

## Estimating Hay Bale Position with Stereo Vision Technique using an Omnidirectional Camera

R. Farrokhi Teimourlou\*. N. Noguchi \*\*

\*Graduate School of Agriculture, Hokkaido University,  
Sapporo, Japan(Tel: 011-706-4147; e-mail: rahman@bpe.agr.hokudai.ac.jp).

\*\* Graduate School of Agriculture, Hokkaido University,  
Sapporo, Japan, (Tel:011-706-3847;e-mail: noguchi@bpe.agr.hokudai.ac.jp)

---

**Abstract:** In path planning of autonomous agriculture vehicles, detecting and identifying obstacles, and taking appropriate collision avoidance measures are critical for safe operation. The goal of this research was to obtain the hay bale distance and position to be used in real-time obstacle avoidance detection for autonomous robot tractors using stereo vision system. The vision system was an omnidirectional camera that have a wide field view and useful for specific applications such as meadow, open field, etc. The estimated hay bales distance from the camera showed that the calibration parameters should be further improved to enable autonomous navigation of a robot tractor in the meadow.

---

### 1. INTRODUCTION

Today, the development of automated and autonomous guidance systems for agricultural vehicles in Japan is deemed a promising alternative to cope with the dwindling farming labor force, in addition to the quest for higher production efficiency and safer operation. The key to these problems is the use of mobile robots, as a substitute for conventional tractor and will play an important role in future (Noguchi, et al., 1993, 1994). In particular, to be economically competitive in the international markets of agricultural machinery products, Japan should promote the development of robot vehicles for agricultural purposes to reduce production costs and to cope with decreasing farming labors.

Past researches studied about a method for creating work schedules or an automated path creation for an agricultural production system, hay transportation on slope by using two transporting robots which mean the dump trucks, and a loading robot which is a tractor with a hay fork was developed to perform their respective task simultaneously (Noguchi and Terao, 1995). However, all of these papers dealt only with the simulation of the hay bale's work schedule, in spite of the importance of detecting and localizing the obstacles, i.e. the hay bales, before making a work schedule or an automated path creation on a real-time application. It is require continuing researches for resolving these difficulties in order to be used the benefits of previous studies by farmers.

First attempt in performing this schedule in the field was real-time detection of hay bales that was investigated in last study to complete the robot system harvesting hay bales in a meadow (Farrokhi Teimourlou and Noguchi, 2007). Large 3-dimensional (3-D) space, automated systems need the ability to detect and track objects. Similarly, autonomous navigation systems must perceive and avoid obstacles. The recognition

of 3-D structure of a large scene is another application which typically requires large amount of data but it should be process in a short time as much as possible. Therefore, an omnidirectional camera was used to detect white-wrapped hay bales that have a wide field of view and produced panoramic images. This study focused on estimating hay bales orientation using stereo vision techniques. By combining these two methods, i.e. detection of the hay bales and estimation of their orientation, it is possible to predict whether the robot tractors can navigate in the meadow autonomously.

Real-time omnidirectional vision has been studied by the computer vision community for years and has emerged as an important method for robot navigation systems, extracting the 3-D structure of a scene and video surveillance. Many algorithms have been developed using this method from a freely moving camera (Maybank, et al. 1992) or using panoramic images by omnidirectional camera with the curved mirrors to compute depth maps (Kang and Szeliski, 1997). However, these cameras are often associated with high distortion and low-resolution images. Nevertheless, the laboratory's omnidirectional camera (Point Grey Research Inc., Ladybug2) which consists of six high-resolution cameras and mounted on a compact unit was used. This camera produces high resolution images that make it interesting to use in open environment and wide field. Panoramic stereo approach has been proposed to form the optical triangulation by an omnidirectional camera mounted on mobile robot to detect and localize objects. The detecting algorithm had been developed previously and the objective of this research is to localize and estimate the distance of hay bales from camera to be used in an autonomous navigation system in meadow application.

## 2. MACHINE VISION HARDWARE

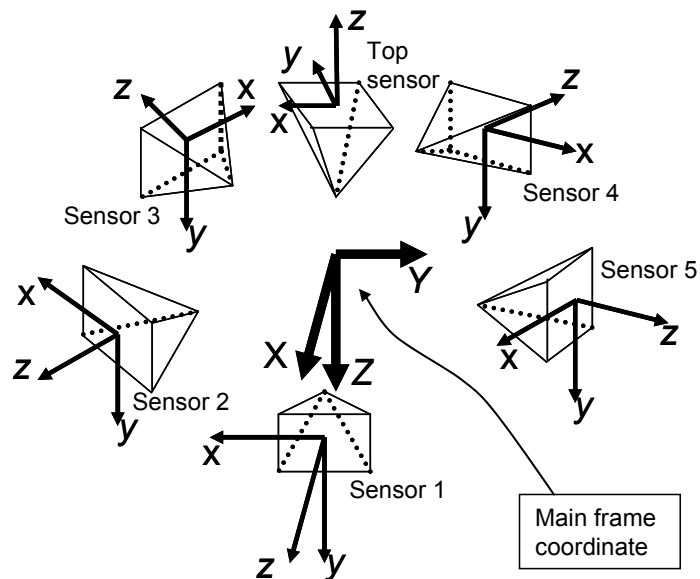
### 2.1 Machine vision

The obstacle detection system was developed based on the high resolution omnidirectional camera, Ladybug2, manufactured by Point Gray Research Inc. (Uyttendaele, *et al.*, 2004). One of the main design objectives of this camera is to capture a high resolution panoramic image. Since a catadioptric camera usually has a limited resolution and a non-uniform distribution of the resolution through the field of view, the Ladybug2 camera employs a multi-camera design to achieve the high resolution requirement (Silpa-Anan and Hartlry, 2007).

The camera head unit of the omnidirectional camera consists of six 1024x768 color CCD image sensors, with five CCD image sensors positioned in a horizontal ring and one is in vertical position pointing upward. The five main sensors give a coverage of 360° horizontally, and 100° vertically; the sixth sensor adds a coverage from the north pole of 50° below the equator. The camera system can collect movies covering more than 75% of the full spherical view with almost the same apparent point of view. The head unit products Bayer tiled images at up to 30-frame per seconds (FPS). The camera system is designed to minimize the effects of parallax by tightly packaging the CCDs and has approximately 4.7M effective pixels and approximately 3800 pixels cover the horizontal circumference. The CCDs in the horizontal ring are in "portrait" orientation to increase the vertical field of view (the vertical resolution of CCD is 1024 pixel while horizontal resolution is 768 pixels). To make a complete panoramic image, the views from the adjacent cameras have a small amount of overlap of approximately 5 to 10 pixels (Point Gray Research Inc, 2003).

### 2.2 Camera coordinate and calibration

The camera software manages the camera coordinate system



by breaking it down into seven right-handed coordinate frames of one of two types: six independent image sensor coordinate frames and a camera coordinate frame. Each of the six image sensor has its own independent coordinate frame. As shown in Fig. 1, the origin of the image sensor coordinate is the optical center of the sensor. Z axis points out of the sensor towards the scene, X axis points to right and Y axis points down. The X and Y axes corresponds to the image  $u$  and  $v$  axes where  $(0, 0)$  is the upper-left corner of the image with the  $u$  axis pointing to the right and the  $v$  axis pointing down. The Camera Head Coordinate Frame presents a unified coordinate frame for the device as a whole. It is based on a spherical camera model. In this model, the camera has one common point in the center where it is a collection of rays in a space. A spherical camera model consists of a common point in the center which is called a center of spherical surface, and a surface of a sphere which is named a spherical image. The Camera Head Coordinate Frame is setup roughly as follows: the origin of this coordinate frame lies roughly in the center of the five horizontally oriented cameras (Point Grey Research Inc., Ladybug2). The Z axis points out through the top sensor. The Y axis is parallel to the X axis of sensor 1 and top sensor, and the direction of the X axis can be derived by applying the right hand rule to the Z and Y axes (Fig.1).

The relationship between a point on a spherical image and a point in a space was then formulated. As illustrated in Fig 2, setting a spherical camera center  $C \in R^3$  to the camera center  $C$ . The intersection of the ray and spherical surface yield a point  $x \in S^2$  (Torii, *et al.*, 2007). Having the equation between  $x$  and  $X$ :

$$x = \frac{1}{|X|} X \quad (1)$$

Since  $x$  is a vector of orientation on a sphere, it is possible to

Fig. 1. An omnidirectional camera, the sensors coordinate and main frame coordinate in systematic diagram

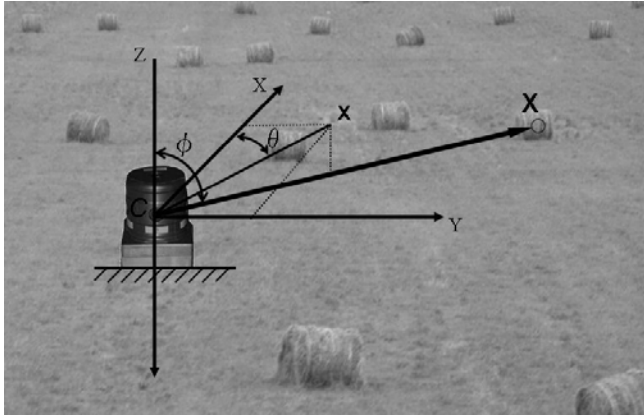


Fig. 2. Two and three dimensional views for spherical cameras

express it in spherical coordinate, such that,

$$\mathbf{x} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} d \cos \theta \sin \varphi \\ d \sin \theta \sin \varphi \\ d \cos \varphi \end{pmatrix} \quad (2)$$

where  $\theta$  and  $\varphi$  are the two spherical angles.  $\theta$  and  $\varphi$  map the panoramic image axes  $u$  and  $v$ .

### 3. METHODS

#### 3.1 Relation between coordinates

The method is applying the least square solution for a few points in 3-D with known world coordinates,  $M_i = [X_i, Y_i, Z_i]^T$  and their corresponding image points in camera coordinates  $N_i = [X'_i, Y'_i, Z'_i]^T$ . The coordinates of 2-D points  $m_i = [u_i, v_i]^T$  have been converted to  $N_i = [X'_i, Y'_i, Z'_i]^T$  in main frame coordinate by using camera

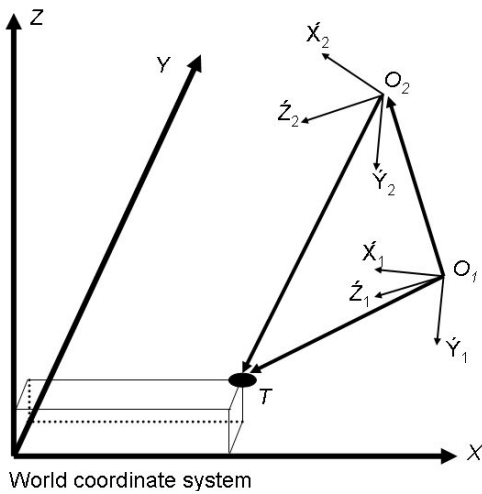


Fig. 3. World and main camera coordinate systems

calibration results,  $P$ :

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = P \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (3)$$

where  $s$  is an arbitrary scale and  $P$  is a 3x4 matrix, called the perspective projection matrix. The world coordinates use to describe the camera position by the rotation  $R$ , camera's main coordinate origin, the position of the camera center and the reconstructed scene. In this case the projection matrix of camera with orientation  $R$  and translation  $t$  is given by (4):

$$M_i = [R | t] N_i \quad (4)$$

Or

$$\begin{bmatrix} X_i \\ Y_i \\ Z_i \end{bmatrix} = \begin{bmatrix} R_1 & R_2 & R_3 & t_1 \\ R_4 & R_5 & R_6 & t_2 \\ R_7 & R_8 & R_9 & t_3 \end{bmatrix} \begin{bmatrix} X'_i \\ Y'_i \\ Z'_i \\ 1 \end{bmatrix} \quad (5)$$

$i$  indicates the number of corresponding point in world coordinate and camera coordinate. By rearranging (5), there are:

$$DQ = L \quad (6)$$

where:

$$Q = [R_1 \dots R_9 \ t_1 \dots t_3]^T$$

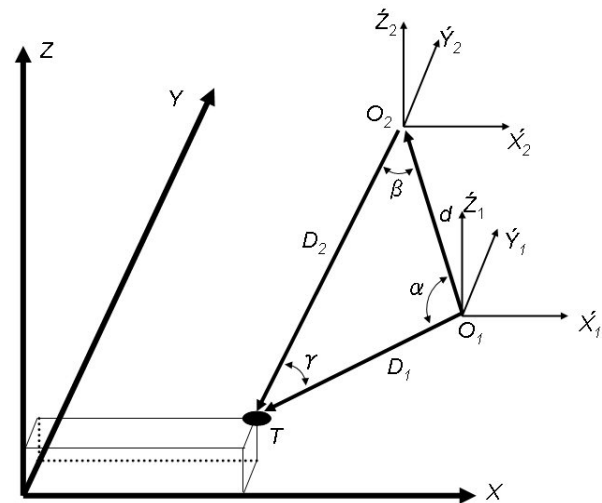


Fig. 4. Rotated camera coordinated and formed triangle between two viewpoints of camera and hay bale



Fig. 5. Omnidirectional camera attached on the tripod

and  $D$  is  $(3i, 12)$  matrix and vector  $L$  is  $(3i, 1)$  that are known. It is possible to determine the unknowns in vector  $Q$  by using the pseudo inverse. By multiplying the above equation by  $D^T$  on the both sides and rearranging (6), vector  $Q$  is determined:

$$Q = (D^T D)^{-1} D^T L \quad (7)$$

Once  $Q$  is determined, the rotation and translation matrix between corresponding 3-D points in camera coordinate and world coordinate are defined.

### 3.2 Estimating obstacle position method

Estimating of obstacles distance is base on stereo vision system. Assume that  $O_1$  and  $O_2$  are the viewpoints of the system, and they can be localized by RTK-GPS in UTM coordinate. Baseline  $d$  is the distance  $O_1 O_2$  between the two sensor viewpoints. Then a triangle  $O_1 O_2 T$  can be formed shown in Fig.3. By using the previously described method, vector of points in two different coordinates convert to world coordinate system and all vectors will be in the same coordinate as shown in Fig.4.

The angle between any two vectors  $v_1 = (a_1, b_1, c_1)$  and  $v_2 = (a_2, b_2, c_2)$  is given in (8):

$$\cos(\theta) = \frac{v_1 \cdot v_2}{\|v_1\| \|v_2\|} \quad (8)$$

where  $\theta$  is the angle between two vectors. Therefore, the three inner angles  $\alpha, \beta$  and  $\gamma$  of the triangle can be calculated. And already the distance between two positions of camera, baseline, is known. Thus, by knowing the three angles and one side of the triangle ( $d$ ), the length of other sides of triangle or distance of camera from hay bale, i.e. the distance between second viewpoint and hay bale,  $D_2$ , is able to calculate by the standard triangulation method as:

$$D_2 = d \sin(\alpha) / \sin(\gamma) \quad (9)$$

or

$$D_1 = d \sin(\beta) / \sin(\gamma) \quad (10)$$

## 4. FIELD EXPERIMENTS TO OBTAIN HAY BALE IMAGES

Images were taken by placing the omnidirectional camera on the tripod between hay bales. The tripod was adjusted so that the camera was at the height of 150cm. In addition, the camera was adjusted to be leveled as much as possible with the ground. The images were taken in different position with 2 meter difference or 10 meter difference. The combination of two images would make the stereo vision system. In other words unlike the usual stereo vision cameras that have two cameras with fixed baseline, this research used one camera with changeable baseline. The baseline varied from 2m to 40m. Figure 5 shows the omnidirectional camera attached on the tripod.

The hay bales were usually in cylindrical shape with a diameter of 170cm and length of 120cm. The axis of cylinders or hay bales was parallel or perpendicular with ground in the different farms.

The real distance of hay bales from camera was measured using the real-time kinematic global positioning system (RTK-GPS) with an error of +/- 2 cm. The position of the omnidirectional camera was determined by putting an RTK-GPS on the top of the camera. As for the hay bales, the RTK-GPS was placed subsequently at the four corners to obtain the real position. The centre of four corners of the hay bales were taken as the middle of hay bales. Hay bale images were taken from different view points within 10m to 110m distance between hay bales and camera.



Fig. 6. Images from two viewpoints

5. RESULT AND DISCUSSION

Figure 6 is an example of two panoramic images of hay bales taken by the omnidirectional camera from two different viewpoints. The position of 4 fixed landmarks in UTM coordinate was used to calculate the rotation matrix between camera main coordinate and UTM coordinate by using the method described in the previous section. This was done for all positions of camera where images were taken. Images were taken from 0m to 10m with 2m step and from 10m to 40m with 10m step. So it was possible to imagine stereo vision system with baseline from 2m to 40m.

Then, the middle of each hay bale in the image was chosen and the orientation of that point was calculated by camera calibration parameter based on camera main coordinate frame. This vector should be converted to UTM coordinate. The conversion was done by using calculated rotation matrix. The orientation of the same hay bale in another position of camera was calculated and then was converted to UTM coordinate too. Afterwards, the angles of virtual triangle which were created by these two vectors and baseline vector were found by (8). The same process was repeated for all the hay bales appeared in each pair images. The results were compared with the corresponding real angles of triangle created by RTK-GPS data. Figure 7 shows the angle error between actual and calculated angles. Apparently, the error is minute,

i.e. significant errors in estimating the distance. This proved that the estimation of the rotation matrix and camera calibration for finding the orientation of hay bales would require more accuracy in order to obtain a precise estimation of distance. The distance of camera from the hay bales were estimated by using (9). The distance error is relatively high, +/- 80m, but for the hay bales that their distance from camera was less than 40 m error is about +/-2.5m. The distance error was shown in Fig. 8.

By looking to Fig. 9, it is clear that the estimation of angle is very critical from 0 to 40 deg and 120 to 180 deg, especially when hay bales are far from camera. It proved that in this method hay bales should be relatively closed to the camera. In Fig. 4 representation of angles,  $\gamma$  is the hay bale angle. If the hay bale angle is bigger than 10 deg, estimated distance error is small but if it's lesser than 10 deg, it's high.

While there are the estimated hay bale's distance and their pose, and also the camera position in UTM coordinate, it was tried to estimate the position of hay bales in UTM coordinate. Figure 10 shows the mean and standard deviation error according on baseline. As expected, the accuracy of estimating distance improves by increasing the baseline length. The RMSE of all baselines is 40.63m.

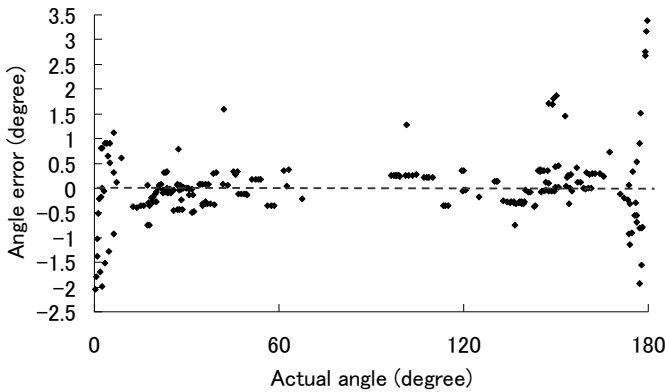


Fig. 7. Estimated angle error between vectors

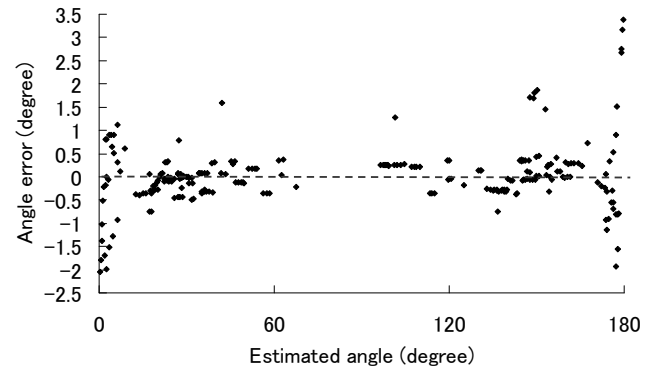


Fig. 9. The estimation distance errors grow up when the angle of triangle is small or big.

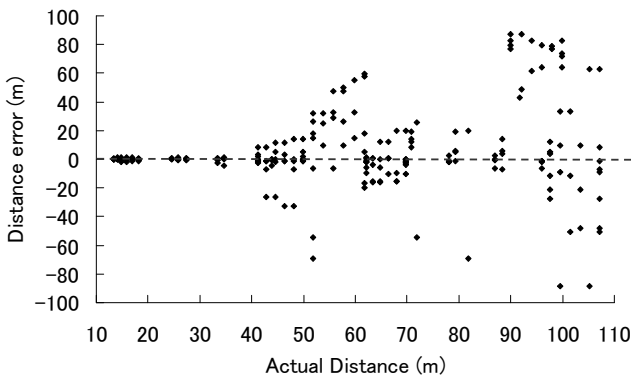


Fig. 8. Estimated distance error

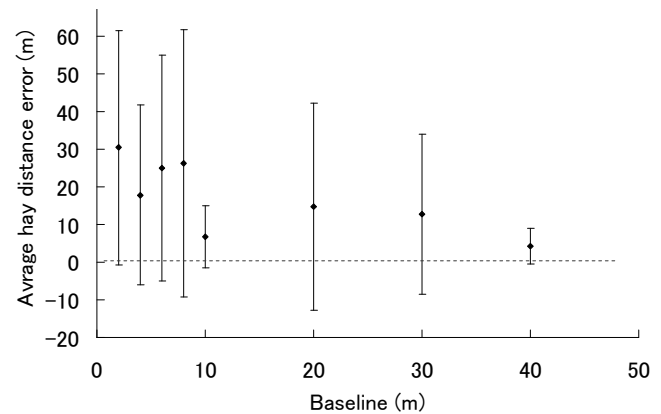


Fig. 10. The mean and standard deviation error base on different baseline.

## 6. CONCLUSIONS

An omnidirectional camera was used as a stereo vision sensor for meadow application. The rotation matrix of main camera coordinate in each position was estimated relative to the world coordinate system. Then, the orientation vector of hay bales were rotated based on this rotation matrix. The correspondent vectors defined a virtual triangle with the baseline vector from first camera position to the second one. Since all angles and one side of triangle are known, the other sides of the triangle could be estimated by applying the standard triangulation method that is the distance of hay bales from camera. The estimated distance was compared with the real distance that was measured by RTK-GPS. The distance errors were quite high because the obstacle distance in this application is relatively far; hence small errors in estimating orientation would result in increasing the estimation of distance error. Nevertheless, it was exhibited the potential of this method in the estimation of obstacle distance. To further improve the result, the camera calibration parameters and the coordinate relations should be optimized.

Uyttendaele, M., Criminisi, A., Kang, S.B., Winder, S., Szeliski, R., Hartley, R., 2004. Image-Based interactive exploration of real-world environments. *IEEE Computer Society*, (0272-1716/04) 52-63.

## REFERENCES

- Farrokhi Teimourlou, R., N. Noguchi (2007) Development of hay bales detection algorithm using omnidirectional multi-camera. *2nd IFAC international conference on modeling and design of control systems in agriculture*, 211-216.
- Kang, S.B. and R. Szeliski, 1997. 3-D scene data recovery using omnidirectional multibaseline stereo. *International journal of computer vision*, 25(2).
- Maybank, S.J. and Faug-eras, O., 1992. A theory of self-calibration of a moving camera. *International journal of computer vision*.
- Noguchi, N., Ishii, K., Terao, H., 1993. Optical control of agricultural vehicles by neural networks (part 1), *Journal of the Japanese society of agricultural machinery*, 55(5), 83-92.
- Noguchi, N., Ishii, K., Terao, H., 1994. Optical control of agricultural vehicles by neural networks (part 2). *Journal of the Japanese society of agricultural machinery*, 56(2), 83-92.
- Noguchi, N. Terao, H., 1995. Creation of route for agricultural vehicle and construction machinery by using genetic algorithm. *Transaction of the society of instrument and control engineers*, 31(10), 1762-1768.
- Point Gray Research Inc. 2003. *Ladybug User manual and API Reference*.
- Silpa-Anan, C., Hartley, R., 2007. Visual localization and loop-back detection with a high resolution omnidirectional camera. <http://www.fieldrobotics.org/~cgeyer/OMNIVIS05/final/Silpa-Anan.pdf>.
- Torii, A., Imiya, A., Ohnishi, N., 2007. Two and three view geometry for spherical cameras. <http://www.fieldrobotics.org/~cgeyer/OMNIVIS05/final/Torii.pdf>.