TECHMIND RESEARCH
ISSN No. 2291-2118
International Journal of Research in Business and Technology
Volume 3 No. 1 August 2013

# A Novel Method to Improve Model fitting for Stock Market Prediction

Sukriti Jain[1], Samarth Gupta[2], Amarjot Singh[3]
[1]Dept of Electronics and Communication Engineering,
Ambedkar Institute of Advanced Communication technologies and research, GGSIPU, India
Email: su_kriti24@yahoo.in
[2]Dept of Mechanical and Production Engineering, IIT Rookee, India.
Email: samarume@iitr.in
[3]Dept of Engineering, Simon Fraser University, Burnaby, Canada.
Email: asa168@sfu.ca

**Abstract-** *Forecasting the trends of stock market is of extreme importance and profitable to stock market traders and also to the researchers who are always trying to find an analogy to describe the behaviour of stock market. Various data mining techniques have been implemented in the recent past to predict the behaviour of stock market. Every method tries to fit a model to the training data to predict the future. As it is obvious the accuracy of prediction depends on the model fitting. We propose a linear model for stock market prediction and further elaborate on improving the fit of the model. The proposed model and the correction method are tested on Istanbul stock exchange. The proposed model fits the dataset with an average error of 12% which is corrected by proposed fitting method to an average of 6%.*

**General Terms-** *Data mining, stock market trend prediction*

**Keywords-** *Stock market, Gaussian Fitting, Linear Models, Technical indicators, Istanbul Stock Exchange Dataset.*

## 1. INTRODUCTION

Stock market forecasting has always fascinated people due to large profits which one can earn in a very short duration of time. Various attempts have been made in the past to predict the behavior of stock market but results obtained were not very useful. A major reason is the inability of the models to fit on the previous data accurately. Although the previous studies have shown that it is possible to forecast the behavior with a good amount of accuracy. We believe the backbone of the algorithm to predict the future depends on its ability to fit the past accurately. Hence, we propose a method which can be used to improve the fitness accuracy on the training data. In order to prove that the improvement in training data, we propose a linear model this will be used to fit the training data. Later, the accuracy of the model fitting the training data is improved using the corrective measure. In this study, we use the Istanbul stock exchange dataset which includes 8 technical indicators available at [1]. These technical indicators are then used to predict the future behavior of the stock market. The dataset used is from Istanbul Stock Exchange, from 5th June 2009 till 22nd Feb 2011.

The paper is divided into following sections. The second section explains the previous work done in the field while the third section elaborates the attributes of the dataset used in the experimentation. The fourth section explains the proposed algorithm followed by results explained in the fifth section. The conclusion is presented by the sixth section and the future work is mentioned in the seventh section.

## 2. LITERATURE SURVEY

Artificial neural networks have been used extensively to forecast the financial time series. In [3], the efficiency of time delay, recurrent and probabilistic neural networks for forecasting of stock trends based on previous data sets has been stated. The data set consists of closing price of the daily shares. In [4] technical indicators were taken as inputs to the neural network for forecasting stock exchange dataset. In [5], to make a decision support system, technical indicators and a neural network was used for exchange traded funds trading. Technical indicators and neural networks were used in [6] to predict the US Dollar Vs British Pound exchange rates. In [7] automatic trading algorithms were presented as a framework. In [8] a back propagation neural network was used to forecast the buy/sell price for a stock and then applied a case based dynamic window to improve the forecast accuracy. In [2] a survey of more than hundred articles which used neural networks and neuro-fuzzy models for predicting stock markets was presented. It was observed that soft computing techniques outperform conventional models in

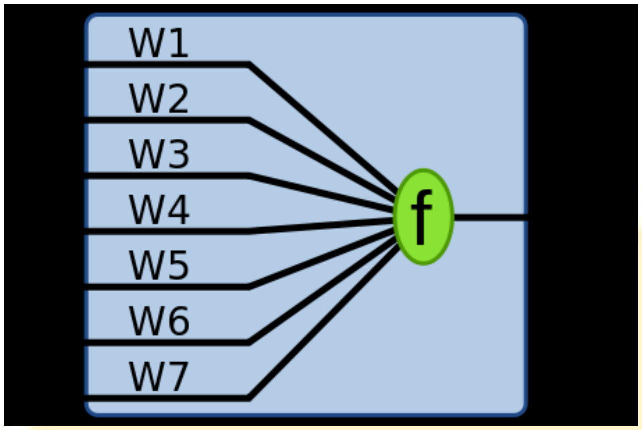


**Fig. 1: The proposed Linear Model**

most cases. Defining the structure of the model is however, a major issue and is a matter of trial and error. In [9], review of data mining applications in stock markets was presented. [10] Used a two-layer bias decision tree with technical indicators feature to create a decision rule that can make recommendations when to buy a stock and when not to buy it. [11] Joined the filter rule and the decision tree technique for a particular stock trading. [12] Presented a hybrid model for stock market forecasting and portfolio selection. In [13] a hybrid fuzzy time series model with cumulative probability distribution approach and rough set rule induction was proposed, to forecast stock markets.

## 3 DATASET ATTRIBUTES

We used the Istanbul dataset for experimentation in this article. The dataset uses 8 attributes to predict the behavior of stock exchange. The attributes of the dataset are described as below.

### 3.1. Istanbul stock exchange National 100 index

The Istanbul Stock Exchange National 100 Index constitutes of national market enterprises resulting into a capitalization-weighted index. The components of National 100 Index are chosen on pre-determined criteria directed for the enterprises to be included in the indices. The base date is January 1986 and base value is 1 for the TL based price.

### 3.2 Istanbul stock exchange Standard 500 return index

The performance of domestic brands is measured on a capitalization-weighted index of 500 stocks. It evaluates the performance by measuring the changes in the aggregate market value of 500 stocks representing important industries. The index was derived with a base level of 10 for the 1941- 43 base periods.

### 3.3. Stock market return index of Germany

The German stock index comprises of 30 important German enterprises trading on Frankfurt Stock. The growth of these is measured on by the operator of Xetra in terms of order volume and market capitalization. The DAX index’s (German benchmark) performance is represented by the L-Dax index after the electronic trading is closed.

### 3.4 Stock market return index of Japan

Tokyo Stock Price Index keeps track of domestic enterprises of the exchange's First Section. In this system the weight of a company depends upon on the total number of shares outstanding to a weighting based on the number of shares available for trading. This measure has a major effect on the weighting of companies because multiple companies in Japan have significant holdings of shares of their business partners involving complicated business alliances, hence these shares are removed from the weight of companies in the index.

### 3.5 Stock Market Index of UK

The Stock exchange of UK or London Stock Exchange came into existence in 1698 and is home to majority of largest companies and best-known companies in the world. The stock exchange comprises of 1,400 companies on the Main Market with a total market capitalization of £3.7 trillion. Companies of all types, nationalities and sizes together represent some 40 sectors.

### 3.8 Stock market return index of Brazil

The Stock market return index of Brazil is a gross total return index weighted by traded volume: consists of the most liquid stocks traded on the Sao Paulo Stock Exchange. The Ibovespa Index has been divided 10 times by a factor of 10 since Jan 1

### 3.7 MSCI European index

The MSCI Europe Index includes 15 developed market (DM) countries across Europe capturing large and average enterprises. Market capitalization across the European Developed Markets equity trade with 436 constituents covering 85% of the free float-adjusted.

### 3.8 MSCI emerging markets index

In 1988, the first comprehensive Emerging Markets Index was launched by MSCI. The MSCI Emerging Markets (EM) Indices have evolved significantly over time moving from around 1% of the global equity opportunity set in 1988 to 13% in 2012.

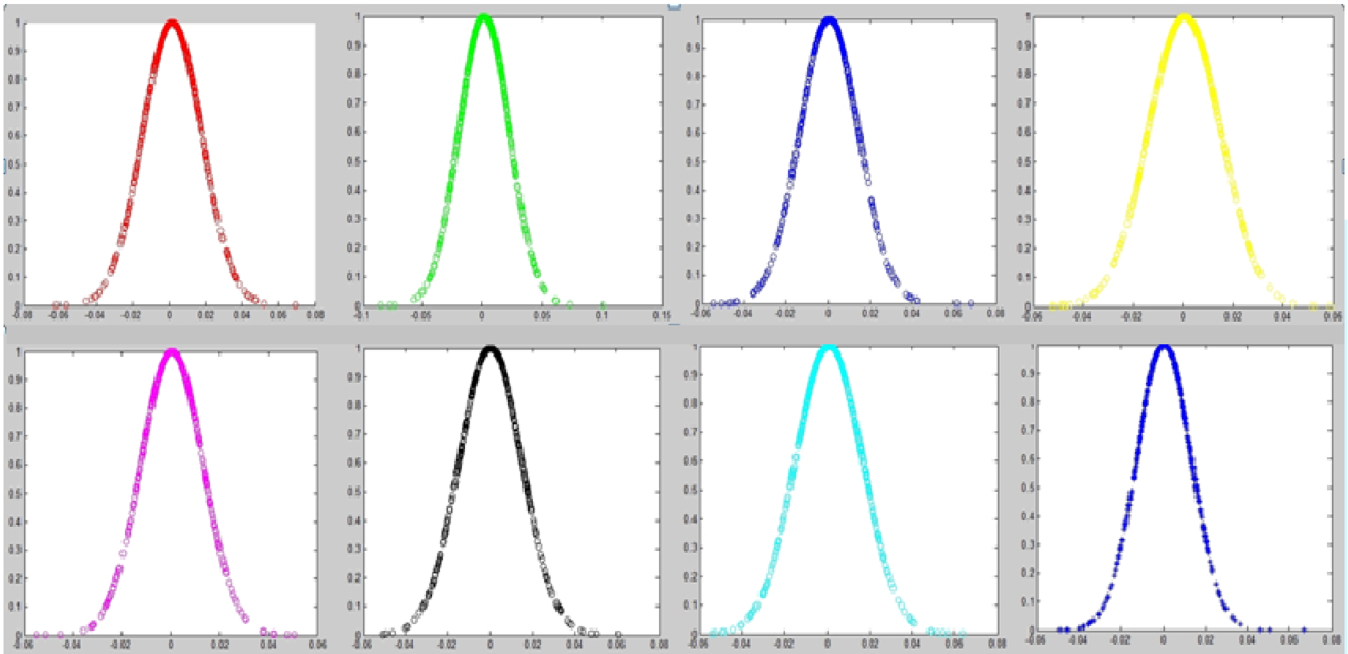

**Fig. 2: Represents each attributes as a Gaussian distribution. The attributes are (a) Istanbul stock exchange National 100 index (In Red) (b) Istanbul stock exchange 500 return index (In Green) (c) Stock market return index of UK (In Yellow) (d) Stock market return index of Japan (In Magenta) (e) Stock market return index of Brazil (In Black) (f) Stock market return index of Germany (g) MSCI European index (In Aqua-Blue) (h) MSCI emerging markets in**

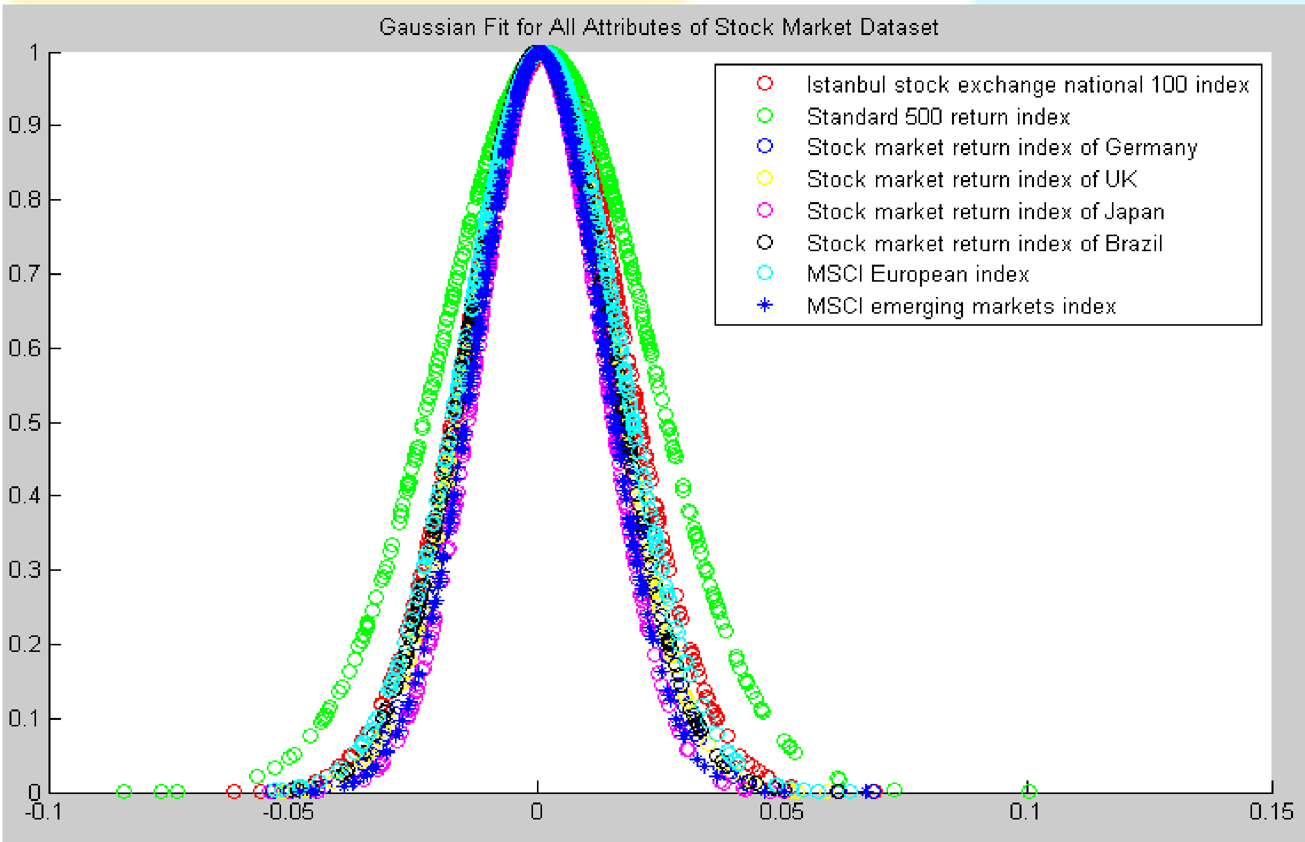


**Fig. 3: Combined Gaussian Fit for All Attributes of Stock Market Dataset**

## 4 PROPOSED ALGORITHM

The proposed algorithm is divided into a two step process. The first section fits a linear model motivated from single layer perceptron on the attributes of Istanbul dataset to compute the weights. The second step fits the Gaussians on the weighted attributes to compute their relative contribution to the output.

### 4.1 Linear Model

The primary aim is to fit a model to the stock market dataset. We develop a linear model motivated from single layer perceptron to fit to the dataset as shown in fig. 1. The model's input are the eight attributes as explained in section: 3. The output of the model is a function defined below:

$$f = \left\{ \sum_{i=1}^{8} X_i W_i \right. \quad (1)$$

Where $X$ represents attribute while $W$ represents weight for the respective attribute. The model uses training data to compute the weight for each attribute.

### 4.2 Gaussian Fitting

Once the weights have been computed using the first step of the algorithm, we compute the weighted contribution of each attribute to the final output: we call it $Wei\_Att$ and is equal to:

$$Wei\_Att = X.*W \quad (2)$$

Once weighted attribute has been compute for each individual attribute, we fit the Gaussian model to every $Wei\_Att$ to get the relative contribution of each weighted attribute. The weighted attributes which contribute least to the output are dropped to obtain an output with smaller error.

## 5 RESULTS

The simulations are performed on an Intel core 2 Duo 4 GB RAM machine. Matlab software was used to perform the simulations. UCI Istanbul Stock exchange dataset is used for experimentation.

In the first step, we train our linear model using the training data taken from the Istanbul stock exchange dataset. After train the model for 67 iterations, we obtained the weights for all the attributes as shown in table. 1

**Table. 1: Represents weights for respective attribute**

| Attribute | Weight |
|---|---|
| ISE | 0.2034 |
| ISE-500 | 0.3503 |
| SP | -0.2127 |
| DAX | -0.1287 |
| FTSE | 1.0438 |
| NIKKEI | -0.0883 |
| BOVESPA | 0.3673 |
| EU | -1.2223 |

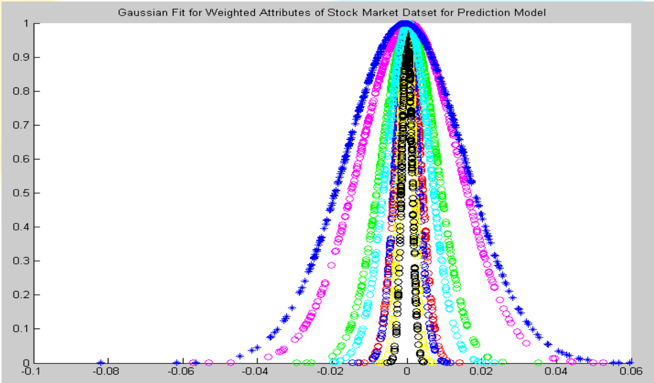


**Fig. 4: Combined Weighted Gaussian Fit for All Attributes of Stock Market Dataset**

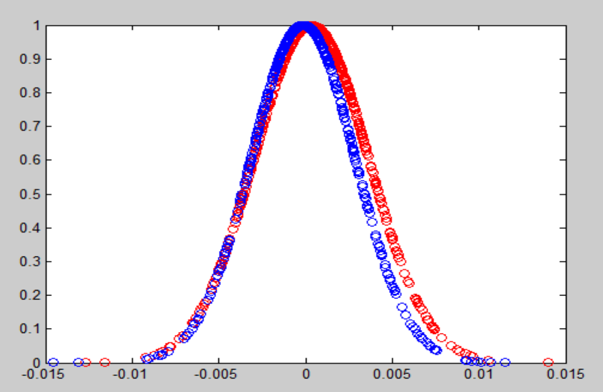

**Fig. 5: Represents high correlation between ISE 100 index and the stock market return index of Germany**

After computing the weights, we computed the weighted attribute for each original attribute of the dataset using the eq. 2. We fit Gaussian distribution to each variable to understand their behavior as shown in fig. 2. As seen from the graphs it is almost impossible to make any judgment of the data based on individual parameters distribution. Hence, we try to analyze the distribution together as shown in fig. 3. The data is highly correlated and can't be understood directly. Further, we fit Gaussian to the data obtained from our linear model as shown in fig. 4. It is evident from the fig. 4, that the stock market return index of brazil has least contribution towards stock market variation while the maximum contribution comes from MSCI emerging markets index.

It is observed that the Istanbul stock exchange National 100 index and the stock market return index of Germany have high correlation as shown in fig. 5. In addition, the distribution that stock market return index of brazil and stock market return index of UK also have high correlation as shown in fig. 6.

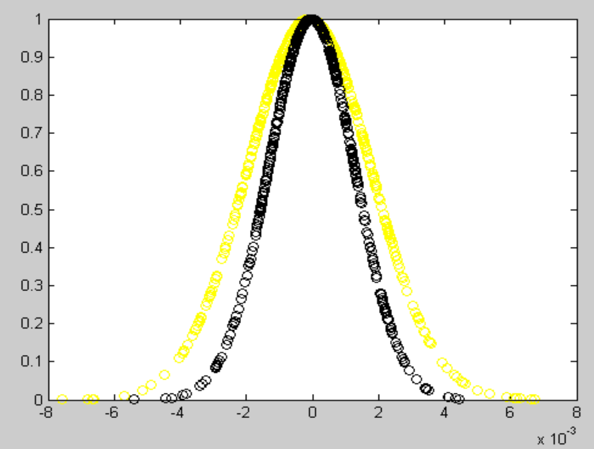


**Fig. 6: Represents high correlation between stock market return index of Brazil and UK**

High correlation indicates similar contribution to stock market variation while low contribution indicates low input towards final stock market variation. As seen from fig. 3, this kind of analysis is extremely difficult if the Gaussian fit is applied to attributes directly and is only possible by our modeling.

Further this information can be used to avoid the attributes with least contribution to the final output. We primarily removed the attributes with minimum contribution from the equation leading to an average enhancement in accuracy of 6%. This proves that the fitting of the model with removal of attributes with minimal contribution increases the accuracy. In order to provide a better understanding of the enhancement in error, we present an example. The weighted attributes of the 6th row are as

**Attributes** =

-0.0059 -0.0148 0.0049 0.0017
-0.0052 0.0043 -0.0198 0.0152

The ground truth Emerging Market Index (EM) is -0.02263. According to our linear model, the weighted attributes should be summed to obtain EM index.

$$Sum\ (Attributes) = -0.0197;$$

$$Error = \frac{(-0.0197) - (-0.02263)}{(-0.02263)};$$

Error = -12.83%;

As discussed above the contribution of NIKKEI is extremely low towards EM and hence should be removed from the linear model. Once NIKKEI has been removed from the sum the value of sum becomes:

$$Sum\_Modified = -0.0197 - 0.0043$$
$$= -0.0240$$

Improving the error from 12.83% to:

$$Error = \frac{(-0.02407) - (-0.02263)}{(-0.02263)}$$

$$Error = -6.19;$$

## 6 CONCLUSION

The article presents a novel method which can help in improving the model fitting which eventually will produce better results for the future predictions. The fitting is improved by dropping the data with least contribution resulting into an improvement on fitting accuracy or accuracy of training data.

## 7 FUTURE WORK

The future work will focus on improving the accuracy by taking into account the correlation of all attributes of data also. In this article we improved the accuracy by only considering the attribute with minimum contribution to the stock prediction.

## REFERENCES

[1] http://archive.ics.uci.edu/ml/datasets/ISTANBUL+STOCK+EXCHANGE
[2] Atsalakis G. S., and Valavanis K. P., 2009 Surveying stock market forecasting techniques – part II. In soft computing methods Expert Systems with Applications.
[3] Saad E. W., Prokhorov D.V., and Wunsch, D.C., 1998 Comparative study of stock trend prediction using time delay, recurrent and probabilistic neural networks. In IEEE Transactions on Neural Networks, Vol. 9, No. 6, pp. 1456-1470.

[4] Lee, C-T., and Chen,Y-P. 2007. The efficacy of neural networks and simple technical indicators in predicting stock markets. In Proceedings of the International Conference on Convergence Information Technology, pp.2292-2297.

[5] Kuo, M-H., and Chen, C-L. 2006. An ETF trading decision support system by using neural network and technical indicators. In Proceedings of the International Joint Conference on Neural Networks, pp. 2394-2401.

[6] Nagarajan,V., Wu,Y., Liu,M. and Wang Q-G. 2005 Forecast Studies for Financial Markets using Technical Analysis. In Proceedings of the International Conference on Control and Automation (ICCA2005), pp. 259-264.

[7] Bansal, Archit, Mishra ,Kaushik, Pachouri, Anshul, 2010 Algorithmic Trading (AT) - Framework for Futuristic Intelligent Human Interaction with Small Investors. In International Journal of Computer Applications, vol. 1, no. 21, pp.01-05.

[8] Chang, P.-C. , Liu, C.-H. , Lin, J.-L. , Fan, C.-Y. , & Celeste, S.P. Ng. 2009 A neural network with a case based dynamic window for stock trading prediction. In Expert Systems with Applications, vol.36, pp.6889–6898.

[9] Setty, Venugopal D., Rangaswamy, T.M. and Subramanya, K.N. 2010. A Review on Data Mining Applications to the Performance of Stock Marketing. In International Journal of Computer Applications, vol. 1, no. 3, pp.33-43.

[10] Wang, J-L. , and Chan, S-H. 2006 Stock market trading rule discovery using two-layer bias decision tree. In Expert Systems with Applications, 30, pp.605–611.

[11] Wu, M-C., Lin, S-Y., & Lin, C-H. 2006 An effective application of decision tree to stock trading. In Expert Systems with Applications, 31, pp.270–274.

[12] Huang, K.Y, and Jane, C.-J., 2009 A hybrid model for stock market forecasting and portfolio selection based on ARX, grey system and RS theories In Expert Systems with Applications, 36, pp.5387–5392.

[13] Teoh, H. J., Cheng,C-H., Chu, H-H., and Chen, J-S., 2008 Fuzzy time series model based on probabilistic approach and rough set rule induction for empirical research in stock markets. In Data & Knowledge Engineering, 67, pp.103–117.

## Author's Biography

**Sukriti Jain** is an Undergraduate researcher who has completed her Bachelors in Electronics and communication Engineering from Ambedkar Institute of Advanced Communication technologies and research (AIACTR), GGSIPU, Delhi, India. She has also completed internships with National Thermal power co-operation, Delhi and Bharat Electronics Limited , Ghaziabad. Her research interests include Financial markets, bond portfolio management and Asset management.

**Amarjot Singh** is Graduate Student and Research Assistant in Laboratory for Robotic Vision at Simon Fraser University, Burnaby. Before joining SFU, I was a research engineer in Acoustic Research Laboratory at National University of Singapore. My research is focused on Image Segmentation, Motion Tracking, Computational Photography and Financial Modeling.

**Samarth Gupta** is an undergrad student at Indian Institute of Technology, Roorkee (IITR). His main areas of interest are application & software development using Java Programming Language.