consequently limits the maximum the intrusions usually based on transfers of privileges.

2. SuperUser model: UNIX Access Control

We summarize briefly the aspects of UNIX access control. Everything in a UNIX system resides in a file. Thus controlling access to files effectively controls access to software and data on the system. In conventional UNIX systems, the root user (also referred to as super user) is all powerful, with the ability to read and write to any file, run all programs, and send kill signals to any process. For this reason, the standard model of UNIX access control is called SuperUser model.

The objects of importance in the UNIX system are files, and the access modes are read, write and execute. There are three sets of three bits labeling each file description in UNIX (rwx rwx rwx): three bits describe the file owner's privileges, three the group's privileges, and three the privileges assigned to others (which is all users of the system). Within each three bit set, one bit says whether or not the read privilege is granted, one says whether or not write is granted, and the third says whether or not execute is granted. If the privilege is not granted, then a "-" appears, e.g:

- rw- r- - r- - 1 ghadi ghadi 28190 2009-03-01 12:41 UNIX-Admin.pdf

These permissions can be maintained and changed by the UNIX commands chgrp, chmod, chown and umask. In addition to the three modes (read, write, execute) we have three additional access modes: the Setuserid, setgroupid and sticky bits. In a directory listing, setuserid is indicated by an s or S replacing the x in the owner's permissions. The sticky bit is indicated by a t or T replacing the x in the others permissions.

Setuserid status means that when a program is executed, it executes with the permissions of the user who owns the program, in addition to the permissions of the user executing it. The effective user id of the process becomes the id of the owner of the executable file. The real user id of the process remains that of the user who initiated the process. This bit is meaningless on non executable files or on directories.

Setgroupid It behaves in exactly the same way as the setuserid bit, except that the program operates with the permissions of the group associated with the file. When a process is executed with setgroupid bit turned on, the effective group id of the process becomes the group id of the owner of the executable file and the program thus executes with permissions of that group. The real group id of the process remains that of the user who initiated the process. This bit is meaningless on non executable files.

On some systems this bit has a special meaning when set on directories. For example, in SunOS, the group id of a file is set to the group id of the directory in which it is created if the setgroupid is set on the directory. Otherwise, the group id of a file is set to the primary group id that the owner belongs to.

sticky If set on an executable binary file, the 'sticky' bit tells the operating system to maintain the image of the executing process in the swap area, even when execution is terminated. If a directory has its sticky bit set, users may not delete or rename files in this directory that are owned by other users. The sticky bit is usually set on world writable directories.

All the user's files are under the user's home directory /home/$USER, including the startup files and possibly other directories. Therefore, the home directory is the ultimate outer defense against any access to the files of a user. The user account could be easily pirated if the access rights of the home directory are badly managed.

3. Role Based Access Control: RBAC Model

A security policy is a set of rules which specify how to manage, protect or distribute information or resources of a system. In the case of the access control, a security policy is the definition of the authorized accesses. This definition will depend on the concept of entities, of the information of safety available as well as characterization of the access.

Hereafter, the keywords used in the field of the modeling of the information systems:

Definition 1

- Subject: A person or automated agent,
- Object: Any system resource, such as a file, terminal, printer, database record, etc,
- Role: Job function or title which defines an authority level,
- Permission: The ability or the right to perform some action on some resource,
- Operation: A level permission that a resource manager uses to identify security procedures,
- Session: A mapping involving subject, role and/or permission.

One generally distinguishes two big classes from security policies: discretionary policies (DAC) and mandatory policies MAC[3] [13].

The principal difference between a DAC and an MAC is the way in which the information of safety of the objects is modified and created. In the DAC model, each object is under the responsibility of one or more subjects. Change of informations of object security is thus carried out at the discretion of one or more persons in charge. Thus, a subject can potentially have access to any object, provided the persons in charge of this last gives him the permission of this object. Conversely, in the MAC model, Each subject has a set of fixed permissions, which it cannot change.

A typical example of DAC model is the security policy employed in the operating system UNIX (SuperUser Model). The Bell-LaPadula [14] model of access control uses mandatory access control.

The analysis of the security policies and the existing models makes it possible to conclude that the control systems of existing accesses are insufficient. Indeed, discretionary access control present of serious disadvantages with respect to the escapes of information and the Trojan horses, while the obligatory access control very rigid and is badly adapted to the systems really distributed.

In computer systems security, RBAC is an approach to restricting system access to authorized users. It is a newer alternative approach to DAC and MAC. RBAC is a policy neutral and flexible access control technology sufficiently powerful to simulate DAC and

[3] MAC: Mandatory Access Control)

MAC. Conversely, MAC can simulate RBAC when the role hierarchy is restricted to a tree rather than a partial order.

RBAC is a access control system to computer or network resources based on the roles of individual users within an enterprise. In this context, access is the ability of an individual user to perform a specific task, such as create, delete or modify a file. Roles are defined according to job competency, authority, and responsibility within the enterprise. Within an organization, roles are created for various job functions. The permissions to perform certain operations are assigned to specific roles.

In RBAC model:

- Subject can have multiple roles,
- Role can have multiple subjects,
- Role can have much permission,
- Permission can be assigned to many roles.

4. Privileges graph

The privileges graph [15, 16] is a graph whose nodes are roles which represent a set of privileges on a set of objects. So, we propose to model the access control system in the form of this type of graph. Thus, we can improve the security policies by the application of a process with the three different phases:

1. Build the privileges graph. The arc of graph is a transfer method of privileges, in other words, the arc
2. is a mean to acquire privileges,
3. Check if the arcs of the graph are licit,
4. Compare the obtained graph with the hoped security policy.

The nodes and the arcs of the graph are created by the application of the rules which compose the authorization scheme. First we start by inspecting the files */etc/passwd* and /etc/group to check off the list of the users and the groups of the system (Node/Role). Then, several programs and scripts allow us to identify all the existing arcs in our system.

The authorization scheme is considered sure if from an initial protection sure state, we cannot reach a unsure state by the application of the rules of this scheme.

The figure 1 represents a sample example of graph of privileges.

In this example, we have 8 roles which correspond to the privileges of the 8 groups of a system. We have two administration roles: R_admin_1 (administration role whose members are R_5 and R_6) and R_admin_2 (administration role with only one member R_5). An analysis of the file system has to detect transfers of the following privileges: R_1 and R_2 are in the file *$HOME/.shosts* of R_8, what is presented in the graph by arc 1. Likewise, R_3 and R_7 are present in *$HOME/.shosts* of R_4. The inspection of the configuration files special to each user reveals that R_3 uses *.xinitrc* file of R_2 (*.xinitrc* is the file which cements all the process of starting of X). This transfer of privileges is presented here by arc 3. The arc 4 reflected the following situation: «the privilege of a role is a subset of privileges of another role".

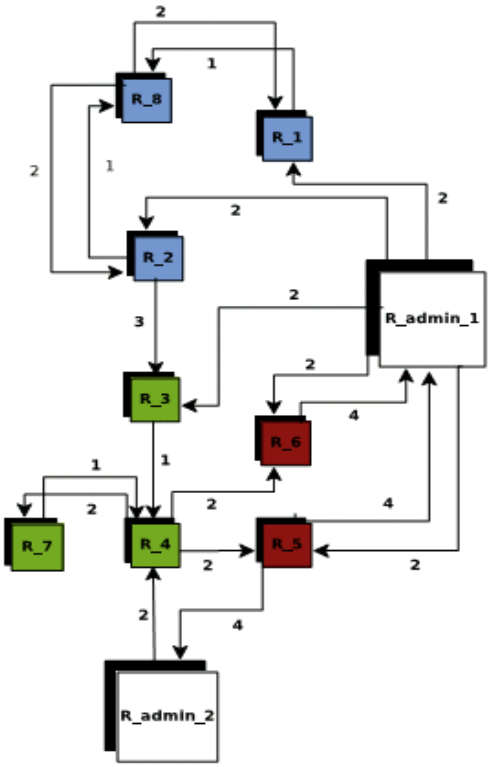


Figure 1. Example of privileges graph

5. Role graph model

Definitions 2

A$\Re$: The set of all access rights of a system;

Privilege: A privilege is a pair p = (*Obj*, A) where *Obj* refers to an object, and A is a non-empty set of access rights for *Obj* (A $\subseteq$ A$\Re$);

Role: A role r is a collection of privileges. r can be represented by a pair (r_{name}, r_{pset}), where r_{name} is the name of r , and r_{pset} represents the set of privileges of r;

$\Re$: Set of roles of a system;

Pv: Set of privileges of a system;

PL: Set of system profiles;

$\Re\Re$: Set of Redundant Roles;

F$\Re$: Father Role;

S$\Re$: Son Role;

UID : Set of identifiers of system users;

GID: Set of identifiers of system groups;

ID = UID $\cup$ GID

G = {g such as g is UNIX group}

$\mathbf{U_g}$ = {u such as u $\in$ g}

Arc One of problems encountered in the system administration of UNIX is that there is no hierarchy between the groups. Consequently, there will be a redundancy of privileges. For this reason, we added the concept of hierarchy of group via the

hierarchy of the roles. In this vision, we regard the arc of the roles graph as methods of transfer of privileges. In other word, an arc between two roles will be a hierarchy relationship. Let r_1 and r_2 two roles, if there is an arc between r_1 and r_2 ($r_1 \curvearrowright r_2$): r_1, we said that r_1 is a son-role of r_2. In this case all the privileges of r_1 will be transmitted to r_2.

5.1. Build of graph

The aim of this section is the build of the elements of the roles graph.

5.1.1 Build of roles

The build of the roles depends on the security policy of the system, and consequently there are more ways of proceeding in this stage. One can give, for example, one of those methods of build of role starting from the groups:

$\forall g \in G\{$
$\forall u \in U_g\{$
extract the privileges of the user u
create a role for this group
$\}$
$\}$

In addition to the roles created via the groups or of the users, there are two special roles to define: R_{root} and C_{ommon}:

R_{root}: Union of all privileges:

$$R_{root} = \bigcup_{p \in P_V} p$$

R_{common} Set of common privileges for all users. For example, the read access to the file */etc/passwd* in order to have the possibility of changing the password.

5.1.2 Build of edges

Definitions 3

1. Let ζ a function enumerating the privileges of a given role:

$$\zeta : \Re \longrightarrow P_V$$
$$r \longmapsto \zeta(r) = r.r_{pset}$$

$r.r_{pset}$ present the privileges of the role r

2. Default privileges (DP): Privileges assigned to a given role during its creation. In other term, are the initial privileges without counting the privileges acquired by hierarchy?

2. Hierarchical privileges (HP): Privileges obtained thanks to the hierarchy of roles [11],

3. Effective privileges (EP): Union of the default privileges of immediate son-roles. Effective privileges of a role r are: $(EP(r) = DP(r) \cup HP(r) = \zeta(r) = r.r_{pset})$

$\forall r_i, r_j \in \Re$. if $\zeta(r_i) \subseteq \zeta(r_j)$ Then an arc between r_i and r_j well be created: $r_i \curvearrowright r_j$

$\forall r_i, r_j \in \Re$, r_j is authorized to accede to the privileges of r_i if and only if $\zeta(r_i) \subseteq \zeta(r_j)$ we say, in this case, that r_i is son-role of r_j.

5.2. Graph properties

Our graph is a graph whose vertexes are roles and edges are the hierarchical relationship between these roles; methods of transfer of privileges between father-role and son-role. The aim of this paragraph is to quote some properties of this graph.

Definitions 4

We say that there exists a path, between two r_i role and r_j and noted $r_i \rightsquigarrow r_j$ if there exists $r_{k_1}, r_{k_2}, \cdots, r_{k_n} \in \Re$ such as $r_i \curvearrowright r_{k_1} \curvearrowright r_{k_2} \curvearrowright \cdots, \curvearrowright r_{k_n} \curvearrowright r_j$ We note the set of all paths in the graph Path, and thus our graph is noted: $(\Re, P_V, P_{ath}, \curvearrowright)$.

Property 1: Reflexivity

$\forall r_i \in \Re \; r_i \curvearrowright r_i$ (A role can be a son-role with himself since $(\zeta(r_i) \subseteq \zeta(r_i))$)

Property 2: Antisymmetry

$\forall r_i, r_j \in \Re$ *if* $r_i \curvearrowright r_j$ *and* $r_j \curvearrowright r_i$ *then* $r_i \equiv r_j$
$(\zeta(r_i) \subseteq \zeta(r_j)$ *and* $\zeta(r_j) \subseteq \zeta(r_i) \Rightarrow \zeta(r_i) = \zeta(r_j))$

This property ensures the no-redundancy of roles.

Property 3: Transitivity

$\forall r_i, r_j, r_k, \in \Re$
if $r_i \curvearrowright r_j$ *and* $r_j \curvearrowright r_k$ *then* $r_i \curvearrowright r_k$
$(\zeta(r_i) \subseteq \zeta(r_j)$ *and* $\zeta(r_j) \subseteq \zeta(r_k) \Rightarrow \zeta(r_i) \Rightarrow \zeta(r_k))$

Figure-2 presents an example of graph of role with a presentation in the form of table (table-1).

Now we can say that our graph. $(\Re, P_V, P_{ath}, \curvearrowright)$ is a partially ordered set (also called a poset).

In the graph of roles, we can have two roles without relation between them; therefore we conclude that our graph is partially ordered set (also called a poset).

- 1-4 it means 1,2,3,4
- Let $\{1-25\} \in P_V$
- The direct privileges of a role are those which are not obtained via transfer of privileges (privileges by defect affected by the administrator),
- The indirect privileges of a role are those which are obtained via transfer of privileges,
- The effective privileges of role are the union of its direct and indirect privileges,
- The path p is such as :

$$R_{commun} \curvearrowright R_2 \curvearrowright R_7 \curvearrowright R_1 1 \curvearrowright R_1 4 \curvearrowright R_{root}$$
$$\Leftrightarrow$$
$$\{1-4\} \curvearrowright \{1-4,7\} \curvearrowright \{1-4,7,8,15\} \curvearrowright \{1-4,7,8,15,21\} \curvearrowright \{1-4,7,8,15,21,26,27\}$$

After this discussion, we can say that our graph is a Lattice. In Mathematics, a Lattice is a poset in which sets of any two elements have a unique supremum (the elements' least upper bound; called their join) an infimum (greatest lower bound; called their meet).

Our graph is too a directed acyclic graph. It means a graph with no directed cycles; that is, for any vertex r, there is no nonempty directed path that starts and ends on r.

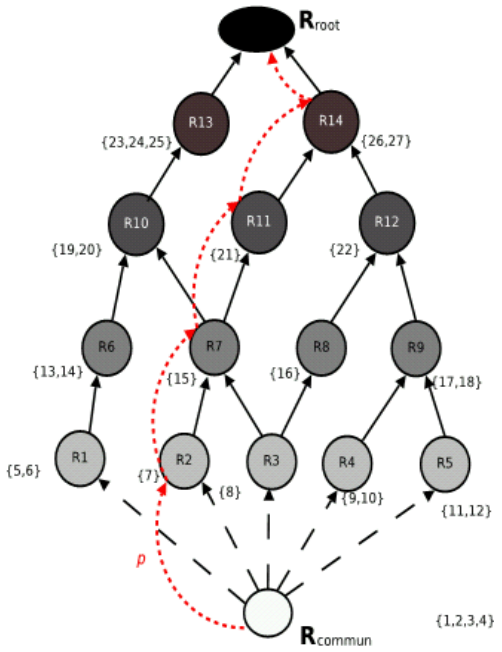


Figure 2. Example of privileges graph

Roles	Direct Privileges	Indirect Privileges	Effective Privileges
R_{common}	1-4	0	1-4
R1	5,6	1-4	1-6
R2	7	1-4	1-4,7
R3	8	1,2,3,4	1-4,8
R4	9,10	1-4	1-4,9,10
R5	11,12	1-4	1-4,11,12
R6	13,14	1-6	1-6,13,14
R7	15	1-4,7,8	1-4,7,8,15
R8	16	1-4,8	1-4,8,16
R9	17,18	1-4,9-12	1-4,9-12,17,18
R10	19,20	1-8,13-15	1-8,13-15,19,20
R11	21	1-4,7,8	1-4,7,8,21
R12	22	1-4,8-12,16-18	1-4,8-12,16-18,22
R13	23,24,25	1-8,13-15	1-8,13-15,23,24,25
R14	26,27	1-8,13-15,19,20	1-8,13-15,19,20,26,27
R_{root}	$P_V - \{1-25\}$	1-25	P_V

Table 1. Transfer of privileges for role-graph of figure 2.

5.1. Role Graph Algorithms

We have developed algorithms to:

- add a role giving its direct privileges, expected juniors and seniors
- add a role giving effective privileges
- add/delete a privilege to/from a role
- add/delete an edge

Creation of the roles starting from the profiles

$\forall p_l \in P_L$
Create role r_{p_l}
assigned permissions to r_{p_l}
add r_{p_l} to $\Re$

Addition of the redundant roles in $\mathfrak{RR}$

$\forall r_1 \in \mathfrak{R}$
{
  $\forall r_2 \in \mathfrak{R}$
    {
      if $(\zeta(r_1) = \zeta(r_2))$ { Add r_1 et r_2 into $\mathfrak{R}$ $\mathfrak{R}$}
      if $(\zeta(r_1) \subset \zeta(r_2))$ { Add r_1 into $CR(r_2)$ }

    }
}

Creation of the heritage relation $\mathfrak{RR}$

$\forall r \in \mathfrak{R}$ {
  $\forall r_s i \in CR(r)$ {
    Add the heritage relation between r and $r_s i$
  }
  Delete the redundant permissions
}

8. Conclusion

In terms of access control, the question which often arises is : What are the users who have permission to access to a given service ?. Modeling the access control system in the form of oriented and hierarchical graph allows visualizing the impacts of the modifications of permissions of a role. The search for all possible ways in the graph makes it possible to filter the transfers of privileges. This research is perfectly feasible by using the algorithms of the theory of graphs.

The model is not of course complete, there is still again a lot of work in order to improve it. There are two essential extensions required to supplement SuperUser model in a UNIX environment: 1. the system file permissions must be modeled, 2. the links between files must be modeled too.

References

[1] D.F. Ferraiolo, R. Kuhn, R. Sandhu (2007), "RBAC Standard Rationale: comments on a Critique of the ANSI Standard on Role Based Access Control', IEEE Security & Privacy, vol. 5, no. 6 (Nov/Dec 2007), pp. 51-53.
[2] J. B. D. Joshi, E. Bertino, and A. Ghafoor, "Formal Foundations for Hybrid Role Hierarchy," ACM Transactions in Information and Systems Security, in print for Nov. 2007.
[3] ISO/IEC 27001:2005, Requirements for Information security management systems, 2005.
[4] ISO/IEC 27002:2005, Code of practice for information security management, 2005.
[5] O. M. Sheyner. Scenario graphs and attack graphs, Thesis of School of Computer Science, Computer Science department, Carnegie Mellon University, Pittsburgh, PA, 2004.
[6] S. Jha, O. Sheyner and J. Wing, Two formal analyses of attack graphs, Computer Security Foundation Workshop, 2002.
[7] G. Vache, Towars Information System Security Metrics, European Dependable Computing Conference 7, Proceedings Supplemental Volume, pp 41-44, Kaunas, 7-9 May 2008.
[8] A..Ghadi, D. Mammass, M. Mignotte and A.Sartout, «Formalism of the access control model based on the Marked Petri Nets". International Journal of u- and e- Service, Science and Technology Vol.1, No.2, Mars, 2009.

[9] M Jaume and C Morisset. “A formal approach to implement access control”, .Journal of Information Assurance and Security, 2:137–148, 2006.
[10] Feng Xiaoning;Wang Zhuo; Yin Guisheng; “HierarchicalObject-Oriented Petri NetModelingMethod Based on Ontology” Internet Computing in Science and Engineering, 2008. ICICSE ’08. International Conference on 28-29 Jan . 2008 Page(s):553-556.
[11] S. Gavrila, J. Barkley, "Formal Specification for Role Based Access Control User/Role and Role/Role Relationship Management" (1998), Third ACM Workshop on Role-Based Access Control.
[12] J. Barkley, "Implementing Role Based Access Control Using Object Technology", First ACM Workshop on Role-Based Access Control (1995).
[13] P. Samarati1 and S. de Capitani di Vimercati, “Access Control: Policies, Models, and Mechanisms“, Foundations of Security Analysis and Design, Springer Berlin / Heidelberg, 2001 Page(s):137-196.
[14] D. Bell and L. LaPadula, ldquoSecure Computer Systems: unified Exposition and Multics Interpretation, rdquo Tech. Rep. MTR-2997,MITRE Co., July 1975.
[15] Paul Ammann, Duminda Wijesekera, and Saket Kaushik, “Scalable, Graph-Based Network Vulnerability Analysis,” Proceedings of the 9th ACMConference on Computer and Communications Security, 2002, pp. 217–224.
[16] H. Ehrig, G. Engels, H.-J. Kreowski, and G. Rozenberg, editors. Handbook of Graph Grammars and Computing by Graph Transformations. Vol. II: Applications, Languages, and Tools. World Scientific, 1999. 236,242.