Deep Learning for Real-Time Vehicle Counting and Categorization in Traffic Videos

AyushKumar Ruplal Burde, Abhinav Vijay Patii, Durvesh Satish Patil,
Name of guide:- T.P.Raju
Tulsiramji gaikwad patil college of engineering, Nagpur, India
ayushburde156@gmail.com , abhinavpatil720@gmail.com , durveshpatil28111@gmail.com

---------------------***--***-------------------------------------***-------------------
---

Abstract **-** *Traffic analysis is a problem that megacity itineraries have been dealing with for times. Smarter styles are developed to anatomize business and speed up the process. Business analysis can record the number of vehicles and vehicle classes in an area at a given time. People have been developing similar mechanisms for decades now, but utmost of them involve using detectors to calculate the direction of moving vehicles and identify vehicles to track vehicle figures. Although this system has progressed over time and is truly effective, they're n't budget-friendly. The problem is that similar systems bear periodic conservation and estimation. thus, this design aims to calculate and classify the vehicle rested on vision. The system involves wharf frames from videotape to descry and count vehicles using Gaussian Mixture Model (GMM) background deduction, also classifying vehicles by comparing figure areas with prognosticated values. A significant donation of the paper is the comparison of two type styles. type is done using Contour Comparison(CC) and Bag of Features(BoF) styles.*

Keywords: - Vehicle counting, Business analysis, Contour Comparison.

1. INTRODUCTION

Moment, countries and governments need a safe and affordable system to automate vehicles and control vehicle theft. adding business on roads and highways, adding business, and problems with being vehicle detectors have led to the development of new vehicle discovery technologies. Computer vision systems are the most common choice, but several problems must be overcome to successfully perform type. Real- time discovery and shadowing of objects or vehicles moving on different roads by intelligent vision systems is important for multitudinous fields of disquisition and technology operations. lodging useful information analogous as business density, object speed, automobilism behaviour and business patterns from these camera systems becomes essential. Manual analysis is no more. Developing intelligent systems that can prize business business and vehicle type data from business control systems is essential for business operation. differently, the monitoring system is also important in automobilism backing operations, as the vision system allows the discovery and type of the vehicles involved in the mugged incident. A figure is a visual representation of commodity.

The time period has several uses in information technology. A print is an snap that's created or copied and saved electronically. An picture may be interpreted as a vector visible or a raster visible. Digital picture processing involves the manipulation of virtual prints the operation of a virtual computer. This is part of the signal and machine, still pay unique attention to the print. DIP makes a thing of developing computer systems that could carry out print processing. The input of the device is a virtual image and the system tactics the image using an important set of rules with the snap as affair. It makes it possible to apply in addition huge algorithms to the enter snap and might keep down from problems similar as noise and signal deformation in the course of processing. Pix may be divided into the posterior three types. A double picture consists of pixels that may be considered one of colours, generally black and white. Double snap shots are called situations or two situations. This way that each pixel is stored as an integer, ie 0 or 1. Gray is an intermediate colour among white and white. It's a unprejudiced shade or achromatic colour, which actually means" white" color because it's suitable to be composed of white and white. It came the achromatism of cloudy skies, dust and lead. A achromatism (virtual) print is a virtual picture that includes colour information for every pixel. This fashion is environmentally affable as it does n't suffer chemical processing. Digital imaging is constantly used to validate and report literal, clinical, and specific events This paper describes a imaginative and visionary- rested device for detecting, covering and classifying shifting motorcars. Four different implicit enterprise pots can be described, but the proposed software program is lissom and the wide variety of agencies

2. LITERATURE SURVEY

Alpatov et al. The hassle of avenue circumstance analysis for enterprise manage and safety is considered. The following photograph processing algorithms are proposed vehicle discovery and counting algorithms, street sign discovery algorithms. Algorithms are designed to reuse pictures taken from desk bound cameras. The superior vehicle discovery and counting algorithms also are tested on bedded smart camera structures. Singing and so forth. Proposed a imaginative and prescient- grounded car identification gadget and counting device. This study posted a highlevel trace dataset with a aggregate of 57,290 reflections on a mixture of eleven,129 photos. Compared to being public databases, the proposed database consists of small gadgets with reflections to offer a entire database for automobile discovery grounded on deep literacy. Neupane et al. Created a education database of about

30,000 samples from seven being car digicam classes. To spoil P2, exceptional- tuning is grounded on this skilled and applied transfer workout in a ultramodern YOLO(You Only Look previously) network. For P3, this paintings proposes amultivehicle shadowing set of rules that snappily calculates every path, classifies, and obtains the vehicle haste. Lin et al. It introduces a business tracking device grounded on igital discovery zones, Gaussian admixture model(GMM) and YOLO to ameliorate automobile counting and bracket effectiveness. GMM and virtual discovery bands are used to rely cars and YOLO is used to classify vehicles. Additionally, car distance and time are used to estimate automobile velocity. In this take a look at, Montevideo Audio and Video Data(MAVD), GARM Road- Traffic Monitoring Dataset(GRAM- RTM) and our series dataset are used to check the proposed gadget. Chauhan et al. It makes use of a complicated Convolutional Neural Network(CNN) item This work evaluates the quiescence, energy and tackle costs of planting exploration grounded on our CNN model, as the growing region lacks strong network connectivity for nonstop videotape streaming from the road to the pall garçon.

Arinaldi et al. presented a business videotape analysis system grounded on computer vision ways. The system is designed to automatically collect important statistical information for policy makers and controllers. These statistics include vehicle counts, vehicle type groups, and vehicle speed estimates from videotape and vehicle operation monitoring. System discovery like and vehicle bracket in vehicle videotape. This work carried out two models for this purpose, the first is MoG SVM network and the second is grounded on Faster RCNN, a new popular deep literacy armature for object discovery in images. Goma et al. presented an effective real-time approach for detecting and counting moving vehicles grounded on YOLOv2 and caching point stir analysis. This work is grounded on the discovery of coetaneous vehicle features and tracking to achieve accurate counting results. The proposed strategy works in two phases; the first is to identify vehicles and the second is to count moving vehicles. For original object discovery, this work uses the fastest literacy object discovery algorithm YOLOv2 before filtering K- groups and KLT shamus . An effective approach is also introduced using the temporal information of theinter-frame discovery and shadowing features to marker and directly calculate each vehicle with its separate line. Oltean et al. It proposes a real- time vehicle counting approach using a small YOLO for shadowing. This program works on Ubuntu with GPU processing and the coming step is to test it on low budget bias like the Jetson Nano. Test results show that this approach achieves high delicacy in real- time speed(33.5 FPS) in real business videotape.

Pico et al. proposed to apply a low- cost system for vehicle identification and bracket using an ARM- grounded platform(ODROID XU- 4) installed with the Ubuntu operating system. The algorithm used is grounded on an open source library(Intel OpenCV) and is enforced in the Python programming language. trials prove that the effectiveness of the enforced algorithm is 95.35, but it can be bettered by adding the training samples. Tituana et al. review colorful former workshop developed in this field and identify the technological styles and tools used in those workshop; In addition, this study also highlights trends in this field. The most applicable papers are reviewed and the results are epitomized in tables and numbers.

3. PROPOSED SYSTEM

This system can be used to descry, fete and track vehicles in videotape images, and also classify the detected vehicles into three different classes grounded on their size. The proposed system is grounded on three modules, videlicet background literacy, focus birth, and vehicle bracket as described in background deduction, a classic approach to capture background images or in other words, moving object discovery.

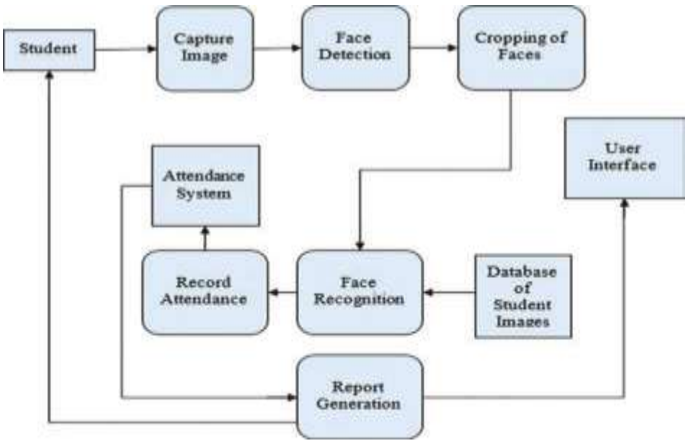


Fig.1: Block diagram of proposed system.

Vehicle Detection and Counting

The third and final module in the proposed system is bracket. After applying the pre-extraction module, the corresponding silhouettes are attained, these figure features are like centroids. Aspect rate, area, size and stiffness are uprooted and used for vehicle bracket. This module consists of three way, videlicet background reduction, image improvement and focus birth. The background is removed to reveal the focus. This is generally done by setting the static pixels of the stationary object to double 0. After background deduction, noise filtering, dilation, and corrosion are used to gain the correct silhouettes of the background object. The affair of this module is first class. Area of interest In the first frame of the videotape, I draw a line near the image and define the ROI. The thing is to fete ROI in after frames, but fete that ROI is n't the primary vehicle. This is only part of the vehicle and can be misshaped, rotate, restate, or indeed fully vanish from the frame. Vehicle discovery A visionary strategy for opting a hunt window for vehicle discovery using image environment, a GMM frame is proposed for vehicle discovery with successional movement with top-down attention. constantly achieved satisfactory performance in relating vehicles clear link box. proposed a methodical hunt strategy for detecting visual vehicles in the image, where the discovery model proposed a deep RL frame to elect the applicable action to capture the vehicle in the image. Vehicle Count This module counts detected vehicles and the result of this count will be streamlined constantly grounded on vehicle discovery, the result will be affair for streaming videotape using OpenCV.

Background Learning Module

The first module in this system is to learn how the background differs from the focus. Also, since the proposed system works on the videotape feed, this module excerpts frames from it and learns about the background. In a business scene captured by a stationary camera mounted on the roadside, moving objects can be considered focus and stationary objects can be considered background. An image processing algorithm is used to learn about the background using the system described over.

Gaussian Mixture Modelling (GMM)

In its simplest form, GMM is a type of clustering algorithm. As the name suggests, each group is modeled according to a different Gaussian distribution. This flexible and probabilistic approach to data modeling means that we've soft assignments rather of hard assignments to groups similar as k- order. That is, each data point can be generated by one of the distributions and associated chances. In fact, each distribution has a specific" responsibility" for generating specific data points.
How can we estimate the appearance of this model? Well, one thing we can do is introduce retired variables Data(gamma) for each data point. This assumes that each data point is generated using some information about the idle variable. In other words, it tells you that the Gaussian generated a certain data point. But in practice we do n't see these retired variables, so we've to estimate them. How to do this? Well, luckily for us, we formerly have an algorithm that works in this kind of situation, the Expectation Maximization(EM) algorithm, and we'll talk about it next.**The EM algorithm:-** The EM algorithm consists of two way, theE-step or Anticipation step and the M- step or Maximization step. Suppose we've some idle variables(unobserved and defined by the vector Z below) and our data point X. Our thing is to maximize the borderline probability of X given our parameters(defined by the vector). In fact, we can find the borderline distribution as the union of X and Z and find the sum of all Z's(rule of probability).

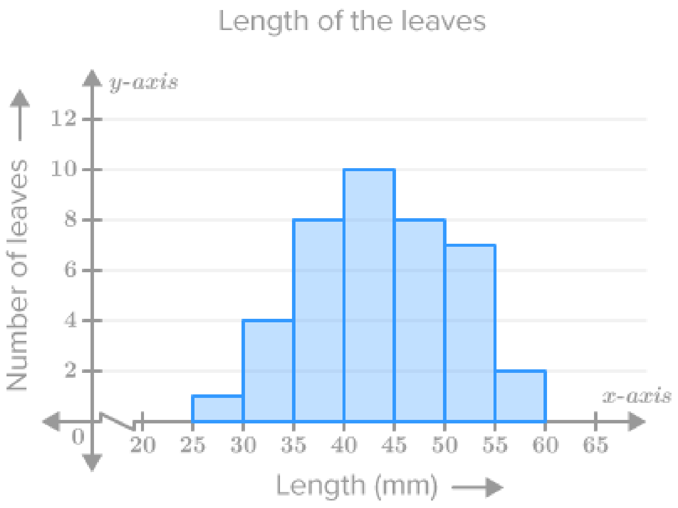
$$In\ p(X|\Theta) = I\eta\{\ \Sigma = (X.Z|\Theta)\}$$

The below equations frequently produce complex functions that are delicate to gauge . What we can do in this case is to use Jensens Inequality to construct a lower set function that's easier to optimize.However, we can compare the original function, If we optimize this by minimizing the KL difference(gap) between the two distributions. This process is described. I've also handed a link to a videotape showing the derivate of the KL divergence for those who want a more rigorous fine explanation.
In fact, we only need to do two way to estimate our model. In the first step(E-step), we want to estimate the posterior distribution of our idle variable in terms of tentative weight(π), our term(μ) and the Gaussian mean covariance(Σ). also we can enter the alternate step(M- step) and use it Increase the liability associated with our parameter parameter θ. This process is repeated until the algorithm converges(the loss function remains unchanged).

Bag of Features Model

Model Visual Features (BOF) is one of the most important generalities in computer vision. We use a visual vocabulary model to classify image content. It's used to make a high volume shadowing system (non-specific, precise). When we classify textures using textures, we rather use a visual vocabulary model. As the name suggests, the conception of" visual bag of words" is actually deduced from the" bag of words" model used in information reclamation (eg, textbook- grounded hunt machines) and textbook analysis. The general idea in Word Bag is to present a" document"(ie a web runner, a Word train, etc.) as a collection of important keywords, fully ignoring the order in which the words appear. Documents that partake the same keyword are considered to belong to each other anyhow of the order of the keywords. Also, because we fully ignore the order of words in a document, we call this representation" bag of words" rather of" list of words" or" list of words" Treating a document as a" bag of words" allows us to efficiently dissect and compare documents. because we do not need to store information about the order or position of words- we count how numerous times a word appears in a document, and also use the frequence number for each word as a way to rate the document. In computer vision, we can use the same conception- only now rather of working with keywords, our" words" are now layers of images and related point vectors.

Given a wordbook of possible visual words, we can also count the number of times each visual word appears and fantasize it as a histogram. This histogram is a veritable bag of visual words. structure visual vocabulary can be divided into three way.



Step# 1: point birth

The first step in erecting a visual bag of words is to prize descriptors and features from each image in our database. point birth can be done in several ways identify crucial points and excerpt SIFT features from crucial regions of our image; using circles in regular intervals(for illustration, solid button sensors) and derivations of other forms of original steady descriptors; or we can prize the average RGB value from a arbitrary image region. The point then's that for each input image we get several point vectors

Step# 2: Dictionary/ Vocabulary Construction

After rooting point vectors from each image in our database, we need to make a vocabulary of possible visual words. Word conformation is generally done through a k- means clustering algorithm that summarizes the point vectors attained from step 1. The centers of the performing clusters(eg, centroids) are considered as visual word wordbooks.

Step#3: Vector Quantization

Given an arbitrary image (whether from our original database or not), we can identify and abstract the image using a bag of image words using this process prize the point vector as in step 1 over. For each uprooted point vector, count its nearest neighbors in the wordbook created in step 2- this is generally done using Euclidean distance. Take the set of nearest neighbor markers and construct a histogram of length k (the number of clusters formed by kverbs), where the i value in the histogram is the frequence of the i-visual word. When modeling an object by distributing prototype vectors, this process is generally called vector quantization.

Classification

One of the intriguing features of our network is its simplicity the classifier is only replaced by a masking subcaste, without any previous or convolutional structure. still, it should be prepared with a large quantum of training data vehicles of different sizes should appear nearly far and wide.

Visual shadowing solves the problem of chancing a target in a new frame from the current position. The proposed shamus stoutly tracks the target with sequence movement controlled by GMM. GMM predicts the stir to run the target from its position in the former frame. The crossroad box moves with the movement prognosticated from the former state, and the coming movement continues to be prognosticated from the moved state. We break the vehicle shadowing problem by repeating this process in a series of tests. GMM excels in RL as well as SL. Online adaption takes place during real shadowing.

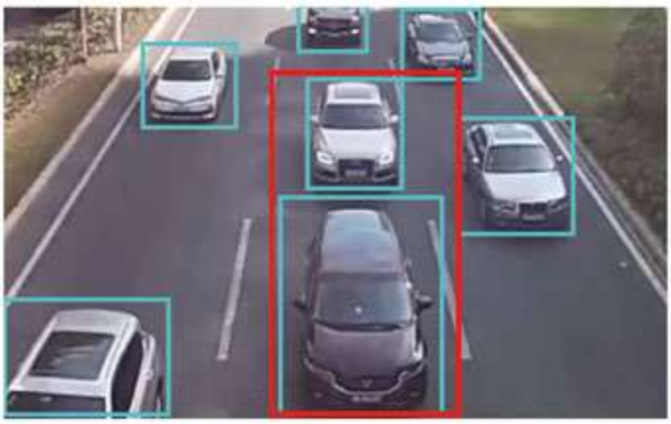
A GMM is designed to induce stir to find the position and size of the target vehicle in a new frame. The GMM algorithm learns a policy that selects the most optimal action to follow from the current situation. In GMM, a policy system is developed in which the input is a abbreviated image subcaste in the former state and the affair is the probability distribution of conduct similar as restatement and scale change. The process of choosing this course of action requires a bit of exploration step more than the sliding window or seeker slice approach. likewise, since our system can localize the target by opting the stir, post-processing similar as box retrogression is not needed.

Advantages of Proposed System

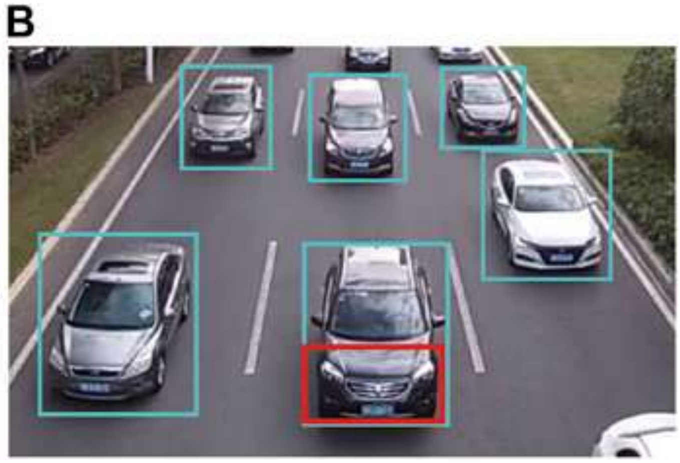
- Identify high- moving vehicles in videotape sequences.
- Vehicle shadowing is uncorked.
- Identify the type of vehicle.
- Calculate the quantum of business through videotape.

4. RESULTS AND DISCUSSION

A



International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 04 | April - 2025 SJIF Rating: 8.586 ISSN: 2582-3930



5. CONCLUSION AND FUTURE SCOPE

5.1 Conclusion

The proposed solution is implemented in python using OpenCV bindings. Camera images from multiple sources are processed. A simple interface was developed for users to select regions of interest for analysis, and image processing techniques were used to count the number of vehicles and classify vehicles using machine learning algorithms. From the experiment, it can be seen that the CC method outperforms the BoF and SVM methods in all results and provides classification results closer to the ground truth value.

5.2 Future Scope

One of the limitations of the system is that it is not effective in detecting vehicle obstacles, which affects the accuracy of counting and classification. This problem can be solved by introducing secondary feature classification, such as color based classification. Another limitation of the current system is the need for human supervision to determine areas of interest. To calculate the vehicle, the user must determine an imaginary line that intersects the center of the contour, so the accuracy depends on the judgment of the human observer. In addition, the camera angle affects the system, so camera calibration techniques can be used to determine the path to see the road better and improve efficiency. This system is not able to detect vehicles at night because it requires the visibility of objects in front for the extraction of contour features, as well as classification features using SIFT features. The system can also be optimized for greater accuracy using more sophisticated image segmentation and artificial intelligence processes.

REFERENCES

[1] **Chauhan, M. S.**, **Singh, A.**, **Khemka, M.**, **Prateek, A.**, & **Sen, Rijurekha**. (2019). Bedded CNN-grounded vehicle bracket and counting in non-laned road business. *Proceedings of the Tenth International Conference on Information and Communication Technologies and Development (ICTD '19)*, Article 5, 1–11. https://doi.org/10.1145/3287098.3287118.

[2] **Butt, M. A.**, **Khattak, A. M.**, **Shafique, S.**, **Hayat, B.**, **Abid, S.**, **Kim, Ki-Il**, **Ayub, M. W.**, **Sajid, A.**, & **Adnan, A.** (2021). Convolutional neural network grounded vehicle bracket in adverse luminous conditions for intelligent transportation systems. *Complexity*, Article ID 6644861, 11 pages. https://doi.org/10.1155/2021/6644861.

[3] **Neupane, Bipul**, et al. (2022). Real-time vehicle bracket and tracking using a transfer literacy-bettered deep learning network. *Sensors (Basel, Switzerland)*, 22(10), 3813. https://doi.org/10.3390/s22103813.

[4] **Gomaa, A.**, **Minematsu, T.**, **Abdelwahab, M. M.**, et al. (2022). Faster CNN-grounded vehicle discovery and counting strategy for fixed camera scenes. *Multimedia Tools and Applications*, 81, 25443–25471. https://doi.org/10.1007/s11042-022-12370-9.

[5] **Jahan, N.**, **Islam, S.**, & **Foysal, M. F. A.** (2020). Re al-time vehicle bracket using CNN. *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 1–6. https://doi.org/10.1109/ICCCNT.49239.2020.9225623.

[6] **Lin, C. J.**, **Jeng, Shiou-Yun**, & **Lioa, Hong-Wei**. (2021). A real-time vehicle counting, speed estimation, and bracket system grounded on virtual discovery zone and YOLO. *Mathematical Problems in Engineering*, Article ID 1577614, 10 pages. https://doi.org/10.1155/2021/1577614.

[7] **Oltean, G.**, **Florea, C.**, **Orghidan, R.**, & **Oltean, V.** (2019). Towards real-time vehicle counting using YOLO-bitsy and fast

stir estimation. *2019 IEEE 25th International Symposium for Design and Technology in Electronic Packaging (SIITME)*, 240–243. https://doi.org/10.1109/SIITME.47687.2019.8990708.

[8] **Pico, L. C.**, & **Benítez, D. S.** (2018). A low-cost real-time bedded vehicle counting and bracket system for traffic management applications. *2018 IEEE Colombian Conference on Communications and Computing (COLCOM)*, 1–6. https://doi.org/10.1109/ColComCon.2018.8466734.

[9] **Balid, W.**, **Tafish, H.**, & **Refai, H. H.** (2018). Intelligent vehicle counting and bracket detector for real-time business surveillance. *IEEE Transactions on Intelligent Transportation Systems*, 19(6), 1784–1794. https://doi.org/10.1109/TITS.2017.2741507.

[10] **Alpatov, Boris**, **Babayan, Pavel**, & **Ershov, Maksim**. (2018). Vehicle discovery and counting system for real-time business surveillance. *1-4*. https://doi.org/10.1109/MECO.2018.8406017.

[11] **Tituana, D. E. V.**, **Yoo, S. G.**, & **Andrade, R. O.** (2022). Vehicle counting using computer vision: A survey. *2022 IEEE International Conference for Confluence in Technology (I2CT)*, 1–8. https://doi.org/10.1109/I2CT.54291.2022.9824432.

[12] **Song, H.**, **Liang, H.**, & **Li, H.**, et al. (2019). Vision-grounded vehicle discovery and counting system using deep literacy in trace scenes. *European Transport Research Review*, 11, 51. https://doi.org/10.1186/s12544-019-0390-4.

[13] **Arinaldi, A.**, **Pradana, J. A.**, & **Gurusinga, A. A.** (2018). Discovery and bracket of vehicles for business videotape analytics. *Procedia Computer Science*, 144, 259–268. https://doi.org/10.1016/j.procs.2018.10.527.

[14] **Khan, A.**, **Sabeenian, R. S.**, **Janani, A. S.**, & **Akash, P.** (2022). Vehicle bracket and counting from surveillance camera using computer vision. In **Suma, V.**, **Baig, Z.**, **Shanmugam, S. K.**, & **Lorenz, P.** (Eds.), *Inventive Systems and Control*. *Lecture Notes in Networks and Systems*, vol. 436. Springer, Singapore. https://doi.org/10.1007/978-981-191012-8_31.

[15] **Gonzalez, P.**, & **Nuño-Maganda, M. A.** (2014). Computer vision-grounded real-time vehicle shadowing and bracket system. *Midwest Symposium on Circuits and Systems*, 679–682. https://doi.org/10.1109/MWSCAS.2014.6908506.