The 2010 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 18-22, 2010, Taipei, Taiwan

Relative Posture Estimation Using High Frequency Markers

Yuya Ono, Yoshio Iwai and Hiroshi Ishiguro

Abstract— **Recently, research fields of augmented reality and robot navigation are actively investigated. Estimating a relative posture between an object and a camera is an important task in these fields Visual markers are frequently used to estimate a relative posture between an object and a camera, but the usage of visual markers spoils a scene. In this paper, we propose a novel method for posture estimation by using kernel regressions and high frequency markers that do not spoil a scene. The markers are embedded in an object's texture in the high frequency domain and it is hard for people to notice the markers. We observe changes of spatial frequency of object's texture by using a high-definition camera to estimate the current posture of an object. We show the effectiveness of our method by experimental results with both simulated and real data.**

I. INTRODUCTION

In recent years, research fields of augmented reality and robot navigation are actively investigated. In augmented reality, many methods have been proposed to provide useful information to users by CG images superimposed on a real images at geometrically right position[1]. In robot navigation systems, visual information from a camera is very important because a camera is a passive sensor. Therefore, there are many researches that estimate a current position and/or a movement path from an image sensor[2]. In the above research fields, estimating a relative position/posture between a camera and an object is still important problem.

Various methods for estimating a relative posture have been proposed until now. These methods are separated into two types: vision based[3] and special purpose sensor based method[4], [5]. GPS, gyro sensor, and magnetic sensor are often used as a special purpose sensor for estimating a posture. Kanbara et al. proposed a system using a GPS and a small inertial navigation sensor to compensate the disadvantages of both sensors[4]. Matsuda et al. proposed a system using a small inertial navigation sensor and a vision sensor[5]. In general, a method using multiple sensors achieves the high accuracy of posture estimation, but a systems is tend to be complicated.

Vision based approach estimates a relative posture from visual information retrieved by feature extraction and/or edge detection. Oe et al. proposed a method for posture estimation by using natural feature points from an omnidirectional image sequence[3]. Lepetit et al. proposed a method for posture estimation by using the correspondence between image feature points and those of a CAD model[6]. The above methods, however, use prior knowledge of an object and/or environment, so it is difficult to apply the methods in unknown environment.

Marker tracking approach is also used for posture estimation of an object. Kato et al. proposed a method for posture estimation in real time by using rectangle marker in an image[7]. As an application of this research, Habara et al. proposed a system that provides user's location in an indoor environment[8]. Marker tracking approach requires that visual markers should be fixed at an object, so it spoils a scene. To avoid the above problem, recursive reflection materials are used as invisible markers[9], but the cost of the materials is still a problem. Tenmoku et al. proposed a method for posture estimation by using a special designed poster[10]. Sato et al. proposed a coded wall paper for posture estimation[11]. These methods take into consideration of scene, but the design of posters or wall papers is highly restricted.

In this paper, we propose a novel method for posture estimation by using kernel regressions and high frequency markers that do not spoil a scene. The markers are embedded in an object's texture in the high frequency domain and it is hard for people to notice the markers because of the ability of human eyes. We observe changes of spatial frequency of object's texture by using a high-definition camera to estimate the current posture of an object. We show the effectiveness of our method by experimental results with both simulated and real data.

II. PROBLEM SETTINGS

In this research, relative posture between a camera and a object is estimated from a image embedded visual markers in the high frequency domain. We explain the problem settings in this section. As shown in Fig. 1, relative posture of an object in the marker coordinate system $O_m - X_m Y_m Z_m$ can be represented by the translation matrix dT and the rotation matrix parametrized by rotation angles $(\theta_x, \theta_y, \theta_z)^T$ in the camera coordinate system $O_c - X_c Y_c Z_c$. In this research, we assume that object rotation can be decomposed as $R = R_x(\theta_x)R_y(\theta_y)R_z(\theta_z)$, and we also assume that the initial posture is $(\theta_x, \theta_y, \theta_z)^T = (0,0,0)^T$. d is the relative depth between the current posture and the initial posture.

The proposed method can estimate the relative rotation parameter $(\theta_x, \theta_y, \theta_z)^T$ and the relative depth d from visual changes of an object in the high frequency domain caused by posture change of an object embedded visual markers.

Graduate School of Engineering Science, Osaka University, 1–3 Machikaneyama, Toyonaka, Osaka 560–8531 Japan
iwai@sys.es.osaka-u.ac.jp

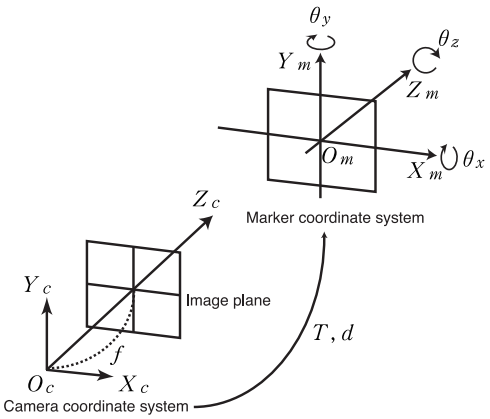

Fig. 1. Coordinates and problem settings

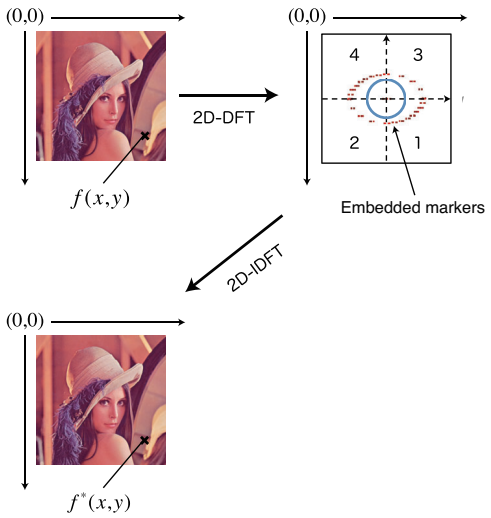

Fig. 2. Overview of marker embedding

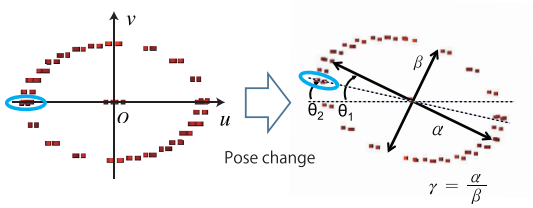

Fig. 3. Features in the frequency domain

III. RELATIVE POSTURE ESTIMATION

A. Marker Embedding

The outline of marker embedding is shown in Fig. 2. When an image is transformed with 2D Fourier transform, the lowest frequencies usually contain most of the information. We embed visual markers in the high frequency domain, which does not influence its appearance so much because of the ability of the human eye. Visual markers are embedded in an elliptical shape in a certain interval in the high frequency domain as shown in Fig. 2, and then a marker image is acquired by the inverse 2D Fourier transform. Relative posture is estimated by observing frequency changes in a marker image caused by object's posture changes. Unlike the conventional researches[13], [14], [15], we need not the assumption or the prior knowledge of characteristics of object texture because of the above approach. Therefore, the proposed method can use printed materials embedded visual markers as marker images without the knowledge of texture information of the printed materials. Moreover, our method uses high frequency domain which is insensitive for human eye, and has the advantage that marker images used in our method do not spoil a scene.

B. Image Feature

As described in the previous section, visual markers are embedded in a elliptical shape in the high frequency domain of a marker image. Let α and β be lengths of the axes of the ellipsoid, and θ_1 and θ_2 be a rotation angle and a phase of the ellipsoid, respectively. α and β are parameters sensitive to a relative distance between a camera and an object, so we introduce the parameter $\gamma = \frac{\alpha}{\beta}$ which is robust for distance changes. Five parameters, $\alpha, \beta, \gamma, \theta_1 and \theta_2$, are used as image features as shown in Fig. 3.

The process flow of the proposed method is shown in Fig. 4. Figure 5 shows an example of power spectrum of a marker image, and Fig. 6 shows an result of image processing for marker detection. Ellipse detection is performed by energy minimization of the following equation:

$$Q^2 = (a \cdot x^2 + b \cdot y^2 + c + 2f \cdot x + 2g \cdot y + 2h \cdot xy). \quad (1)$$

α, β, and θ_1 can be calculated from the above parameters.

Phase parameter θ_2 can be estimated from the maximum angle of the correlation between an ellipsoid in an input image and one of the initial posture. We consider an indicator function $s(\theta)$ that becomes 1 if a marker exists in θ direction and becomes 0 if a maker does not exist in θ direction. Let $s_1(\theta)$ be an indicator function of the initial posture and $s_2(\theta)$ be an indicator function of the current posture. The correlation between $s_1(\theta)$ and $s_2(\theta)$ is defined and the maximum correlation angle is estimated by the following equation:

$$\theta_2 = \arg\max_{\Delta\theta} \int_0^{2\pi} s_1(\theta + \Delta\theta) \cdot s_2(\theta) d\theta. \quad (2)$$

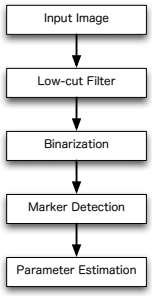

Fig. 4. Process flow of parameter estimation

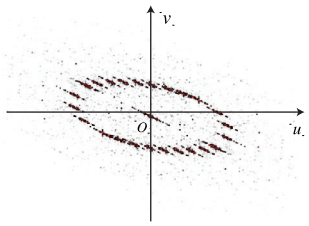

Fig. 5. Power spectrum

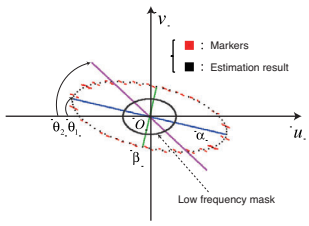

Fig. 6. Estimation result

Even if the scales of two ellipsoids are different with each other, we can estimate θ_2 because of the indicator function $s(\theta)$.

C. Relative Posture Estimation by Kernel Regressions

Relative posture, $\theta_x, \theta_y, \theta_z$, between a camera and an object is estimated from image features, $\alpha, \beta, \gamma, \theta_1, \theta_2$, by kernel regressions that map image features to posture parameters. In this paper, we use a support vector regression (SVR)[16] as a kernel regression. SVR is expressed by the following equation:

$$y = f(\mathbf{x}) = \sum_{i=1}^{D} w_i K(\mathbf{x_i}, \mathbf{x}) + b, \tag{3}$$

where D is the number of dimension of image features, w_i $(i = 1, \cdots, D)$ is a weight, and b is an offset. We use a polynomial kernel as a kernel function K:

$$K(\mathbf{x}, \mathbf{y}) = (s \cdot \mathbf{x}^T \mathbf{y} + o)^d, \tag{4}$$

where s, o, and d are hyper parameters and are determined by a preliminary experiment. SVR outputs the displacements $\Delta\theta_x, \Delta\theta_y$, and $\Delta\theta_z$ from the previous posture $\theta_{x,t-1}, \theta_{y,t-1}, \theta_{z,t-1}$ and image features: $\gamma, \theta_1, \theta_2$, and then update the current posture $\theta_{x,t}, \theta_{y,t}, \theta_{z,t}$

The relative depth between a camera and an object can be estimated from the previous position because of the following equation:

$$\frac{\alpha_1}{\alpha_2} = k \cdot \frac{d_1}{d_2}, \tag{5}$$

where d_1 and d_2 are depths in the previous and current posture, respectively. α_1 and α_2 are image features in the previous and the current posture, respectively.

IV. EXPERIMENTAL RESULTS

A. Hyperparameters

In order to use SVR for posture estimation, hyperparameters s, o, and d should be determined in advance. Therefore, we generated synthesized images rotated at random with $\theta_x, \theta_y, \theta_z \in \{0, 10, 20, \ldots, 50, 60\}$ [deg], extracted image features from the images, and then applied SVR learning algorithm to the training data.

From this preliminary experiment, the number of dimension d has a larger effect to the training error than another parameters, so s and o are fixed to be 1. Figures 7 and 8 show the results of the preliminary experiment. From this result, the number of dimension d is fixed to be 5 for posture estimation, and fixed to be 4 for depth estimation in experiments in the following section.

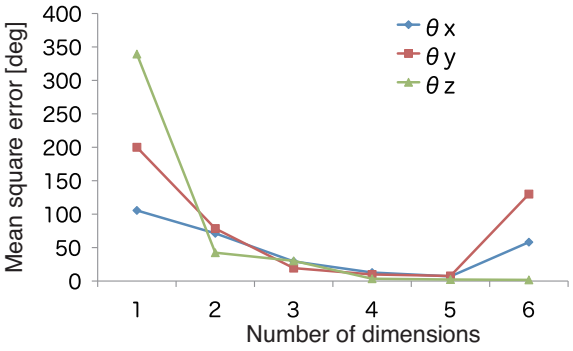

Fig. 7. Training errors of $\theta_x, \theta_y, \theta_z$

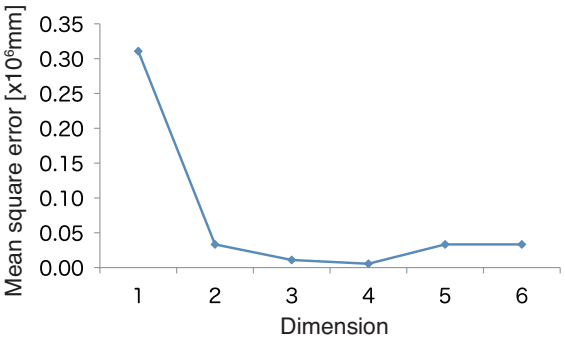

Fig. 8. Training error of d

B. Experiment with Simulation Data

Table II shows estimation errors (generalization errors) by using training data in Table I. From this result, estimation errors are about 5 degrees and are enough for posture estimation.

In order to estimate various posture changes, we make sequences of synthesized images rotated at random with $\theta_x, \theta_y \in [-60, 60], \theta_z \in [-180, 180]$, and then SVR is learned by this training data. Figures 9,10 and 11 show estimation results of one sequence. RMSE of $\theta_x, \theta_y, and \theta_z$ in this experiment are 3.08, 2.63, and 2.28, respectively. From the

TABLE I
TRAINING DATA

Training data	range of posture
Set 1	$\theta_x \in [0,60], \theta_y \in [0,60], \theta_z \in [0,60]$
Set 2	$\theta_x \in [0,60], \theta_y \in [0,60], \theta_z \in [-60,0]$
Set 3	$\theta_x \in [0,60], \theta_y \in [-60,0], \theta_z \in [0,60]$
Set 4	$\theta_x \in [0,60], \theta_y \in [-60,0], \theta_z \in [-60,0]$

TABLE II
ESTIMATION ERROR

parameter	rooted mean square error [deg]
θ_x	4.00
θ_y	4.19
θ_z	1.04

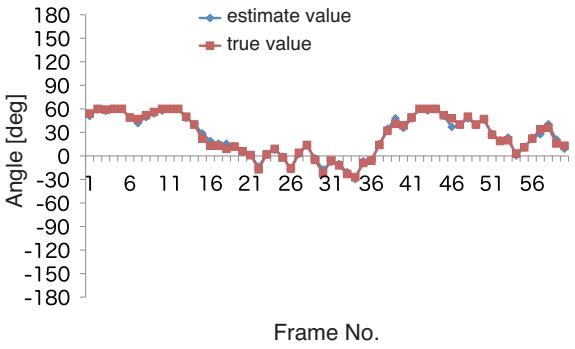

Fig. 9. estimation results of θ_x

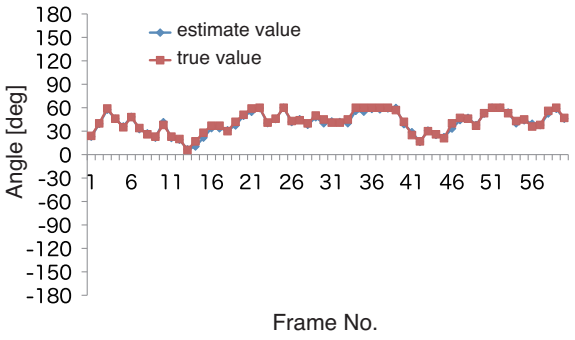

Fig. 10. estimation results of θ_y

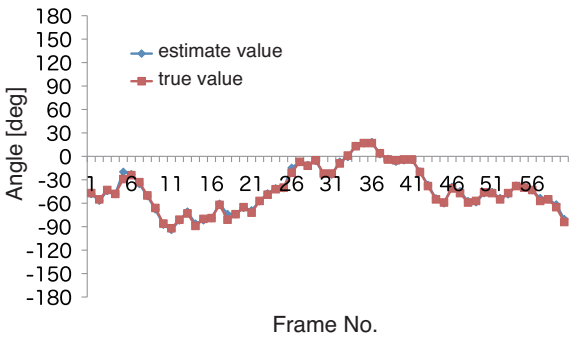

Fig. 11. estimation results of θ_z

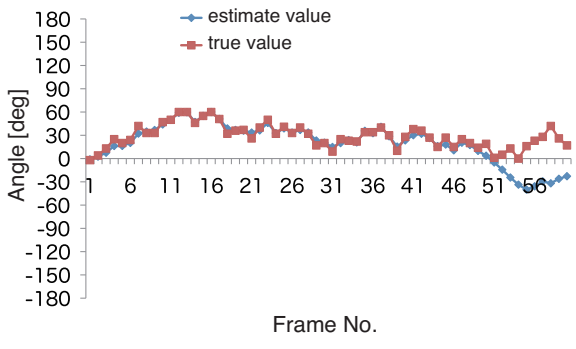

Fig. 12. estimation results of θ_x

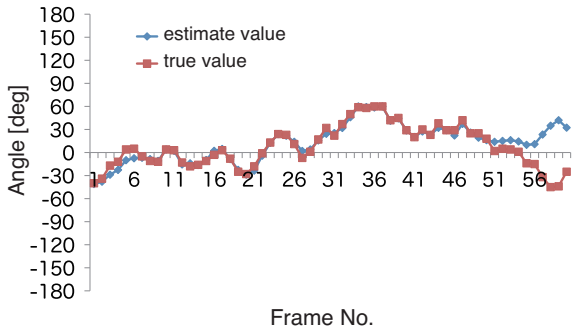

Fig. 13. estimation results of θ_y

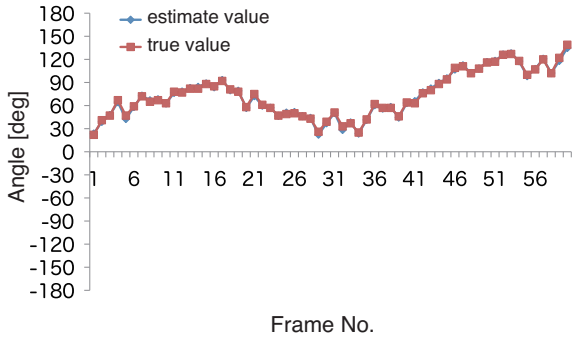

Fig. 14. estimation results of θ_z

results, the accuracy of posture estimation is enough for virtual reality application. However, as shown in Figs. 12, 13 and 14, the sign of some results of posture estimation is reversed. This is because the appearances of images are almost same nevertheless the rotation angles are different as shown in Figs. 15 and 16. In that case, the accuracy of the proposed method becomes worse because γ, θ_1 and θ_2 are almost equal if there is no difference between two images.

C. Experiment with Real Data

We also conducted experiments in the real data to show the effectiveness of our method. We made a B1 size wall paper for this experiment as shown in Fig. 17. The markers of a motion capture system are attached to the wall paper, and motion data from a motion capture system are used as true values. The specification of a camera used in this experiment is shown in Table III. The wall paper is moved at the distance about 3000 [mm] from the camera. Figures 18, 19 and 20 show estimation results of posture parameters, and Fig. 21 show estimation results of the distance between a camera and an object. The rooted mean square errors of estimated parameters are shown in Table IV. From the experimental results, posture parameters, θ_x, θ_y, and θ_z, can be estimated within ± 3.3 [deg], and the estimation error of the distance is about $\pm 2.3 \times 10^2$ [mm]. Estimation error of the distance becomes worse if the distance between a camera and object becomes small. This is because visual markers are shifted to the lower frequency domain and visual changes in the

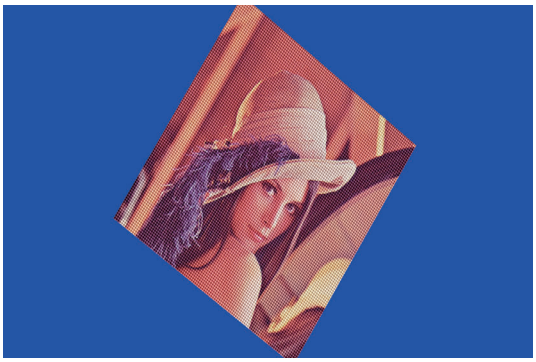

Fig. 15. $(\theta_x, \theta_y, \theta_z) = (20, 30, 40)$

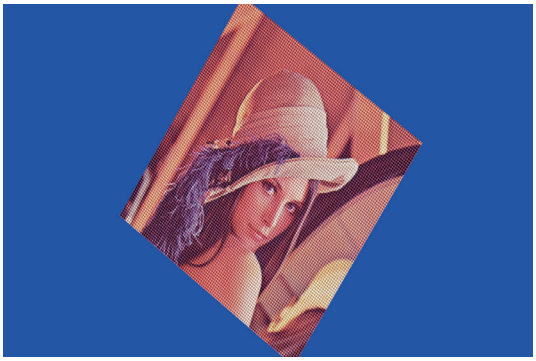

Fig. 16. $(\theta_x, \theta_y, \theta_z) = (-20, -30, 40)$

lower frequency domain are smaller than those in the higher frequency domain.

We have also conducted another experiment with the training data in Table I. Figure 22 show examples of posture estimation. From this figure, the accuracy of posture estimation is enough for applications of augmented reality. We investigate estimation errors between true rotation R_{true} and predicted rotation R by Frobenius norm. The values of Frobenius norm in simulation data and real data are shown in Table V.

From this table, the scaling factor m of estimation error of rotation matrix in the real experiment can be estimated by

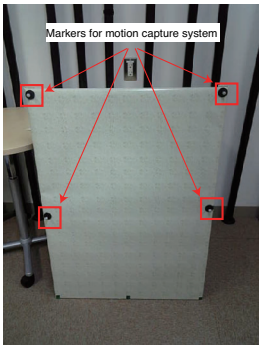

Fig. 17. Wall paper used in this experiment

TABLE III
CAMERA SPECIFICATION

device	FOVEON X3 (CMOS)
size	20.7 × 13.8 [mm]
focal length	200 [mm]
image size	2640 × 1760 [pixels]

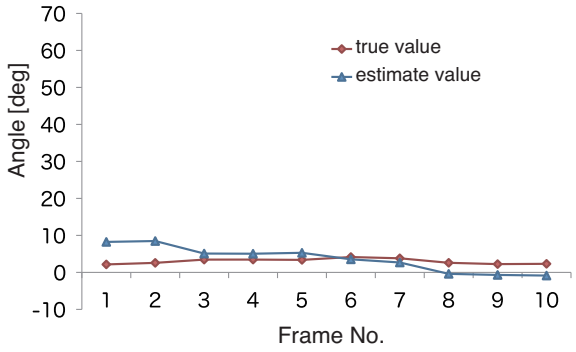

Fig. 18. Estimation results of θ_x

the following equation:

$$m = \frac{1.49 \times 10^{-2}}{0.959 \times 10^{-2}} \simeq 1.55. \tag{6}$$

Let e_{rmse} be a rooted mean square error of an axis, estimation error of rotation angle can be calculated as follows:

$$e = \arccos(1 - m \cdot (1 - cos(e_{\text{rmse}}))). \tag{7}$$

By the above calculation, the estimation error in real scene can be estimated as shown in Table VI.

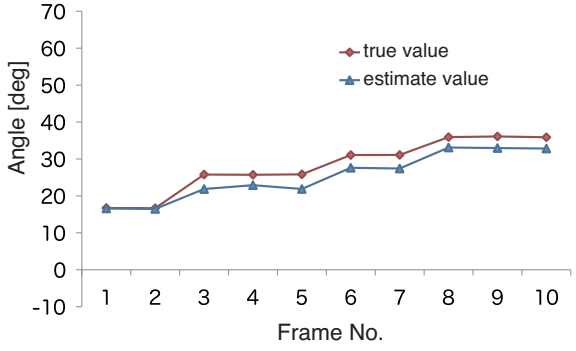

Fig. 19. Estimation results of θ_y

TABLE IV
EVALUATION OF ESTIMATION RESULTS IN REAL IMAGES

parameter	rooted mean square error
θ_x	3.32 [deg]
θ_y	3.03 [deg]
θ_z	0.77 [deg]
d	2.25 $\times 10^2$ [mm]

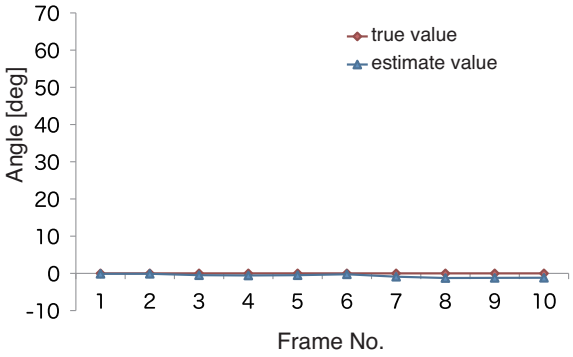

Fig. 20. Estimation results of θ_z

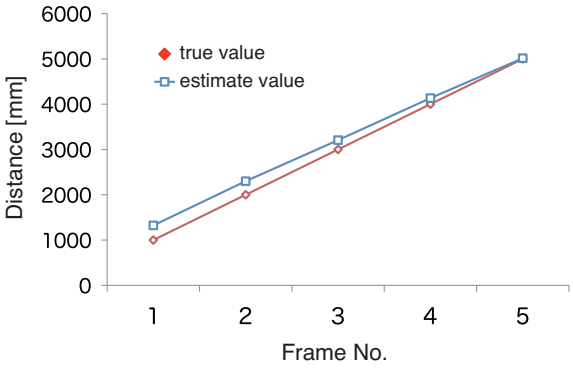

Fig. 21. Estimation results of distance

V. CONCLUSION AND FUTURE WORK

This paper proposes a method for relative posture estimation by using visual markers which are embedded in an object's texture in the high frequency domain, and we show that relative posture can be estimated by kernel regressions. We embedded visual markers as an elliptical shape in this experiment, but our method does not restrict an elliptical shape. Our method requires one of parametric shape or pattern for posture estimation. In future work, we investigate the best shape for embedding which is robust for motion and focal blur.

TABLE V

FROBENIUS NORMS OF ESTIMATION RESULTS

	$\|\|R_{true} - R\|\|_f$
simulation	0.959×10^{-2}
real scene	1.49×10^{-2}

TABLE VI

PREDICTION VALUES OF ESTIMATION ERRORS

estimation parameter	rooted mean square error
θ_x	4.98 [deg]
θ_y	5.22 [deg]
θ_z	1.34 [deg]

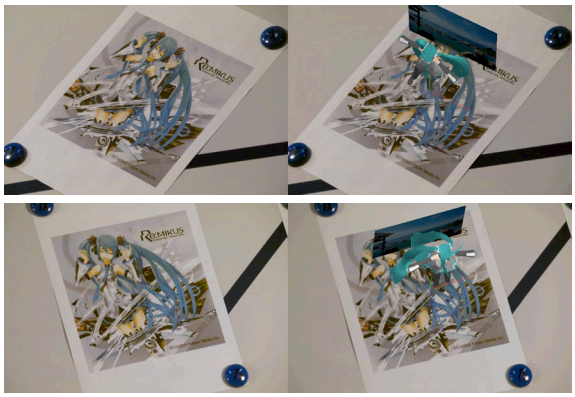

Fig. 22. examples of posture estimation results: input image (left) and superimposed image (right)

REFERENCES

[1] F. Zhou, Henry B.L. Duh, and M. Billinghurst: "Trends in Augmented Reality Tracking, Interaction and Display: A Review of Ten Years of ISMAR", *Proc. ISMAR 2008*, pp. 193–202 (2008-9)

[2] Y. Matsumoto, M. Inaba, and H. Inoue: "View-Based Approach to Robot Navigation", *Journal of the Robotics Society of Japan*, Vol. 20, No. 5, pp. 506-514 (2002-7)

[3] M. Oe, T. Sato, and N. Yokoya: "Camera Position and Posture Estimation Based on Feature Landmark Database for Geometric Registration", *Trans. of the Virtual Reality Society of Japan*, Vol. 10, No. 3, pp. 285–294 (2005)

[4] M. Kanbara and N. Yokoya: "Outdoor Augmented Reality System Using RTK-GPS and Inertial Navigation System", *Technical report of IEICE. PRMU*, Vol. 104, No. 572, pp. 37-42 (2005-1)

[5] K. Matsuda, S. Ikeda, T. Sato, and N. Yokoya: "Robust Estimation of Position and Posture of Camera with High-speed Rotation Using Landmark Database and Inertial Sensor", *IEICE technical report*, Vol. 106, No.535, pp. 5–10 (2007-2)

[6] V. Lepetit, L. Vacchetti, D. Thalmann, and P. Fua: "Fully automated and stable registration for augmented reality applications", *Proc. 2nd IEEE/ACM Int. Symp. on Mixed and Augmented Reality*, pp. 93–102 (2003-10)

[7] H. Kato, M. Billinghurst, K. Asano, and K. Tachibana: "An Augmented Reality System and its Calibration based on Marker Tracking", *Transactions of the Virtual Reality Society of Japan*, Vol. 4, No. 4, pp. 607–616 (1999-12) (in Japanese)

[8] T. Habara, T. Machida, K. Kiyokawa, and H. Takemura: "Wide-Area Indoor Position Detection using Fiducial Markers for Wearable PC", IEICE technical report. Image engineering, Vol. 103, No. 643, pp. 77-82 (2004-1)

[9] Y. Nakazato, M. Kanbara, and N. Yokoya: "A User Localization and Marker Initialization System Using Invisible Markers", *Trans. of the Virtual Reality Society of Japan*, Vol. 13, No. 2, pp. 257-266 (2008-6)

[10] R. Tenmoku, A. Nishigami, F. Shibata, A. Kimura, and H. Tamura "A Geometric Registration Method Using Wall Posters as AR-Markers", *Trans. of the Virtual Reality Society of Japan*, Vol. 14, No. 3, pp. 351–360 (2009)

[11] S. Sato, T. Tanikawa, and M. Hirose: "Indoor position tracking system using interior decorations with coded patterns", *Technical report of IEICE. Multimedia and virtual environment*, Vol. 106, No. 91, pp. 1-6 (2006-5)

[12] C. Cachin: "An information-theoretic model for steganography", *Information and Computation*, Vol. 192, pp. 41–56 (2004-7)

[13] A.P. Witkin. "Recovering surface shape and orientaion from texture", *Artificial Intelligence*, Vol. 17, pp. 17–45 (1981)

[14] B.J. Super and A.C. Bovik: "Planar surface orientation from texture spatial frequencies", *Pattern Recognition*, Vol. 28, No. 5:729–743 (1995-5)

[15] T.Tsuji and Y. Yoshida: "Estimation of rotation and slant angles of a textured plane using spectral moments", *IEICE Trans. on information and systems*, Vol. 85, No. 3, pp. 600 (2002-3)

[16] A.J. Smola, and B. Schölkopf: "A tutorial on support vector regression", *Statistics and Computing*, Vol. 14, No. 3, pp.199–222 (2004-8)