Deep Learning based approach for Range Estimation

Samvram Sahu
Department of Avionics
IIST
Trivandrum, India
samvram.iist@gmail.com

Anurag Shukla
Department of Avionics
IIST
Trivandrum, India
aanurag97@gmail.com

V. Adithya Krishnan
Department of Avionics
IIST
Trivandrum, India
v.adi_1997@yahoo.co.in

Pradipta Roy
Directorate of EOTS
Integrated Test Range, DRDO
Chandipur, India
pradipta32@gmail.com

***Abstract*—Range analysis is one of the most sought after topics in missile technology. Deterministic and statistical methods have been used for estimation of the range of a missile. The present industrial implementation involves use of redundant sensors and apparatus at two far-off sites which has always been a problem. In this paper, we find a method to estimate the range of a missile with the use of a single camera. Deep Learning based depth mapping has provided appreciable results for shorter ranges, but this method remains unexplored for longer ranges. A Convolutional Neural Networks(CNN) based approach fares equally well with the other methods of monocular depth estimation such as linear regression, support vector machine based regression, decision tree regression by treating the range estimation problem as a regression problem.**

***Index Terms*—CNN, Monocular Vision, Range Estimation, Regression**

I. Introduction

Missiles are the backbone of air defence sector and have been a major budget consumer of any country's defence line-up. But more than designing a missile and successfully launching it, tracking of a missile has been a major challenge faced by the defence sector. Tracking provides the information relating to the functioning of the missile and its trajectory. Conventionally this has been achieved by the use of RADARs or Telemetry applications. But with the advent of Computer Vision and its implementation into modern Defence Systems, methods like Electro-Optical Tracking System(EOTS) have become more popular.

EOTS is basically a camera mounted on a platform which is compatible of detecting a missile in its scope, adhering a lock-on and then tracking it, mostly using Infrared(IR) imaging. During the tracking phase the camera needs the object of interest to stay in sight and hence it moves its optical axis to align with the object. The control of this device is dependent on the feed from other EOTS devices, and/or RADARs which track the missile. Usually another EOTS device is used to triangulate the position and hence find the range of the missile. This is a deterministic method and hence the error introduced are mostly attributed to measurement errors. However, the high accuracy comes at the cost of an immense number of sensors/devices used for this method.

This paper however intends to overcome the above stated problem by using monocular vision for range estimation by using various regression techniques. The paper covers all the work that has been done on making a robust regression algorithm which can estimate approximately how far a missile is in a given image. It is an application specific design that adheres to the industrial requirements of Integrated Test Range(ITR), Defence Research and Development Organisation(DRDO). The various Regression designs use techniques like Support Vector Machine(SVM), Tree Regressor and Linear Regressor. Deep learning approach is taken in the form of CNN and ANN(Artificial Neural Networks) Regressors.

The collaborating lab and its apparatus have been discussed further. An array of methods were tested and an intuitive comparison is provided towards the end.

A brief overview of our work is presented in this paper. *Related Work* consists of the preliminary work we reviewed in order to work on this topic, *Methodology* covers the various approaches to this problem, which is *Deterministic Approach*,*Statistical Approach* and *Deep Learning based Approach*, which in turn contains *ANN* and *CNN* based methods. in *Results & Discussion*, we discuss the observations and their corresponding inferences.

II. Related Work

A lot of research and methods are being developed for the depth estimation problem due to its use in various fields like robotics, defence, Computer Vision, and many more. Log mapping method uses complex log mapping to measure the distance between the camera and the object. Two images from two different camera positions are used in this method. Ratio between the objects sizes projected on the two images that are moved on the cameras optical axis is used for the calculation. Another such intuitive method for object distance calculation is based on an objects position on the image. This method does not depend on the objects size. This was further researched upon by changing the pitch angle of the camera. Photometric visual servoing is a new technique to overcome the problem of the object tracking process. In photometric visual servoing, the tracking process is no longer required, since the image intensity (the pure luminance signal) is sufficient to control the robots motion. Image gradient and image entropy have the same approaches as photometric visual servoing. The image gradient technique is based on the extraction of information of an image which is located in its high frequency areas(contours). Marchand and Collewet [9] applied a method to use the square norm of the gradient obtained from all of the pixels in an image as visual features in visual servoing. Wang and Liu [8] proposed a new visual servo control technique for the robotic manipulator, whereby a back propagation neural network would make a transition from the image feature to joint angles. Yet another method was the triangulation technique(active or passive). The active triangulation method emits a signal and then measures the reflected signals, whereas the passive triangulation method uses the background illumination.

III. Methodology

A. Deterministic Approach

The method requires some additional prerequisites before being applied as the range estimation requires few values for the geometrical analysis. The height of the camera is required for calculating the range. The angle of the camera with respect to the horizontal should be known. These two parameters are very important in the calculation of the range. This maybe considered as a demerit as the height of the camera or the height of the object may not be available.

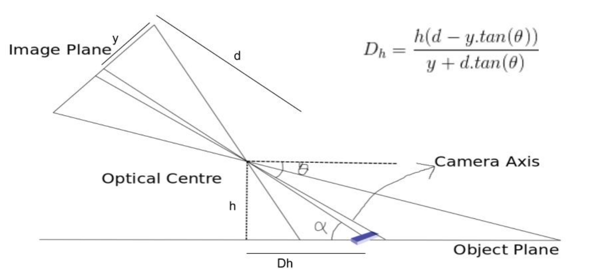

Fig. 1. Experimental setup showing the arrangement of camera and the object

The experimental setup is shown in Figure 1, Image is formed on the $Image\ plane$ where the CCD sensors reside, where y is distance of the object's image from the bottom of the sensor in the image plane, d is the focal length of the lens used in the camera, or simply the distance between the centre of the lens and the $Image\ plane$, θ is the angle of the depression of the optical axis from the local horizontal, and α is the angle of elevation from the object towards the centre of the lens.D_{Hy} is the y-direction or vertical distance from camera bottom, h is the height of the camera from ground, y is the image centroid distance from image centre in image plane, f is the focal length of the camera lens.The object is later shifted to various other positions and a dataset of ranges are calculated using our geometrical method. The ranges were checked manually also later. Theta(θ) is kept constant through out the experiment and it is assumed that the parameters f and d remain constant for all the images.

$$\begin{aligned} tan(\alpha) &= \frac{h}{D_{Hy}} \\ tan(\psi) &= \frac{y}{f} \end{aligned} \tag{1}$$

$$\begin{aligned} \theta + \psi + 90^\circ - \alpha &= 90^\circ \\ \alpha - \psi &= \theta \end{aligned} \tag{2}$$

$$tan(\theta) = \frac{tan(\alpha) - tan(\psi)}{1 + tan(\alpha) * tan(\psi)} \tag{3}$$

Solving equations $1, 2, 3$ we get -

$$D_{Hy} = \frac{h(d - y * tan(\theta))}{y + d * tan(\theta)} \tag{4}$$

Similar method can be used for distance along x-direction or the perpendicular plane. The equation comes out to be following:

$$D_{Hx} = D_{Hy} \frac{x - 0.5x}{f} \tag{5}$$

The main problem with geometric method is its inability to deal with long range objects. This can be seen mathematically in equation 6.

$$\begin{aligned} &\theta \to 0 \\ &y \times tan(\theta) \approx d \times tan(\theta) \to 0 \\ \Rightarrow D_{H_y} \approx &\frac{h \times d}{y} \end{aligned} \tag{6}$$

As object is placed far away from the camera, to capture the object in frame, the angle of depreciation need to be close to 0. Taking the limit θ tends to 0, we observe that the values h and d are fixed and of finite length. But y being the distance of centroid from image centre, will tend to 0 as object will lie more and more closer to the centre of the image frame. This will result in undetermined fraction form will shows that as the range increases, the error will diverge.

B. Statistical Approach

In this approach, we estimate the range based on some prior features. However we needed a set of features which perform as an input corresponding to which we can use the range data as an output, for devising a function or a transform.

Features used: We need to form a feature vector(F_v) from $M_{321\times321}$, and use them as inputs for regression problems. We designed a MATLAB function named featureExtractor which takes an image input and gives a feature vector as the output.

The output essentially consists of a vector which has 4 elements, i.e.

- Size of the object
- Average Intensity of the object
- Relative Intensity of the object
- Eccentricity of the object

1) Linear Regression: We optimize a weight vector(w), s.t. $w^T F_v \approx R$. The cost function is Root Mean Square Error which has to be minimized.

For the case of Linear Regression we get an optimal RMSE of 25.831 kilometres which is pretty huge and this is an unacceptable regressor for our case.

2) Support Vector Machine(SVM)) regression: Support Vector Machine can also be used as a regression method, maintaining all the main features that characterize the algorithm (maximal margin).

RMSE for linear SVM is more or less similar to the Linear Regressor.

3) Decision Tree Regression: Decision tree builds regression or classification models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes. Decision trees can handle both categorical and numerical data.

The RMSE of this regressor is the least and is in the order of 4 kilometres.

C. Deep Learning Approach

CNNs are good at learning most complex functions. Using a similar approach as the human brain, we design an ANN based regressor which can estimate the range given an image based on the previous values.

1) Artificial Neural Network(ANN): An ANN is based on a collection of connected units or nodes called artificial neurons which loosely model the neurons in a biological brain. Each connection, like the synapses in a biological brain, can transmit a signal from one artificial neuron to another. An artificial neuron that receives a signal can process it and then signal additional artificial neurons connected to it. We try various architectures and freeze the optimal architecture with 1 hidden layer. The number of neurons in the hidden layer is varied till we obtain the optimal RMSE which occurs for the case of 20 neurons.

2) CNN based Approach: CNNs extract features used for regression themselves from the image. Mainly two constraints on architechture are taken from [7].

1) The ratio of real filter size of convolution to the receptive field size should not fall below $\frac{1}{6}$.
2) The receptive field size of the last neuron should not exceed the input image size.

Constant hyper-parameters while tuning

- Mini Batch Size - 100
- Validation Frequency - After every epoch
- Optimizer used - Stochastic Gradient Descent with Momentum
- Maximum Epochs - Experimentally determined till when network does not overfit
- Initial Learn Rate - 0.001

With an initial network we get a RMSE of around 20 kilometres which is still unacceptable, and hence we increase the depth of the architecture which can capture more complex features. The modified network clearly fails for small value of ranges where the network predicts wrong values. We expect this network to outperform the rest. However the results suggest otherwise, this network produces the worst results among all the networks and we infer that depth is not a necessary criterion for regression which concurs with results in [7]. With this we take a hint that the depth should be reduced in order to deal with the problem exactly. Hence we come up with the following architecture.

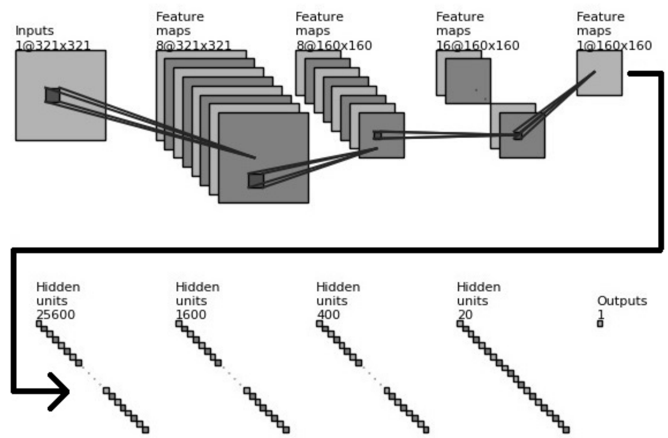

Fig. 2. Tuned Network for Application

When the network in figure 2 is trained, the best solution amongst all neural networks is obtained.

The architecture goes as follows,

- Input Layer
- Convolution Layer - 8
- Convolution Layer - 16
- Max pooling - 1600
- Max pooling - 400
- Max pooling - 20
- Max pooling - 1
- Regression Layer

TABLE I
METHOD COMPARISON

Specification	**Decision Tree**	**ANN**	**CNN**
Feature Extraction	Manual	Manual	Auto
Training Time	Less	Comparable	High
Scalability	Low	High	Highly scalable
Adaptability	Low	Comparable	Highly adaptable
Data Set Required	Medium	Large	Large

*Discrptions are comparitive and not to be taken absolute.

Few methods prove to be superior than others in some aspects, whereas others perform better in different conditions, table I briefly shows the different performance aspects.

IV. Results & Discussion

The different methods stated before, we analysed their performance on same set of data to get a comparative results for all the methods thus conclude the best suited method keeping in mind the constraints and extend of the problem statement.

The linear regressor(refer to figure 3) as expected didn't give a good results thus indicating towards the fact that the problem statement does not follow a linear behaviour.

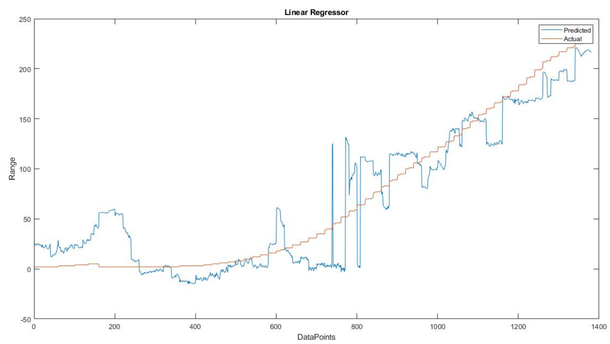

Fig. 3. The predicted range of linear regressor vaguely follows the actual range traversed, this shows non-linear dynamics in system

The use of SVM was then tried out in two parts. At first, we used a linear SVM regessor(refer to figure 4) which like the linear regressor, didn't give a good result but some improvement in the RMSE values were observed. The second part was done using a non-linear SVM regressor(refer to figure 5) which showed a little bit of improvement but still not satisfactory. This is because the root problem with SVM that it does not very well deal with scalability of features of the data set.

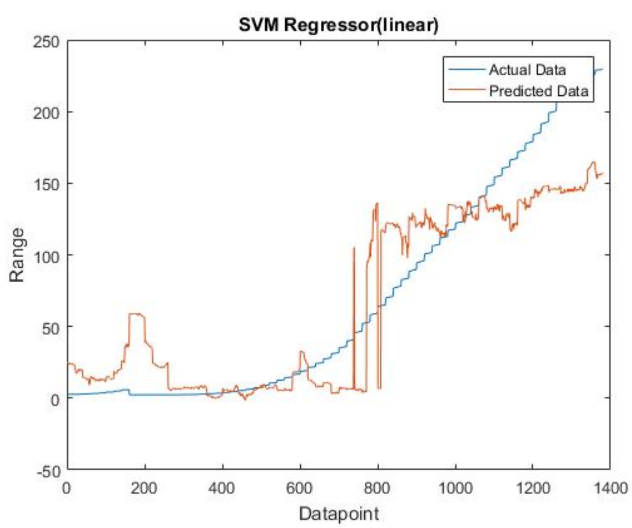

Fig. 4. SVM is a good choice for learning complex functions, using kernel tricks. Linear SVM's predicted range is better than linear regressor, but fails at higher ranges

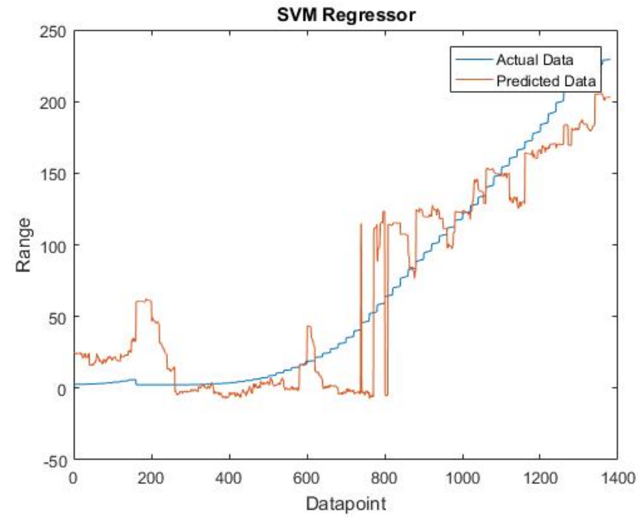

Fig. 5. Non-linear SVM performs similar to the linear SVM with improved performance at farther ranges, both fail at mid range(stage separation is bottleneck)

The decision tree regressor(refer to 6) used gave the best results among all other deterministic methods. This regressor divides the problem space into sub-regions till it can approach it for a solution hence attaining the level of accuracy. But, this is extremely sensitive to small perturbations in the data: a slight change in the data fed(from a different site) can result in a drastically different tree. Also it can easily over-fit which is an issue.

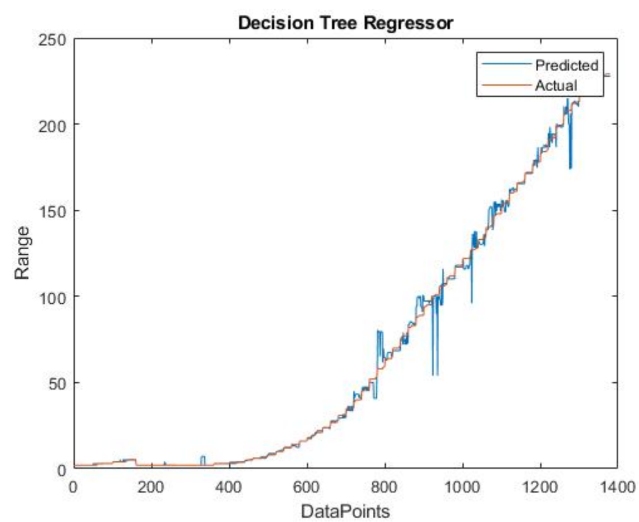

Fig. 6. Dividing the problem space into smaller regions provide high enough accuracy yet validation error is too high, i.e. it does not regularize well

Moving on to statistical approach, an ANN with different number of neurons was tested. The optimal number of neuron was found out to be 20 (varied from 5 to 100). The results were decent compared to other methods(refer to 7).

The second statistical method used was the use of CNN. After analysing different architectures, the best results are shown here(refer to figure 8).

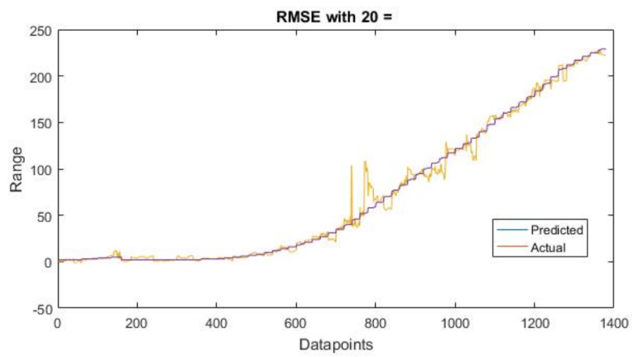

Fig. 7. The best result for manually extracted features can be seen by ANN with 20 hidden neurons in a single hidden layer

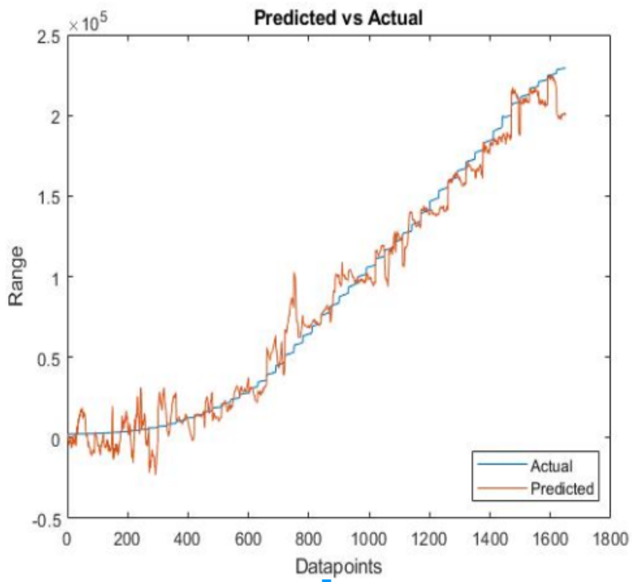

Fig. 8. On proper tuning of the network architecture we get a robust design which regularizes well as well as provides a decent estimate of the range

TABLE II
ERROR RESULTS

Method	RMSE(Km)
Linear Regressor	25.086
Linear SVM	23.108
Non-Linear SVM	21.186
Decision Tree	4.1946
ANN	7.0915
CNN	10.35

*For CNN, best result is shown.

The table II provides the Root Mean Squared Error(RMSE) values of predicted ranges from different methods used.

V. CONCLUSION

In this paper, we have seen how good are various methods that can be used for range estimation from an image. This helps us to know the criteria for choosing a particular method for task specific operations. The results of different regressors were compared to give a good idea about each method. One can easily try out these regressors for their specific task, allowing an easy exploration range estimation. The versatility of deep learning helps to extend the problem statement to great limits. The future work may include getting better optimised architecture and extending the problem statement to more general conditions including weather and atmospheric effects on feature extraction. There is also a prospect of piece-wise regressors i.e. depending on the performance of regressors in each range interval we can implement different algorithms to further enhance and improve the performance of our regressor.

ACKNOWLEDGMENT

We acknowledge the opportunity provided by Integrated Test Range, DRDO, also for the lab provided and maintained by Mr. Pradipta Roy, Scientist - E, ITR-DRDO our guide. We also acknowledge the guidance provided by Dr. Deepak Mishra, Associate Professor, IIST and Mr. Litu Rout, Scientist - C, ISRO during the research.

REFERENCES

[1] Peyman Alizadeh, "Object Distance Measurement Using a Single Camera for Robotic Applications", Laurentian University, Canada
[2] Siddharth Mahendran, Haider Ali, Rene Vidal "3D Pose Regression using Convolutional Neural Networks",
[3] Philipp Fischer, Alexey Dosovitskiy, Thomas Brox "Image Orientation Estimation with Convolutional Networks", GCPR-2015 bibitemb4 Z Said, K.Sundaraj, M.N.A.Wahab "Depth Estimation for a Mobile Platform Using Monocular Vision", Procedia Engineering Volume 41, 2012, Pages 945-950
[4] J. Schennings "Deep Convolutional Neural Networks for Real-Time Single Frame Monocular Depth Estimation" - Uppsala University, December 2017
[5] A Shibata, A Ikegami, M Nakauma, M Higashimori "Convolutional Neural Network based Estimation of Gel-like Food Texture by a Robotic Sensing System", MDPI December, 2017
[6] Zhenxing Niu, Mo Zhou, Le Wang, Xinbo Gao , Gang Hua "Ordinal Regression with Multiple Output CNN for Age Estimation" CVPR - 2016
[7] Xudang Cao, "A practical theory for designing very deep CNNs"
[8] Wang, H. B. and Liu, M. , Design of Rob otic Visual Servo Control Based on Neural Network and Genetic Algorithm , International Journal of A utomation and Computing , vol. 9, no. 1 , pp. 24 - 29, 2012.
[9] Collewet, C. Marchand, E. Photometric visual servoing IEEE Trans. on Robotics 274 828-834