

## UUV DEPTH MEASUREMENT USING CAMERA IMAGES

### **Rogério Yugo Takimoto**

Graduate School of Engineering  
Yokohama National University  
79-1 Tokiwadai, Hodogaya-ku, Yokohama, 240-8501, JAPAN  
takimotoyugo@vento.shp.ac.jp

### **Tsugukiyo Hirayama**

Graduate School of Engineering  
Yokohama National University  
79-1 Tokiwadai, Hodogaya-ku, Yokohama, 240-8501, JAPAN  
hirayama@ynu.ac.jp

### **Fabio Kawaoka Takase**

Departamento de Engenharia Mecatrônica e de Sistemas Mecânicos  
Escola Politécnica da Universidade de São Paulo  
Av. Prof. Mello Moraes, 2231, 005508-900, São Paulo, Brasil  
fktakase@usp.br

**Abstract.** *In some cases, the control of an unmanned underwater vehicle (UUV) running near the free surface is required. For such purpose, accurate measurement of the depth of the UUV is needed. This can be performed through different techniques. Among them, it is possible to mention the pressure sensor and the sonar. Each technique has its own set of unique features, advantages and disadvantages. The sonar, for example, delivers accurate range but can be disturbed by reflections. This work proposes an alternative solution for the accurate depth measurement problem at short range for control purposes. Through the use of computer vision techniques, the depth of an UUV near the free surface is evaluated. First the CAMSHIFT tracking algorithm, a color based tracking algorithm, is applied to determine the displacement in the image plane and then a projective transformation is applied in order to evaluate the depth of the UUV in the world plane. The projective transformation is a plane to plane homography, implemented using the DLT (Direct Linear Transformation) algorithm equations. The system was implemented in the C programming language and extensive experimental tests were performed to evaluate the quality of measurements.*

**Keywords:** *unmanned underwater vehicle, depth measurement, computer vision*

### **1. Introduction**

Nowadays, it is possible to find computer vision applications in many areas because there is an increased need for the automatic processing of the huge amount of digital images generated by cameras. The availability of faster computers has opened the way for new applications in the computer vision area.

Applications that provide industrial inspection, traffic monitoring measurement become available and there is also some application in the naval architecture area. Projects using computer vision techniques for the naval architecture area are related to 3D reconstruction of underwater scenes from 2D images or to the underwater vehicle position control from real time processing of video data. These projects are useful for many undersea applications including scientific, commercial and military applications, such as the ocean floor exploration, the marine life study and underwater structures inspection and repair.

Computer Vision techniques can also be used to provide accurate depth measurement at short ranges and control an UUV. Such a feature is an important resource for underwater navigation. For such purpose, this work shows the applicability of computer vision techniques to recover the depth of an UUV.

The implementation of such functionalities represents the first step in the development of a perceptual interface system, one of the most promising applications in the computer vision area. Perceptual interfaces are ones in which the computer is given the ability to sense and produce analogs of the human senses, such as allowing computers to perceive and produce localized sound and speech, giving computers a sense of touch and force feedback, and in our case, giving an ability to see and extract useful information from the scene. In order to provide this feature in advance, this research proposes an object tracking and distance measurement application.

The proposed system represents the first step in the development of a complete navigation system for an UUV since it is possible to track and determine the size of a known object in the scene. This system makes use of commercial products such as personal computers and video cameras.

In order to recover the UUV depth, this work uses tracking techniques to identify the image plane displacement and projective transformation techniques to estimate the distance in the world plane. In addition, the measurement in the world plane is compared to the real depth to quantify the result of estimation algorithm.

The objective of this paper is to describe the algorithm used in the software development and the results obtained in the experiments to measure the depth of an AUV near the free surface.

## 2. Homography

According to Hartley and Zisserman (2000), a camera is a mapping between the 3D world (object space) and a 2D image and it can be modeled using a pinhole camera. This model is the simplest, the most general linear camera model and is modeled from a 3D point in space projected onto the image plane. The camera image formation process can be represented by the intersection of the visual ray connected to camera center (C) and the 3D point (X) on the image plane corresponds to the image point (x) (Fig. 1).

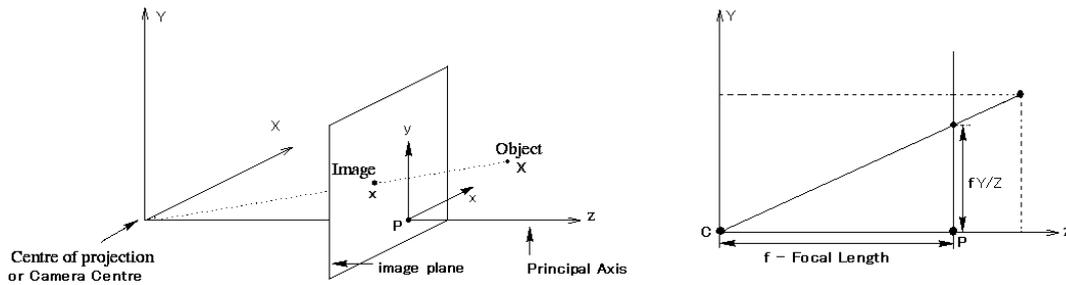


Figure 1. The pinhole camera model

This model can also be used to map points from two 2D planes. The mapping between two 2D planes (plane to plane) or the 2D projective mapping is also known as *projectivity* or *projective transformation* or *homography*. From the camera model for perspective images of planes it is possible to map points on a world plane to points on another plane (Semple and Kneebone, 1979). It is possible to notice that the shape is distorted under perspective imaging because parallel lines in the image tend to converge to a finite point. The distance in the real world (world plane) can be evaluated after removing the projective distortion. This can be done through a projective transformation or homography to change the visualization plane, to correct the geometric shapes from the objects in the image and to eliminate the projective distortion on the determined points. After that, the mapping of an image point to a world point is straightforward through the use of an Euclidean relationship between the desired points. Figure 2 shows a workflow illustrating the projective transformation and the process to evaluate the real distance between two points selected on the image plane.

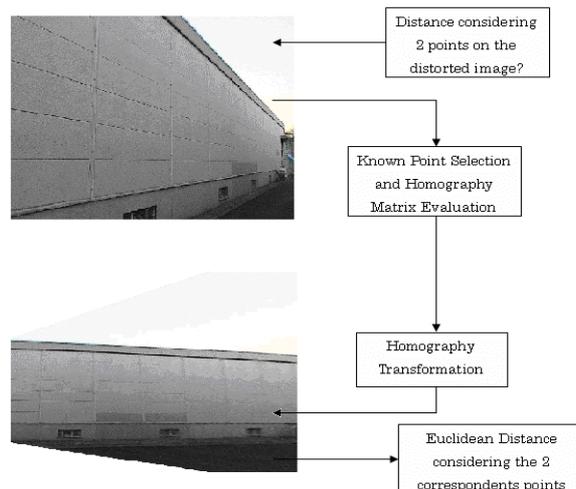


Figure 2. Homography Transformation Workflow

If a coordinate system is defined in each plane and points are represented in homogeneous coordinates (Fig. 3), then the homography can be expressed by (1).

$$X = Hx \tag{1}$$

or by (2)

$$\begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \quad (2)$$

Where  $X$  and  $x$  represent the points coordinates in each plane and  $H$  represents the *homography matrix* and is described by a 3x3 non-singular matrix. The homography matrix ( $H$ ) has 9 elements and only 8 degrees of freedom because the scale of the matrix does not affect the equation.

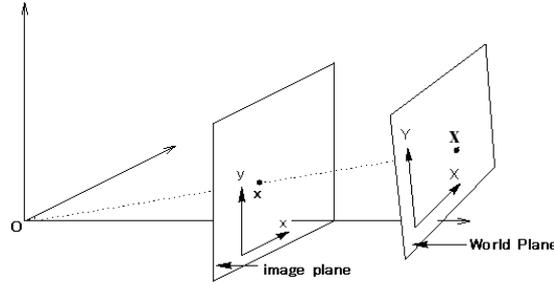


Figure 3. Homography Transformation

There are two methods to evaluate the homography matrix. One method uses the camera internal parameters and the relative positioning of the two planes and camera center. The other one evaluates the homography matrix directly from image to world correspondence eliminating the need to calculate the camera's parameters. In this method, at least four image to world feature correspondences are necessary because each point-to-point correspondence is responsible for two constraints and a 2D point has two degrees of freedom (Hartley and Zisserman, 2000). If the number of feature correspondences increase, we have an over determined solution and  $H$  is determined by finding the transformation  $H$  that minimizes some cost function. To evaluate the homographic matrix it is necessary that the four points must be in "general position", which means that no three points are collinear.

Among the available algorithms to estimate the matrix  $H$ , it is possible to mention the *Direct Linear Transformation* (DLT) algorithm, a widely used algorithm because it is accurate and is relatively simple to implement.

### 2.1. The Direct Linear Transformation (DLT) Algorithm

Taking the Eq. (2) and considering a pair of matching points  $(X,Y)$  and  $(x,y)$  in the world and image plane respectively, the following transformation can be written.

$$\begin{aligned} X(h_{31}x + h_{32}y + h_{33}) &= h_{11}x + h_{12}y + h_{13} \\ Y(h_{31}x + h_{32}y + h_{33}) &= h_{21}x + h_{22}y + h_{23} \end{aligned} \quad (3)$$

where  $X = \frac{x_1}{x_3}$ ,  $Y = \frac{x_2}{x_3}$ ,  $x = \frac{x_1}{x_3}$  and  $y = \frac{x_2}{x_3}$  correspond to the inhomogeneous coordinates in the world and image plane respectively.

Expressing the transformation given by the Eq. (3) in terms of vector product ( $X \times Hx = 0$ ) enable a simple linear equation to be derived. It is possible to write the Eq. (3) in terms of vector product because, since this is an equation involving homogeneous vector, the 3-vectors  $X$  and  $Hx$  are not equal (they have the same direction but may differ in magnitude by a non zero scale factor) (Hartley and Zisserman, 2000).

From the vector product, it is possible to write the Eq. (3) as

$$Ah = 0 \quad (4)$$

where for  $n$  points  $A$  is a  $2n \times 9$  matrix and  $h$  is a 9-vector made up of the elements of the matrix  $H$ .

The Eq. (3) is a homogeneous set of equations and  $h$  can only be determined up to a non-zero scale factor. After selecting a set of four correspondents points, it is possible to solve this equation through the use of the least square

method to find out the solution for  $h$ , that is the singular vector corresponding to the smallest singular value of  $A$ , or the last column of  $V$  in the SVD ( $A = UDV^T$ ). This solution minimizes  $\|Ah\|$  subject to the condition  $\|h\|=1$ . The SVD (singular value decomposition) is a factorization of  $A$  as  $A = UDV^T$ , where  $U$  and  $V$  are orthogonal matrices and  $D$  is a diagonal matrix with non-negative elements (Criminisi, 1999; Hartley and Zisserman, 2000).

This algorithm known as the basic DLT algorithm can be improved by using a method of data normalization for the DLT algorithm, consisting of translation and scaling of image coordinates. This algorithm is known as DLT algorithm with data normalization. The objective of the normalization is to translate the coordinates of the image to new set of points such that the centroid of this new set of points is the origin. The coordinates are also scaled to adjust the average distance from the origin to the same average magnitude.

### 3. Tracking

The ability to tracking moving objects in real time is very useful in many different areas for purposes such as robot navigation, surveillance, video conferencing, etc. In the real time tracking, a process that follows a moving object (or several ones) in time using a camera is required. This process analyses each video frame and outputs the location of moving targets within the video frame based on a feature of this particular object.

Among the available tracking algorithms, this work uses a color based algorithm to output the location of an object to evaluate the AUV depth. To implement this algorithm a HSV color system is used to separate the hue (color) from the saturation (how concentrated the color is) and from the brightness. This algorithm creates a probability distribution image of the desired color in the video taking 1D histogram from the hue channel in HSV space. Using the color histogram as a model the incoming video is converted to a corresponding probability image of the tracked color and the position of the desired image in the video is evaluated.

In this work a color-based tracking algorithm known as Continuous Adaptive Mean Shift (CAMSHIFT) algorithm (Bradski, 1998) was chosen. Since the color distribution changes over time for each video frame, an algorithm that searches this distribution changes is required. The algorithm that satisfies this condition is the modified mean shift algorithm or the CAMSHIFT algorithm because it adapts the probability distribution of the object dynamically to track the object in each video frame.

The CAMSHIFT algorithm uses a robust non-parametric technique for climbing density gradients to find the mode of probability distribution.

#### 3.1. CAMSHIFT and Mean Shift Algorithm

The mean shift algorithm is a non-parametric technique developed by Fukunaga and Hostetler (1975) that climbs the gradient of a probability distribution (Fig. 4) to find the nearest dominant mode – peak. Since the objective of this technique is finding the densest region in the image, it is possible to notice that the algorithm always converges to the local maximum region in the probability density function.

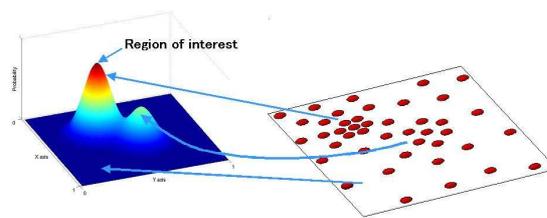


Figure 4. Probability Density Function Mapping

To analyze the algorithm convergence let's consider  $n$  data points  $x_i, i = 1, \dots, n$  on a  $d$ -dimensional space  $R^d$ . According to Silverman (1986) the multivariate kernel density estimate obtained with kernel  $K(x)$  and window radius  $h$ , computed in the point  $x$  is defined as

$$f(x) = \frac{1}{nh^d} \sum_{i=1}^n K\left(\frac{x-x_i}{h}\right) \quad (5)$$

where the  $d$ -variate kernel  $K(x)$  is nonnegative and integrates to one. A widely used class of kernels is the radial symmetric kernels.

$$K(x) = c_{k,d} k\left(\left\|\frac{x}{h}\right\|^2\right) \quad (6)$$

where the function  $k(x)$  is called the profile of the kernel, and  $c_{k,d}$  is a normalization constant that makes  $K(x)$  integrates to 1 (Wand and Jones, 1995).

The gradient of the density estimator (Eq. (5)) is

$$\nabla f(x) = \frac{2c_{k,d}}{nh^{d+2}} \left[ \sum_{i=1}^n g\left(\left\|\frac{x-x_i}{h}\right\|^2\right) \right] \left[ \frac{\sum_{i=1}^n x_i g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{i=1}^n g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)} - x \right] \quad (7)$$

where  $g(s) = -k'(s)$ . The first term is proportional to the density estimate at  $x$  computed with kernel  $G(x) = c_{g,d} g\left(\left\|\frac{x}{h}\right\|^2\right)$  and the second term

$$m_h(x) = \frac{\sum_{i=1}^n x_i g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)}{\sum_{i=1}^n g\left(\left\|\frac{x-x_i}{h}\right\|^2\right)} - x \quad (8)$$

is the *mean shift* (Cheng, 1995). The mean shift vector always points toward the direction of the maximum increase in the density and is proportional to the normalized density gradient. The convergence proof is discussed in Cheng (1995) and is guaranteed to converge to a point where the gradient of density function is zero ( $\nabla f(x) = 0$ ).

The mean shift algorithm iteratively performs

- ◆ computation of the mean shift vector  $m_h(x^t)$ ,
- ◆ update the current position  $x^{t+1} = x^t + m_h(x^t)$

The mean shift algorithm process is illustrated in Fig. 5.

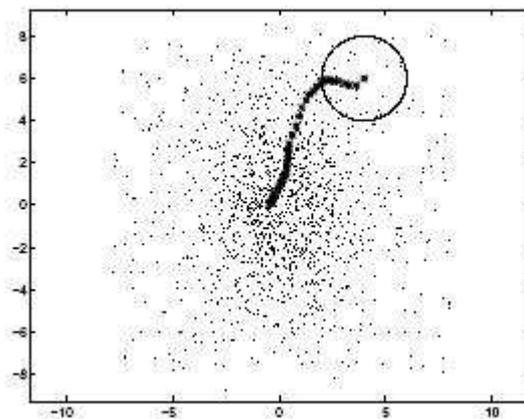


Figure 5. Mean Shift Algorithm

Since the size and location of the probability distribution changes in time due to the object movement in the video it is necessary to modify the mean shift algorithm because it is designed for static distributions. The algorithm that satisfies this requirement is the CAMSHIFT algorithm because it is designed for dynamically changing distributions since the search window size is adjusted in the course of its operation. The CAMSHIFT algorithm relies on the zeroth moment information to continuously adapt its window size within or over each video frame. Thus, window radius, or height and width, is set to a function of the zeroth moment found during search.

## 4. Experimental Results

### 4.1. Preliminary Results

In order to validate the distance measurement module of the developed software some tests were made before the underwater experiment. The test was made marking some points on a white board. The Fig. 6 shows an image of the scheme used to measure the real distance from two points using the developed software. After selecting four known points in the image, the homography matrix can be evaluated and the real distance considering two points on the white board image can be evaluated.

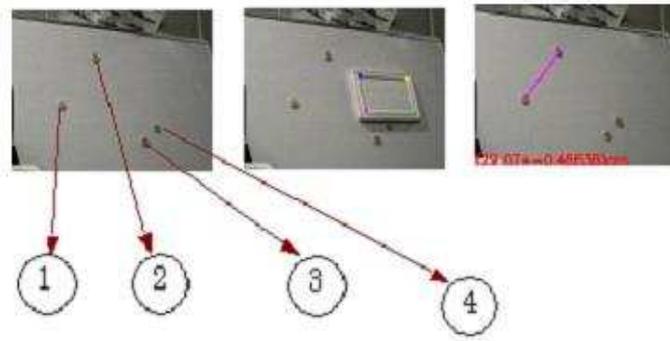


Figure 6. Distance Evaluation Test

The Tab. 1 shows the results of the preliminary experiments. In order to compare the real distance with the measured distance, the absolute and the relative error are also presented.

Table 1. Preliminary Experiment Results.

Points	Real Distance (cm)	Measured Distance (cm)	Absolute Error (cm)	Relative Error (%)
1-2	29	$29.07 \pm 0.466$	0.07	0.24
2-3	39.6	$39.91 \pm 4.22$	0.31	0.78
3-4	9.7	$9.88 \pm 0.08$	0.18	1.85
1-4	40	$39.88 \pm 3.07$	0.12	0.3
1-3	45.6	$44.57 \pm 2.80$	1.03	2.25
2-4	40.5	$41.11 \pm 3.44$	0.61	1.5

### 4.2. Calibration

In order to enable the absolute depth measurement of the AUV it is necessary to perform an initial calibration. The software calibration step can be divided in two parts, the first part involves the calibration to remove the perspective distortion of the measurement plane and the second part involves the calibration to determine the relationship between the depth of the AUV and the position of the light reflection in the camera image.

To remove the perspective distortion (homography matrix evaluation) of the measurement plane it is necessary to find out the displacement plane of the light reflection. In our case this plane is the perpendicular line to the wave surface passing through the source of light. The Fig. 7 shows the scheme for the plane calibration.

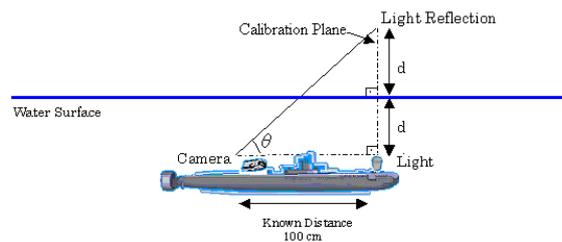


Figure 7. Calibration Plane and Light Position

After removing the perspective distortion of the measurement plane, the software is ready to measure distance between any two points in the image. Thus it is possible to evaluate the AUV vertical displacement (relative depth) considering an initial depth.

To measure the real depth (absolute depth) it is necessary to find out the relationship between the depth of the AUV and the position of the light reflection in the camera image. After some preliminary tests, it was possible to conclude that each depth corresponds to a point coordinate in the image because as the depth of the AUV increase the angle of incidence ( $\theta$ ) in the camera also increase. Then point coordinate and the real depth of the AUV was determined from this relationship.

### 4.3. AUV Depth Measurement experiment

After the calibration step, underwater tests were performed in still water condition, regular wave condition and irregular wave condition. Taking the most general and real situation case, the results are as follow.

The Fig. 8 shows the depth variation in case of irregular waves.

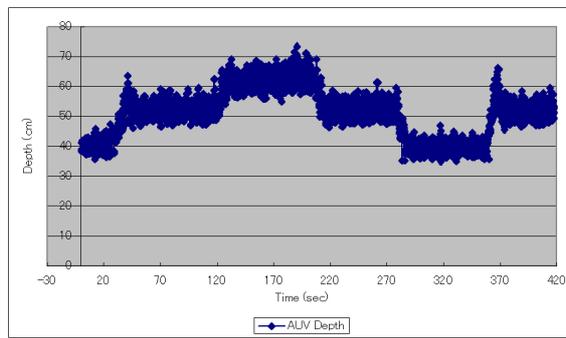


Figure 8. AUV Depth Time History in Irregular Waves Condition.

In order to compare the irregular wave influence on the measurement the output measured depth was compared to the real one. The Tab. 2 shows a comparison between the measurement depth data output and the real value.

Table 2. AUV Depth Measurement in Irregular Wave Condition.

Time Interval (sec)	Real Depth (cm)	Mean Depth (cm)	Minimum Depth (cm)	Maximum Depth (cm)	Absolute Error (cm)	Relative Error (%)
0.10 - 26.9	40	40.35 ± 0.22	35.49 ± 0.121	47.36 ± 0.278	0.35	0.875
86 - 116.9	50	52.03 ± 0.88	46.86 ± 0.237	59.42 ± 2.569	2.03	4.06
146 - 206.9	60	62.62 ± 3.913	54.79 ± 1.25	73.33 ± 10.55	2.62	4.36
236 - 276.9	50	52.33 ± 0.941	47.17 ± 0.253	61.33 ± 3.237	2.33	4.66
326 - 356.9	40	39.58 ± 0.020	34.97 ± 0.147	44.94 ± 0.115	0.42	1.05
386 - 417.6	50	52.04 ± 0.876	46.50 ± 0.20	59.48 ± 2.552	2.04	4.08

Since the irregular wave is often a combination of many wave components, the definition of the wave characteristics (height, period, etc.) must be statistical or probabilistic, indicating the severity of wave conditions. Among these statistical parameters it is possible to mention the significant wave height (mean of the largest 1/3 of waves measured during the sampling period) and the average period. In our experiment these parameters are 2.369 cm for the significant wave height and 1.080 sec for the average period.

It was possible to notice that the presence affects negatively the depth measurement increasing the uncertainty of the measurement. In this situation there is a perturbation on the water surface and it is possible to notice two phenomena that affect the AUV depth measurement.

First, the wave changes the height of the wave surface displacing the reflection light coordinate in the image when compared to the position without waves. Considering a given instant, the measured depth will depend on the wave height in the same instant. Second, the light reflection suffers a deformation due to the water surface movement affecting the tracking and the light reflection coordinate identification

From the experimental results it was possible to analyze the tracking algorithm performance in wave condition (Fig. 8). The Fig. 9 shows the areas where the tracking algorithm performance is good when comparing the depth of the AUV and the wave type used in the experiment.

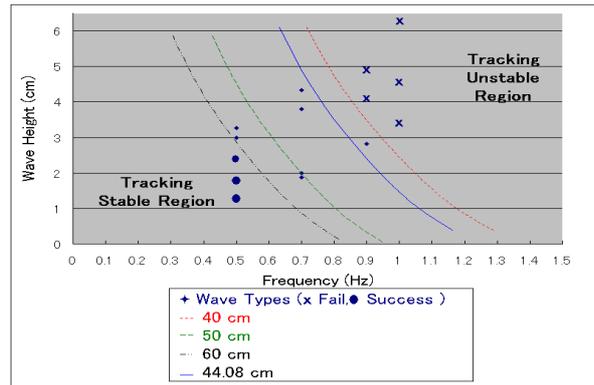


Figure 9. Tracking Stability Analysis

## 5. Conclusion

As an alternative solution for the AUV or UV depth measurement problem, the experimental results show that the mean measured value is very close to the real depth value. This proximity shows that, although the discrepancy of the measured value depth and the real depth is high for a given instant  $t$ , it is possible to evaluate the depth of the AUV with a good approximation taking the mean value in a coherent interval of time. The relative error considering the mean measured value is less than 10%.

It was possible to notice also that two factors affect the tracking performance: the tracked object size and the wave type. The tracking performance is better for bigger objects than for small one. In our case since the tracked object size depends on the depth, the tracking performance depends on the depth. The tracking performance also is better when the wave height and the wave frequency are low (Fig. 9).

From the experiments it was possible to conclude that, the homography transformation represents a good approach for the distance measurement when the displacement plane is known and the tracking performance decrease as the wave height and the wave frequency increase.

## 6. References

- Bradski G. R., 1998, "Computer vision face tracking for use in a perceptual user interface", Intel Technology Journal, Q2:1-15.
- Cheng Y., 1995, "Mean shift, mode seeking, and clustering", IEEE Trans. Pattern Anal. Machine Intell, 17:790-799.
- Criminisi A., 1999, "Accurate Visual Metrology from Single and Multiple Uncalibrated Images", Ph.D. thesis., Dept. Engineering Science, University of Oxford, Oxford, UK.
- Fukunaga K. and Hostetler L. D., 1975, "The estimation of the Gradient of a Density Function, with Applications in Pattern Recognition", IEEE Trans. Info Theory, vol. IT-21, 32-40.
- Hartley R. and Zisserman A., 2000, "Multiple View Geometry in Computer Vision", Chap. 1,2,3,4,5. Cambridge University Press.
- Semple J. and Kneebone G., 1979, "Algebraic Projective Geometry", Oxford University Press.
- Silverman B. W., 1986, "Density Estimation for Statistics and Data Analysis", New York: Chapman and Hall.
- Wand M. P. and Jones M. C., 1995, "Kernel Smoothing", Chapman & Hall, London.