Estimation of Depth Information from a Single View in an Image

S Murali N Avinash
P.E.S. College of Engineering, Mandya - 571 401
nymurali@yahoo.com avinash@witty.com

Abstract

Distance measurement in real world has always been one of the challenging tasks to be performed in the field of computer vision. Photographic images are two-dimensional depiction of three-dimensional real space. On methodical observation of the same, one can identify some third dimensional properties (like depth information). In this paper we have proposed a method, which employs perspective projective geometrical tool, and henceforth bring out a new idea of computing the distances between the edges in the object, which are parallel to the image plane in photographs of actual scene. The technique presented employs uncalibrated images with no knowledge of the internal parameters of the camera (as focal length & aspect ratio) or it's pose (position & orientation with respect to viewed scene). Geometric characteristics (perspective projections) of the scene are employed.

Keywords: Perspective projections, True height, Vanishing point, Picture plane, Geometric cues.

1. Introduction

Images or sequence of images potentially carry a tremendous amount of geometrical information about the scene represented. The aim of the work presented in this paper is to extract this information in a quantifiable form. Perspective projection has been employed as the tool, and henceforth computation of the distances between the objects (vertical edges in our case) in photographs of actual scene is done, as well as the orientation of the same with the picture plane.

The theory developed in this paper proceeds from a strong and reliable mathematical basis, Projective Geometry [2].

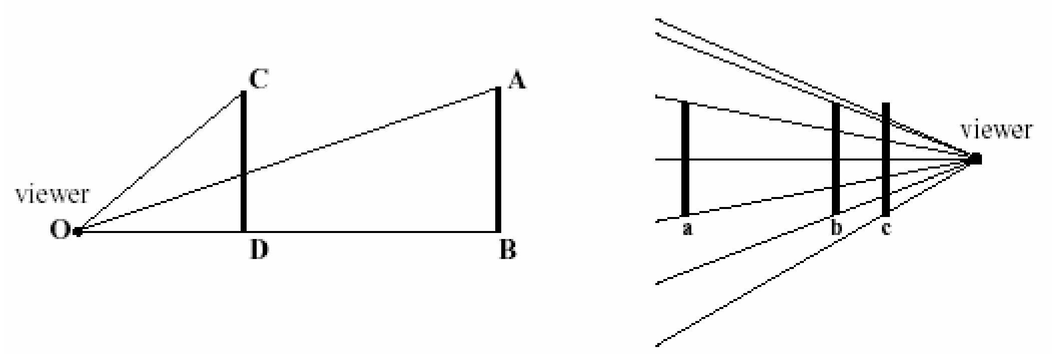
The technique represented requires no knowledge of the camera's internal parameters (focal length, aspect ratio [1,8,9], principal point) or external ones (position and orientation); i.e. no internal or external camera calibration is needed [3]. Camera calibration is replaced by the use of scene constraints (often referred to as scene calibration) such as planarity of points and parallelism of lines and planes. These geometric cues are inferred directly from the images; no scene markers or specialized sensors are required. A new geometric technique to calculate distance measurements is shown to address a range of different cases. The techniques span situations, where no metric information is known about the world (completely uncalibrated camera), to cases where some reference distances are known but they are not sufficient for a complete camera calibration [11] (partial calibration). This leads to the development of extremely flexible algorithms, which can be applied to a wide range of images such as: photographs of buildings and interiors, aerial images, archived images, photographs of huge buildings taken as separate pieces and later to be stitched to bring down as one image. An earlier work on this area of research can be found in the photogrammetry literature [4]. Both single and multiple view configurations are employed. In the past, stereo vision [5] has been used to compute depth. But, in this work, monocular vision proves to be sufficient. This shall be brought into limelight in the subsequent sections to flow by.

This paper is organized as follows. In section 2, an overview of perspective geometry is discussed. In section 3, the different geometric cues along with the role of each feature and their importance and relationships are addressed. In section 4 the actual methodology of the working of our algorithm is discussed. Experimental results in section 5 with conclusion in section 6 are to follow.

2. Perspective Geometry

2.1 Perspective Distortion and Views

Perspective projection is the representation of an object on a plane surface, called the picture plane, as it would appear to the eye, when viewed from a fixed position. It may also be defined as the figure formed on the picture plane when visual ray from the eye to the object cut the picture plane. Thus the image obtained from any camera is nothing but the form of perspective view of the actual scene. Perspective distortions occur during the image acquisition stage. For instance, objects that are far away from the eye (or camera) look smaller than objects, which are close (This basic intuition has been formalized first by Euclid in his Optica [6]). Example for perspective distortion is shown in figure 2.1 (a) and 2.1 (b). Among objects of equal size, one that is the most remote from the eye will appear the smallest. Leonardo da Vinci (1452-1519), B.M.19r. Figure 2.2 shows the box with edges highlighted are of equal measurements in the actual world co-ordinates, though they don't appear as such in the image due to the perspective distortion.



(a) (b)

Fig. 2.1: First proofs of perspective effect**:** (a) Euclid's proof (ca. 300 B.C.); (b) Leonardo's proof:

Fig. 2.2: Perspective distortion in the image acquisition process

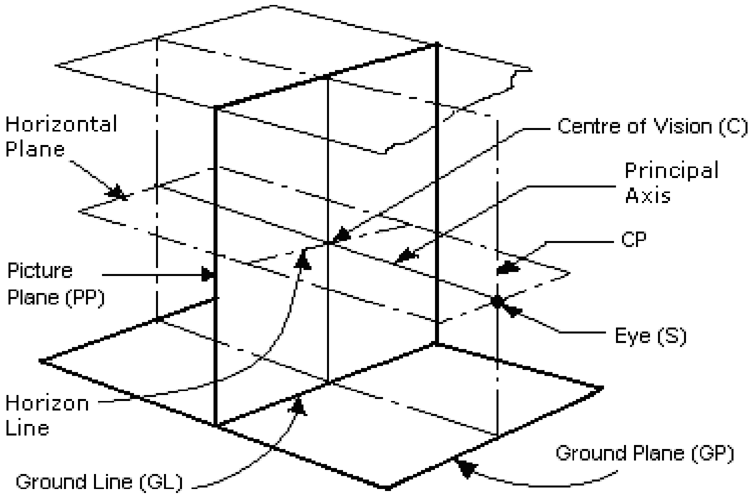


Fig. 2.3: Perspective view planes and points, which are important to be noticed.

The position of the picture plane relative to the object determines the size of the perspective view. The perspective shows the object reduced in size when it is placed behind the picture plane. If the object is moved nearer the picture plane, the size of the perspective will increase. When the picture plane coincides with the object, the perspective of the object will be of its exact size. When the object is placed in front of the picture plane, its perspective when projected back will show the object enlarged in size.

2.2 Vanishing Points

Vanishing points are imaginary points at infinite distance away from the station point. In practice, the point at which the visual ray from the eye to that infinitely distant point pierces the picture plane is referred to as the vanishing point [10].

Drawing a line parallel to the top view of that line from the top view of the station point lots the vanishing point for any horizontal line (in reference to the fig. 2.4). The point at which this line intersects the top view of the picture plane is then projected on the horizon line. This point on the horizon line is the front-view position of the vanishing point. Thus perspectives of all horizontal lines, if produced, pass through their respective vanishing points on the horizon line. Perspectives of all horizontal parallel lines converge to a vanishing point on the horizon line.

Vanishing points for lines perpendicular to the picture plane (when seen in fig. 2.3) are obtained by drawing a line through the top view of the station point, and perpendicular to the picture plane. It lies on the horizon line and coincides with the center of vision. Perspectives of all lines perpendicular to the picture plane converge to the center of vision on the horizon line. This is shown in fig. 2.4.

Lines (fig. 2.3) that are parallel to the picture plane will have no vanishing points. They vanish at infinity. Therefore, perspectives of vertical lines are vertical; perspectives of horizontal lines that are parallel to the picture plane remain horizontal; and perspectives of lines inclined to the ground plane and parallel to the picture plane will remain inclined in the same direction.

2.3 Measuring line or Line of heights

When a line is on the picture plane, it is seen in its true length in its perspective. When a line is behind the picture plane, it is foreshortened in its perspective view.

In the fig. 2.5, *ab* is the top view of a rectangle *ABCD* whose surface is vertical and inclined to the picture plane, the edge *CD* is on the ground plane and the edge *AD* is on the picture plane. In its perspective view *A'B'C'D'*, *A'D'* is equal to the true length *AD*, while *B'C'* is shorter. *A'D'* lies on a vertical line drawn through *a*. The length *B'C'* is derived from *A'D'*.

Thus the measuring line or the line of heights is the trace or the line of intersection with the picture plane, of the vertical plane containing the point or points whose heights are to be determined. Heights of points lying in different vertical planes can be measured from their respective lines of heights. Heights on this line may be measured directly with a scale or may be projected from the front view.

3. Geometric Cues(Feature identification)

The aim of geometric cue detection is to isolate features of the image that can be used to decipher the type of projective distortion it has gone through. In this work, totally seven main patterns are being noted to be sufficient which are relevant to view the measurements. Vanishing points are primary geometrical cue used to determine projective distortion. Horizon line is the line formed by the vanishing points of all the planes that have a horizontal component. As well as being useful geometric cues, the horizon line can also assist in the error analysis of the vanishing points using [3].

In reference with figs. 2.4, 2.5 and 2.6, the ratio of the distances between the vanishing point to the image that is just touching the picture plane and to the base of the same is selected as the first feature to be tabulated ($Feature_1$=e/l). The ratio of the lengths of the two edges is considered and is tabulated as the second feature ($Feature_2=h_n/h_{n+1}$). The angles of the two projection lines of the edges top and bottom points are the third feature ($Feature_3 = \alpha$). Another feature is the angle between the horizon line to the base of the point touching picture plane and that is the angle ($Feature_4=\theta$). Fifth feature is the visible distance of the two edges to be considered for the particular sample ($Feature_5$=Dn). The sixth and seventh features are the features, which are tabulated for the output. Tabulation for the output values while training is also done due to the fact that the training is a 'supervised learning process' in the adopted feed back neural network. Hence they are the actual distances between the edges ($Feature_6= d_n$) and the orientation in some angle with respect to the picture plane ($Feature_7= \phi$).

4. The Method

Assumptions: A certain data set of extreme points to which our system is ready to accept values for, is observed carefully. Within those ranges, a number of samples are jotted down in a tabular form and those images are first plotted on a drawing sheet, which simulates the real image and are ready for measurements. This helped us in avoiding lots of image processing work initially, as our main aim of work was to test the algorithm. Thus the images are measured for all the features and the table of input and corresponding outputs data are completed. This table (as shown in Table 5.1) provides us with the complete input and output data set, which is necessary to train the system using a feed back neural network.

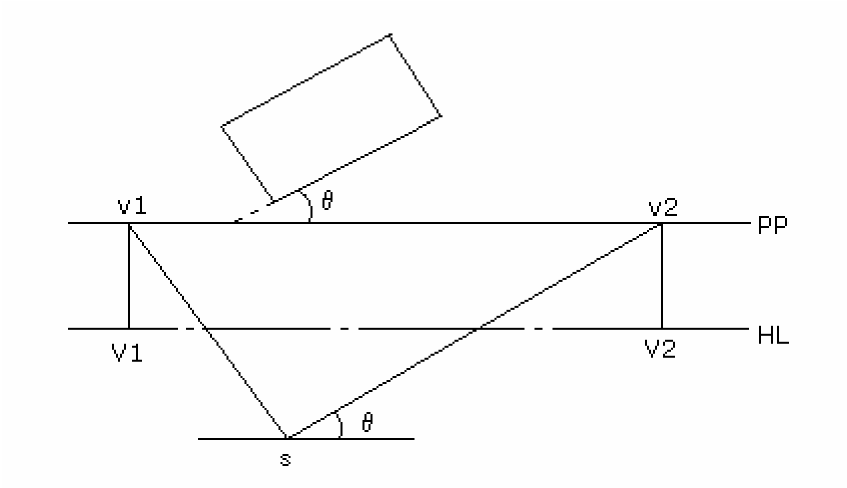


Fig 2.4: Depiction of Vanishing Points

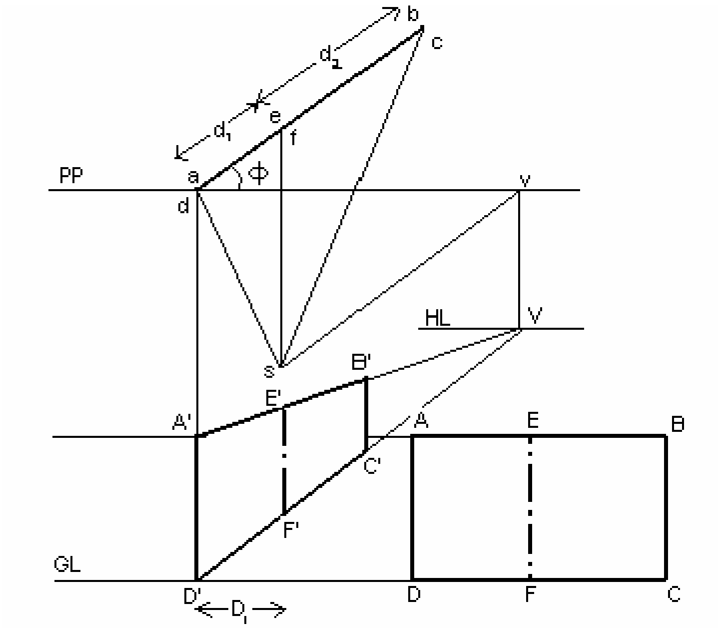


Fig 2.5: Perspective projection depicting Line of Heights.

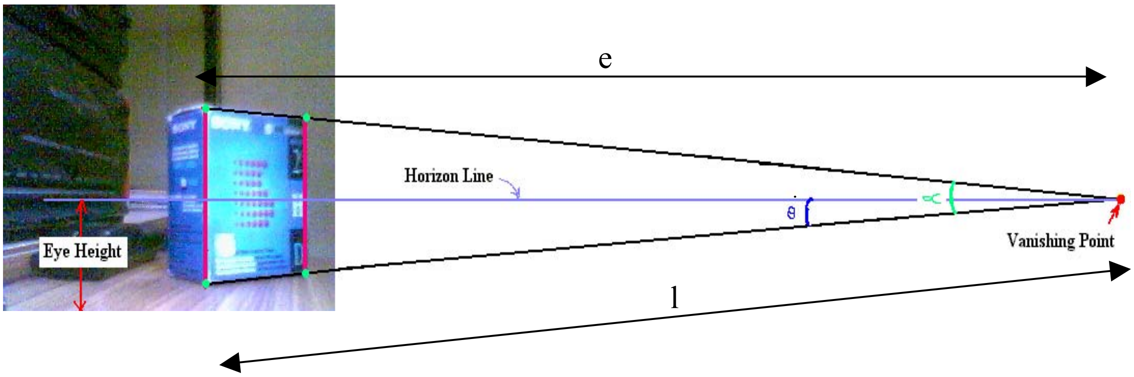


Fig. 2.6: The Image taken and the edges dimensions being calculated using the above method.

Once the neural network with the following specification is trained, an output sample of only the input set is passed to the neural network. The exact output is obtained for our data. The method is as explained below.

Experiment: The implementation part consists of two phases. The first phase mainly consists of the training the system and the second phase consists of actual recognition of a real time image.

In the first phase of the work, the following are carried out in order.

a. Generation of control images.

b. Measurements of geometric cues from control images.

c. Tabulation of features.

d. Designing the system for training

In the second phase of the work, the following are carried out in order.

a. Vanishing point detection.

b. Initial measurements.

c. Recognition from the trained system.

In order to train the system for different values it is found suitable to consider for a set of samples (hand generated images in this case). Various range of data of the picture plane varying from $\phi = 10^0$ to 70^0 are being plotted. It is thus easy to distinguish between different geometric cues.

In the implementation of neural network, a feed forward network [7] with back propagation is used. The advantages and the reason why back propagation is made use in this work is because, feed back networks have the feature of generalization and function approximation. The need for the use of feed back neural network is, we get the wide range of data set initially for the purpose of training them. At the time of recognition, if we have to supply the data set well within the training sample set, then the function, which has been approximated while training would suffice for the result to be calculated.

Lines in world co-ordinates, which are parallel and are mapped through projective transformation to a picture plane, are no longer parallel unless the vanishing point is at infinity. These lines will infact cross and this crossing point defines the vanishing point. The 'Hough Transform' is a widely used algorithm for determining lines and other geometrical primitives in computer vision. For the purpose of vanishing point detection, the Hough Transform is optimized in order to detect lines that would be parallel in the world co-ordinates of a scene. This holds good for the case of occlusion present in the scenes.

Lines that are known to be parallel in the scene and that are also parallel in 2 D image of the scene have vanishing points that lie infinitely far away. Only the direction of these vanishing points can be computed. This approach requires manual input and therefore its accuracy is dependant on the accuracy of the measurement of parallel lines. Vanishing points are often far off the image.

Thus, the work involves how best the geometry is represented and computer from a single perspective image. After the algorithm has been validated using generated images, the real world applicability needs to be determined. Thus after finding all the geometric cues, the trained system is now ready to takes new inputs to give out the result.

5. Experimental Results

The inputs to the network are the tabulated data set from the control images. The inputs are as mentioned before. The network is trained for 23 such control images and for 190 samples. The learning rate is 10^{-5} and is trained in 2,00,000 epochs.

A set of 7 data is tabulated here in Table 5.1, which are corresponding to the figures shown in appendix. The input data set as said earlier consists of both the inputs and their corresponding outputs for training.

After training, a test sample of certain input set is supplied to the network, which came out with the appropriate result, as follows.

Test Sample = [0.58 ; 1.6666 ; 36 ; 55 ; 1.2]

Obtained result from the trained network, output = [58.8954 ; 3.98700]

Output that was plotted gave, output = [60 ; 4]

A few example control images are tagged in the appendix of this paper. The dark lines that we can see in them are considered to be the edges of the image. We can also observe the vanishing point in them. The measurements for the purpose of verification of the algorithm are done manually.

6. Conclusion & Future Possibilities

A theoretical framework employing techniques drawn from projective geometry has been shown to compute measurements from single uncalibrated view. The process of measuring has been treated as a true engineering task, therefore particular attention has been paid in predicting the final measurements from the input data based on the trained samples.

The further enhancement to this work could take place in finding the information about time of capturing images having shadows at particular cases. Lot of redundant data could be obtained if the scene is reconstructed in orthographic view. So the future work could be in the field of Data Compression as well.

Input Data Set					Output Data Set	
e / l	**h_n / h_{n+1}**	α	θ	**D**	ϕ	**d**
0.9929	1.0714	6	8	1.9	10	2
0.9583	1.5625	17	17	5.6	20	9
0.9340	2.2727	32	19	4.8	30	12
0.7564	1.3888	40	40	1.7	40	3
0.7812	2.1428	27	39	2.7	45	8
0.8461	1.9230	39	31	3.1	50	9
0.4857	2	108	60	0.8	60	3
0.1639	3.3333	35	80	0.7	70	6

Table 5.1: Only a few of the input and their corresponding output data set shown here of various range.

If the obtained geometrical information from one scene could be depth information with dimensions, then with the integration of views, this could also be thought of in the field of object recognition.

Image mosaicing is a branch of cognition technology, which stitches several images into a single image. Thus this work stands bright in the image mosaicing field as well.

References

[1] O. D. Faugeras. Three - Dimensional Computer Vision: a Geometric Viewpoint. MIT Press, 1993.

[2] A detailed Notes on Projective Geometry by BalazsCsikos. http://www.cs.elte.hu/geometry/csikos/proj/proj.html

[3] Single view metrology by Antonio Criminisi http://www-2.cs.cmu.edu/~ph/869/papers/Criminisi99.pdf

[4] G. Gracie. Analytical photogrammetry applied to single terrestrial photograph mensuration. In XI International Conference of Photogrammetry, Lausanne, Switzerland, July 1968.

[5] O. D. Faugeras. Stratification of three-dimensional vision: Projective, affine, and metric representation. Journal of the Optical Society of America, A12:465-484, 1995.

[6] Elements of Photogrammetry – by Paul R Wolf. Mc Graw-Hill International edition – II edition.

[7] (neural nets book)

[8] Y. R. Tsai. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. IEEE Journal of Robotics and Automation, RA-3(4):323-344, August 1987.

[9] B. Caprile and V. Torre. Using vanishing points for camera calibration. International Journal of Computer Vision, pages 127-140, 1990.

[10] Elements of Engineering Drawing – N. D. Bhatt. Charotar Publishing House – Twenty ninth revised edition in first angle projections.

[11] R. I. Hartley. Estimation of relative camera positions for uncalibrated cameras. In Proc. European Conference on Computer Vision, LNCS 588, pages 579-587. Springer – Verlag, 1992.

Appendix

Fig. A.1 Corresponds to $\phi = 10^0$

Fig. A.2 Corresponds to $\phi = 20^0$

Fig. A.3 Corresponds to $\phi = 30^0$

Fig. A.4 Corresponds to $\phi = 45^0$

Fig. A.5 Corresponds to $\phi = 60^0$

Fig. A.6 Corresponds to $\phi = 70^0$

Fig. A.7 Corresponds to Test Data (different) and Result of the same

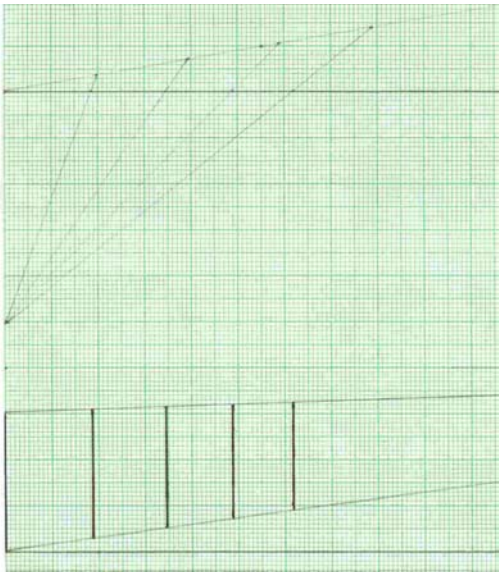


Fig. A.1

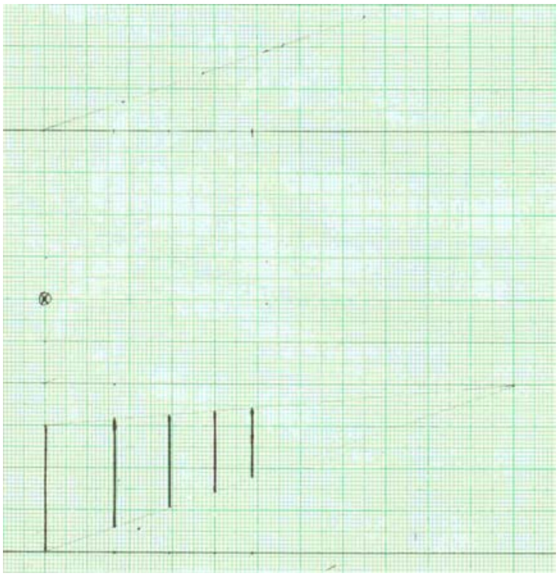


Fig. A.2

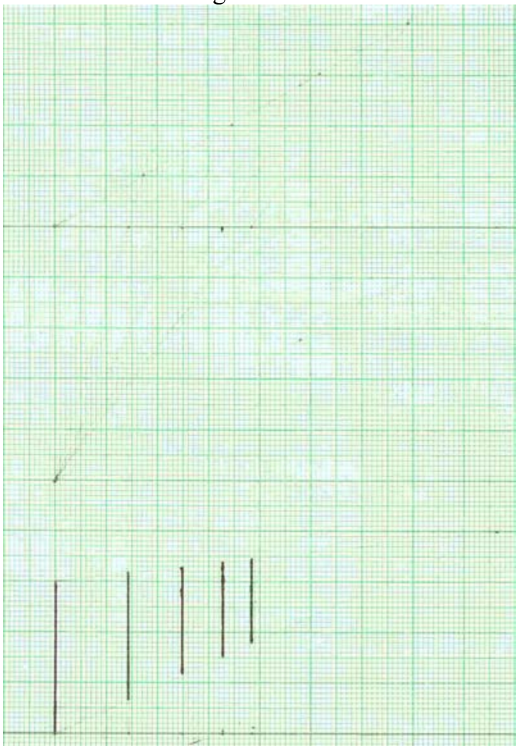


Fig. A.3

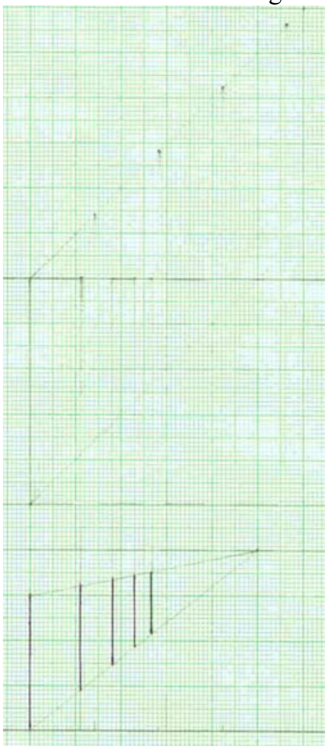


Fig. A.4

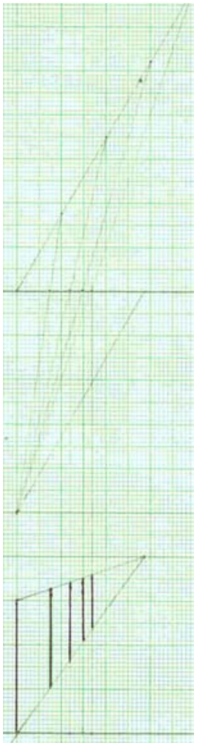


Fig A.5

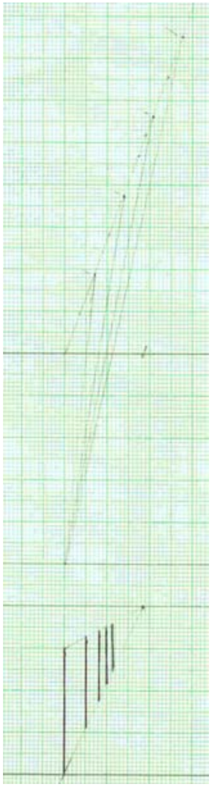


Fig. A.6

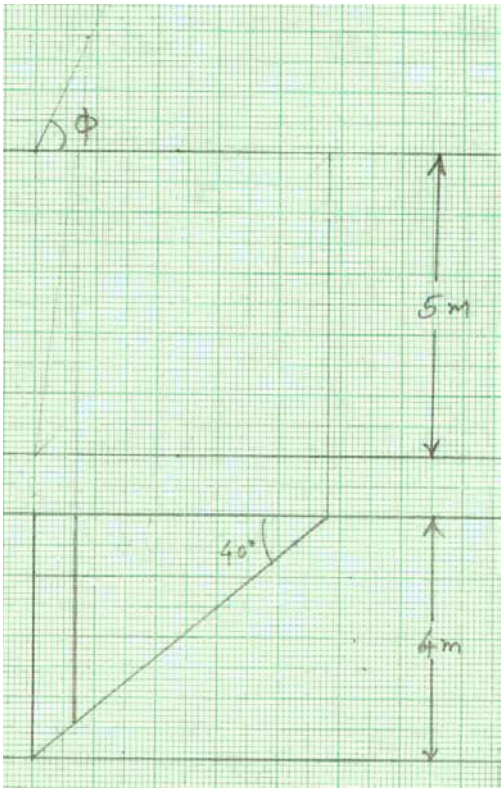


Fig. A.7