Genetic Programming Evolved Spatial Descriptor for Indian Monuments Classification

Malay S.Bhatt
Research Scholar
Rai University
Ahmedabad, India
malaybhatt202@yahoo.com

Tejas P. Patalia
Head, Dept. of Computer Engineering
V.V.P. Engineering College
Rajkot, India
pataliatejas@rediffmail.com

***Abstract*— Travel and tourism are the largest service industries in India. Every year people visit tourist places. and upload pictures of their visit on social networking sites or share via mobile device with friends and relatives. Millions of such photographs are uploaded and it is almost impossible to manually classify these pictures as per the monuments they have visited. Classification is helpful to hoteliers for development of new hotel with state of the art amenities , to travel service providers, to restaurant owners, to government agencies for security etc. The proposed system extracts Genetic programming evolved spatial descriptor and classifies the Indian monuments visited by tourists based on linear Support Vector Machine(SVM). The proposed system is divided into 3 main phases: preprocessing, genetic programming evolution and classification. The Preprocessing phase converts images into a form suitable for processing by genetic programming system using Generalized Co-Occurrence Matrix. The second phase generates best so far spatial descriptor in the form of program based on the fitness. The Fitness is calculated using SVM. Once program is obtained as output it can be utilized for classification. The proposed system is implemented in MATLAB and achieves high accuracy.**

Keywords— ***Classification, Genetic Programming, Generalized Co-Occurrence Matrix, Monument, Movie, Spatial Descriptor, Support Vector Machine***

I. INTRODUCTION

Since many years, tremendous amount of multimedia data in the form of images, audio and video have been generated due to availability of cost-effective electronic devices like camera, mobile, handy cam and shared, uploaded or emailed to relatives and friends staying away to make them feel that they have not missed the precious moments. In January 2013, the India is at 3rd position with 62.6 million facebook members [15].

Video is most popular among all as it comprises of visual cues in the form of images supported by audio information. Multimedia data is generated by folks during its business or vacation trip, to revive celebration of birthday or success party over a period of time. The press media spawns multimedia data to facilitate the news and the message they want to convey to the audience thus plays a vital role in nurturing the mindset of the society.

978-1-4673-7437-8/15/$31.00©2015 IEEE

The Bollywood and the Television industry are also considered as major sources of multimedia data (Movie stills, Posters etc.) as they release more than 2000 movies, lot of music videos, present serials, organize events for 100 years [8]. It is observed that pilot scenes and song sequences consists of the monuments in the background and thus movies promote tourism and increase the revenue of the country as a whole. Fig. 1 depicts the presence of the monument in scenes of popular movies.

Travel and tourism industry incorporates heritage, medical, business and sports tourism. The main objective of this sector is to develop and promote tourism, maintain competitiveness of India as tourist destination and improve and expand existing tourism products to ensure employment generation and economic growth [13].

The Government is also promoting tourism through advertisements, campaigns and takes special interest to preserve the beauty of these monuments. Every year lot of foreign delegates and tourists visit India too. As per India Tourism Statistics 2013, 6.97 million foreign tourists arrived with annual growth rate 5.9% and approximately 1145 million domestic tourists with annual growth rate of 9.6% were observed [14].

TABLE I. BOLLYWOOD MOVIES AND MONUMENTS

Movie Name	Release Year	Monument
Tevar	2015	Taj Mahal
Jeans	1998	Taj Mahal
Jhoom Barabar Jhoom	2007	Taj Mahal
Leader	1964	Taj Mahal
Mere Brother ki Dulhan	2011	Taj Mahal
Namastey London	2007	Taj Mahal
Youngistaan	2014	Taj Mahal
Fanaa	2006	Qutub Minar
Jannat 2	2012	Qutub Minar
Rang De Basanti	2006	India Gate
Rang De Bassanti	2006	Golden Temple
Rab Ne banadi Jodi	2008	Golden Temple

Fig. 1. (a) 'Tevar' Movie (b) 'Mere Brother Ki Dulhan' Movie (c) 'Leader' Movie (d) 'Rang De Basanti' Movie (e) 'Fanaa' Movie (f) 'Jannat 2' Movie (g) 'Rab Ne Bana Di Jodi' Movie and (h) 'Rang De Basanti' Movie.

II. LITERATURE SURVEY

The literature survey covers applications of Genetic Programming in the fields of image processing and computer vision.

Liu et al. [1] have focused on human-action recognition which is a potential application of computer vision and evolved Spatial-Temporal descriptor using genetic programming. The KTH and Weizmann datasets are combined; each action is normalized to 100x100x70 block and taken as input. Output of Genetic Programming Tree (GP Tree) is also a 100x100x70 block. Output is divided into 10x10x50 sub-blocks and mean value of each sub-block is concatenated into 500D vector and taken as input to linear support vector machine. Ten-fold cross validation approach is selected to evaluate the fitness of the GP Tree. The Ramped Half-and Half method is applied to generate Initial population of size 100. Crossover and Mutation genetic operators are applied with 0.90 and 0.10 probabilities respectively. MATLAB along with GPlab is used for implementation and accuracy achieved is 96.9%

Poli [2] investigated applicability of Genetic Programming for segmentation of the brain images and achieved optimal filters.

Streater, J [3] explains the design and implementation of genetic programming system for construction of feature descriptor for skin lesion images. 6 Generalized Co-occurrence Matrices (GCMs) of size 64x64 are constructed using RGB color space with 5 inter-pixel distances for 100 images. Highest accuracy achieved is72%. Fisher's Discriminant Ratio, naïve bayes classifier are also used to compare the results obtained by SVM.

Liu et al. [3] have constructed Spatio-Temporal descriptor for Hand gesture recognition task. Experiments are conducted on the Cambridge Hand Gesture benchmark ataset which consists of 900 video clips of 9 hand gesture classes and Northwestern Hand Gesture dataset. The two different tiers are used to achieve the goal. Max-pooling tier constructs max-pooling filters and Filtering tier consists of 3D Gabor, Gaussian, Laplacian and Wavelet filters and basic arithmetic operations. The accuracy achieved is 85%. Results are compared with other hand crafted 3D descriptors like HMHI, Histogram of Oriented Gradient (HOG)/ Histogram of Optical Flow (HOF), HOG and Scale Invariant Feature Transform (SIFT).

Singh et al. [4] used genetic programming approach for the purpose of image segmentation for bio-medical object detection. They implemented the system in MATLAB and compared the result with GENIE Pro. Function set consists of Morphological operators, Filters, Arithmetic operators, Histogram Equalization for segmentation and thresholding is used for post-processing. Segmentation accuracy achieved is 98% which is approximately 4% higher than GENIE Pro.

Extraction of information from movies is focused since last 10 years. Hollywood Movie Database 51[10], YouTube[11] and the Hollywood datasets[12] are challenging datasets used widely for action recognition. Action recognition from movies is a subset of general human recognition activity.

Laptev et al. [5] highlighted limitations of human action dataset in controlled environment and also described the difficulties faced during the recognition of real movie actions. They identified similarities and dissimilarities between action recognition from movies with object recognition in still images. The first task accomplished with the accuracy of 60% is the automatic annotation of the human action using the movie scripts. Inaccuracy is due to script video misalignment. Classification of human-action is the main goal achieved in the paper with 91.8% accuracy. Experiments are carried out for 8 different actions. HoG, HoF, Spatio-Temporal Bag of Features (BoF) and combination of above are used with non-linear support vector machine to achieve the desired result.

Lei Chen et al. [6] depict a top-down approach based on rules for video editing and audio cues to extract dialogue and action scenes. Finite state machine with an audio-based support vector machine (SVM) classifier is applied for detection of scene type. Classifier uses three features namely: variance of zero crossing rate, silence ratio, and harmonic ratio. The precision and recall rates achieved are 76.56% and 81.6% respectively.

Doudpota et al.[7] focused on impact and popularity of Bollywood movies in South Asia, Middle East, UK, USA and other parts of the world. They mined song sequences from the Bollywood movies. They used in the first part, Zero Crossing Rate, Spectrum Flux and Short Time Energy as features in Support Vector machine for binary classification of extracted segment into music and non-music. In the second phase, extracted music segments are further classified into song and non-song sequences using Probabilistic Timed Automata (Song Grammar). Experiment was carried out on 10 Bollywood movies having 74 songs and out of which 69 were successfully extracted. Recall achieved is 93.24% while the Precision is 87.34%.

III. METHODOLOGY

A domain-independent machine learning methodology to automatically generate high-level spatial descriptor for recognition of Indian monuments using Genetic Programming(GP) is proposed. A group of 2 dimensional operators are clubbed together to generate problem-oriented descriptor which is capable of selectively extracting features from the set of given images. The evolved descriptor can both extract meaningful features and generate compact representation. We learn our proposed system over a training set, in which spatial descriptors are evolved by maximizing the recognition accuracy through a support vector machine as fitness function, and further evaluate the GP-selected one over a testing set to demonstrate the performance of our method.

Generally, tree structure is adopted to represent genetic programs. GP evolves by selection, crossover or mutation through sexual reproduction with pairs of parents being chosen stochastically but biased in their fitness on the given task and finally select the best performing individual as the terminal solution. Basic framework of GP system is shown in Fig. 2.

Start

Create initial population of operation sequences from Terminal and Function Set.

Repeat

1. Evaluate the fitness of the individual.
2. Select individuals with a particular probability biased in their fitness.
3. Create new generation of individuals through crossover and mutation operation

While(Ach_fitness < Req_fitness **or** Max. generation)

Stop

Fig. 2. Genetic Programming Algorithm

A. Preprocessing

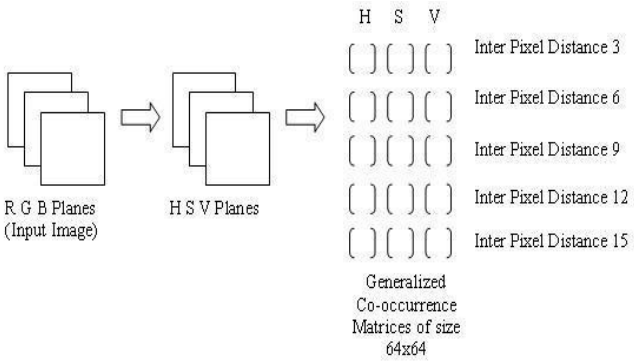
A set of 100 input images is re-sized into 100x100 and blurred by 'Gaussian Filter' with a 15x15 mask and standard deviation '1'. Blurring is essential to eliminate minor intensity variations in the neighborhood. Resulted blurred images are converted from RGB color space into HSV color space.

Fig.3 (a) depicts the whole preprocessing activity. Fig.3 (b) covers sample blurred image while Fig.3(c) shows obtained HSV image after conversion. It is known that HSV color space has more discriminative power as compared to RGB color space.

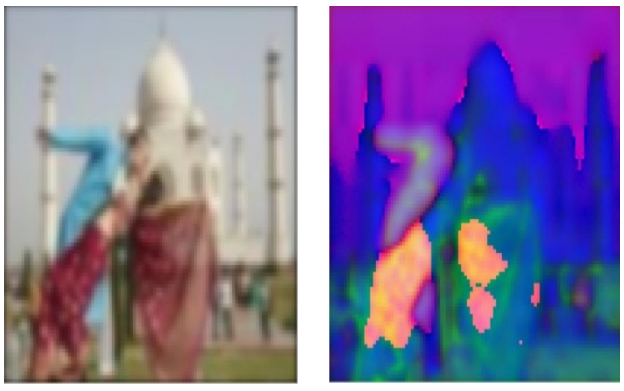
B. GPlab Framework

All things cited by Koza, creator of genetic programming, as characteristics defining suitable genetic programming languages are included in the GPlab1. The basic infrastructure of the Genetic Programming algorithm implemented here is provided by GPlab, a free and open source Genetic Programming toolbox for MATLAB [5]. Although GPlab has been extremely useful, providing the nuts and bolts of population representation, creation, evolution, and visualization, it is also ultimately very inadequate for the purpose of image processing task.

It was necessary to heavily modify and extend GPlab so that, among other supportive changes, GPlab could load, manipulate, and evaluate large matrices as terminals in feature equations



(a)



(b) (c)

Fig. 3. (a) Preprocessing (b) Blurr image and (c) HSV Image

C. Terminal Set

1	1	2	3	4
2	3	4	4	1
3	4	1	1	2
4	1	2	2	3
1	2	3	3	4

(a)

	1	2	3	4
1	2	4	0	0
2	0	1	3	0
3	0	0	1	4
4	3	0	0	1

(b)

Fig. 4. (a) Sample Image (b) GCM in horizontal Direction with inter- pixel distance '1'

The Generalized Co-occurrence Matrices (GCM) of size 64x64 are calculated for inter-pixel distances 3, 6, 9, 12 and 15 in horizontal direction for H, S and V planes. Total of 15 (3 planes x 5 inter- pixel distance) GCM matrices each of size 64x64 forms the terminal set of our GP System. Fig.4(a) describes the working of GCM for a sample image of size 5x5 with 4 gray levels while Fig. 4(b) contains calculated GCM in horizontal direction with inter-pixel distance '1'.

D. Function Set

Nature of the problem at hand decides the function set (internal nodes of the tree) of the GP.

TABLE II. FUNCTION SET

Wrapper function created in MATLAB	Inputs	Function Description
Myadd( )	2	Addition of 2 GCMs of size 64x64
Mysub()	2	Subtraction of 2 GCMs of size 64x64
Mymult()	2	Element by element multiplication of 2 GCMs of size 64x64
Mydiv()	2	Element by element division of 2 GCMs of size 64x64.
Mysqure()	1	Square of individual elements of GCM of size 64x64
Mysine()	1	Sine( ) of individual elements of GCM of size 64x64
Mycosine()	1	Cosine( ) of individual elements of GCM of size 64x64

The choice of functions is based on the following principles[1]:

1) Functions can extract meaningful information.

2) To minimize the execution of the GP, all the operators in the function set must be simple and efficient.[1]

3) To ensure operator closure [9], we have only used functions which map one or two 64 × 64 2D blocks to a single 64 × 64 block. In this way, a GP tree can be an unrestricted composition of function nodes. Always semantically legal tree will be produced in this way.

E. Fitness Function

The most important part of genetic programming is the fitness function which determines how well a program is able to solve the problem. To evaluate the candidate GP evolved descriptors, we here adopt the classification accuracy calculated by a linear SVM classifier on the training set as the fitness function. Our GP architecture, for any of the input sequences, we can obtain an output sequence with an identical size as the input due to the enclosure property.

We further convert output obtained by GP Evaluation into 4096D vector as the input to the linear SVM. To obtain a more reliable fitness evaluation, we adopt ten-fold cross-validation for each new GP tree using SVM. We divide the GP training set randomly into 10 equal-sized parts and perform ten repetitions of training the SVM on 9/10 of the set and testing on the remaining 1/10. The overall fitness 'Er' is the average of the ten-fold cross-validation accuracy. Fitness function is defined as follows:

$$Er = (1 - (\sum_{i=1}^{n} (SVM[accuracy(i)])/n)))*100\ \% \qquad (1)$$

In our case, value of n is 10. accuracy(i) represents accuracy of fold i by the SVM.

IV. EXPERIMENTS AND RESULTS

Our goal is to classify above mentioned monuments from large repository of photographs uploaded on the social networking websites.

A. Implementation

We evaluate our proposed method using 64-bit MATLAB 2013a (with the genetic programming toolbox GPLAB[1]), 8GB of RAM running on the Windows 8.1 OS with i7 5th generation processor. For Genetic Programming evolution a Tournament selection method has been applied in the implementation. Initial Tree depth is considered as 8 and maximum permitted tree depth is considered as 10. We use a population size of 75 individuals with the initial population generated with ramped half-and-half method. During GP run, we fix tree crossover probability to 50% and mutation

[1] http://gplab.sourceforge.net/download.html, A Genetic Programming Toolbox for MATLAB

probability to 50 %. We utilize the 'keepbest' scheme for GP. Maximum number of generation (100 generation) considered as stopping criteria or error rate should be less than or equal to 5%.

B. Dataset

Most important task is the collection of data as no direct dataset is available for the task at hand. In our data set 4 different classes are considered namely: 'Taj Mahal', 'Golden Temple', 'Qutub Minar', and 'India Gate'.

For each mentioned class, 25 images are collected from various websites and a total of 100 images are collected. An individual tourist, family and tourist groups are considered in various poses for both training and testing. To achieve diversity, for each class front view (day), front view(night), Far View, Left-View(45 degree), right-side view(45 degree), People posing (sitting or standing etc.) in front of monuments are considered. Some example images are shown in Fig. 6.

C. Gp Tree

Best GP Individual in the form of LISP after 100 generation is given below and tree representation of the same is shown in Fig.7. Fitness Evaluation graph shows improvement in fitness over 100 generations in Fig. 5.

mydiv(mycos(myadd(myadd(mysquare(X13),myadd(mydiv(mydiv(X5,X7),X4),X12)),X12),myadd(myadd(X2,mydiv(X1,X7)),myadd(myadd(mysquare(mysine(mymult(myadd(X6,mysquare(X8)),mysquare(mycos(mysquare(X13)))))),myadd(X10,X9)),X7))) (2)

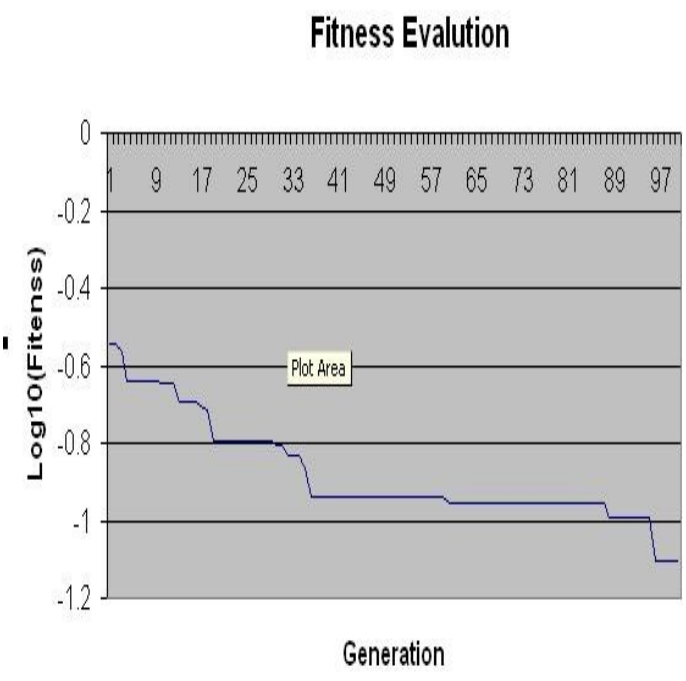


Fig. 5. Fitness Evaluation Graph

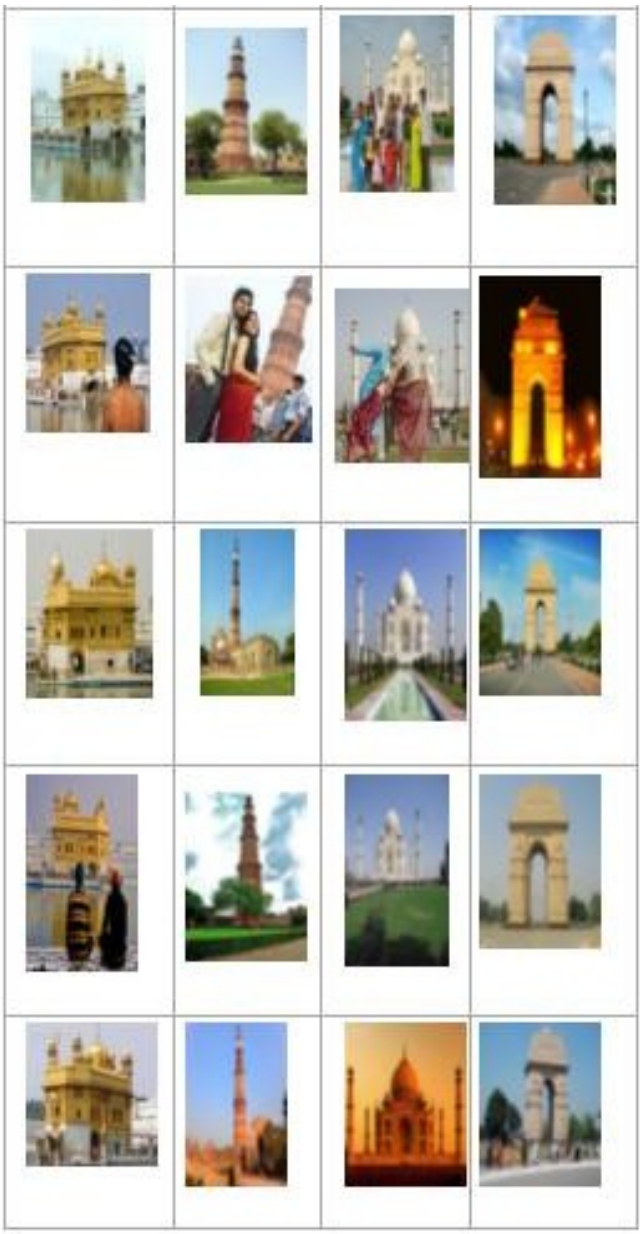


Fig. 6. Images From Sample Database (a)Golden Temple (b) Qutub Minar (c) Taj Mahal and (4) India Gate

V. CONCLUSION

The Indian tourism industry plays a vital role in the economic growth of the nation. Presence of monuments in the bollywood movies and its impact on tourism industry are highlighted as tourists prefer such places for visit. Genetic Programming evolved spatial descriptor using 15 Generalized Co-Occurrence matrices as terminals is implemented with 7 functions and 2 operators. The system takes around 7 Hrs for execution to reach to 100^{th} generation. Evolution of spatial descriptor is one time cost. Once descriptor is obtained in terms of program it can be readily used to create a descriptor for test images and classification is possible within minutes. We have achieved the accuracy of 92.75%.

VI. FUTUREWORK

The proposed system can be further extended by adding other international and national monuments. System can also be used with minor modification for classification of desired entity from large online image repository. We have used GCMs as input to GP system. Other popular feature descriptors like Harris, HoG, SIFT, Local Binary Pattern (LBP) can also be used as input to GP system.

REFERENCES

[1] Liu, L., Shao, L., & Rockett, P., "Genetic Programming-Evolved Spatio-Temporal Descriptor for Human Action Recognition," *BMVC,* pp.1-12, 2012.

[2] Poli, R, "Genetic programming for image analysis," *In Proceedings of the 1st annual conference on genetic programming*, pp. 363-368. MIT Press, 1996.

[3] Liu, L., and Shao, L, "Synthesis of spatio-temporal descriptors for dynamic hand gesture recognition using genetic programming," *2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)*, pp.1-7, 2013.

[4] Streater, J, "Genetic Programming for the Automatic Construction of Features in Skin-Lesion Image Classification," *University of Edinburgh*, 2010.

[5] Laptev, I., Marszalek, M., Schmid, C., and Rozenfeld, B, "Learning realistic human actions from movies," *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, pp. 1-8, 2008.

[6] Chen, L., Rizvi, S. J., & Ozsu, M. T, "Incorporating audio cues into dialog and action scene extraction,". *Electronic Imaging*, pp. 252-263, International Society for Optics and Photonics, 2003.

[7] Doudpota, S. M., & Guha, S, "Mining movies to extract song sequences," *Proceedings of the Eleventh International Workshop on Multimedia Data Mining*, p. 2, 2011 .

[8] B. Bhavan, "Central Board Of Film Certification: Annual Report,". *Ministry of Information and Broadcasting, Government of India*, 91-E, Walkeshwar Road, Mumbai, India, 2009.

[9] R. Poli, W.B. Langdon, and N.F.McPhee, "A field guide to genetic programming," *Lulu.com,* 2008

[10] Kuehne, H., Jhuang, H., Garrote, E., Poggio, T., & Serre, T, "HMDB: a large video database for human motion recognition," *2011 IEEE International Conference on Computer Vision (ICCV)*, pp. 2556-2563, 2011.

[11] Liu, J., Luo, J., and Shah, M, "Recognizing realistic actions from videos "in the wild," *IEEE Conference on Computer Vision and Pattern Recognition*, CVPR, pp. 1996-2003, 2009.

[12] Marszalek, M., Laptev, I., and Schmid, C, "Actions in context," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2929-2936, 2009.

[13] www.india.gov.in/topics/travel-tourism

[14] www.tourism.gov.in/writereaddata/CMSPagePicture/file/marketresearch/publications/IndiatourismStatistics2013_1.pdf

[15] www.socialbakers.com/facebook-statistics

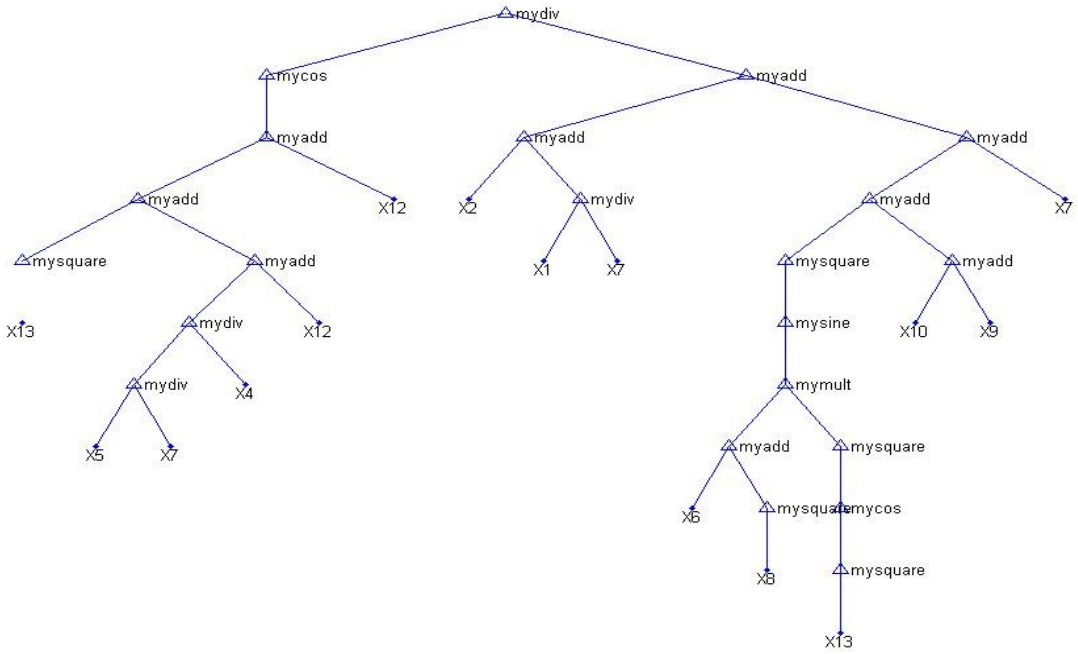


Fig. 7. Evolved GP Tree after 100 generations