Performance Evaluation of a Low Cost Vision-Based Traffic Flow Monitoring System

Rose Nakibuule
Makerere University
Kampala, Uganda
rnakibuule@cis.mak.ac.ug

John Quinn
Makerere University
Kampala, Uganda
jquinn@cit.ac.ug

***Abstract*—Traffic flow monitoring systems aim to provide accurate, complete and timely data for effective management of congestion. Conventional traffic flow monitoring systems are not suitable for deployment in crowed cities in the developing-world, due to the often chaotic nature of traffic there and typically limited budgets. In this paper we present the design and implementation of a low cost vision-based traffic flow monitoring system for such contexts using a mobile phone camera as a unit for data capture and transmission. We describe its architecture and hardware platform as well as the procedure used in traffic flow speed estimation. We also evaluate the system with performance using field experiments in Kampala, Uganda. From the experiments, we observe that the system provides a feasible method of continuously monitoring traffic congestion while reducing deployment costs drastically compared to other technologies in current use. The system also performs well according to other criteria such as ability to operate without maintenance.**

I. Introduction

Many cities in the developing world are experiencing problems of increased traffic and congestion as a result of rapidly increasing population. For example population in Kampala has increased by about 600% and the population of vehicles by about 1400% between 1948 and 2012 [1]. As a result traffic and transportation managers have been challenged with how to manage the increasing traffic flows and congestion, as well as how to improve roadway capacity and efficiency with limited budgets and space for expansion. The problem is compounded by lack of real-time information on traffic flow, congestion levels and lack of resources for optimizing these levels. Lack of information is particularly problematic because traffic flow and congestion levels exhibit spatial and temporal unpredictability. They are spatially unpredictable in that we can find two parallel roads when one is heavily congested while the other is almost empty. In the temporal sense, heavy traffic flow and congestion levels occur at unpredictable times such that traffic flow and transportation managers as well as road users can not predict the traffic flow and congestion levels until they occur.

Although there exist a variety of solutions for automatic traffic flow and congestion monitoring, these solutions are not suitable for chaotic traffic flows which characterize traffic found in crowded cities in developing-world [2], as well as having prohibitively high costs in terms of purchase, installation and maintenance [3]. In an endeavor to alleviate the problems of traffic flow, transportation managers in developing-world cities have resorted to use of closed-circuit television camera (CCTV) networks to monitor traffic flow and congestion levels. This has been hindered by high costs of CCTV cameras, costs of installation and maintenance [4], [5]. In addition, they lack sufficient staff to process and analyze the data from the CCTV networks and sufficient knowledge to maintain the networks. For example in Kampala about 50 CCTV cameras were installed in 2007 but traffic flow management is still a problem due to lack of sufficient staff to process and analyze the data. At the time of writing, many of the installed CCTV cameras are currently nonfunctional due to inadequate maintenance.

In this paper we present a prototype for vision-based traffic flow monitoring which employs a mobile phone camera as a unit for data capture and transmission. The paper expands on our previous work by providing field experiment results for assessing accuracy of the prototype in terms of traffic flow speed estimation, prototype cost (assembling and installation), its deployment feasibility and maintenance.

II. Related work

Several attempts to design and implement systems for chaotic traffic flow monitoring have been made. The use of cellular phones as probes to monitor traffic flow and congestion have been reported in [6], [7]. Cellular phones probes provide rich information on traffic flow but raises a lot privacy concerns from the participants since they may end-up revealing their private information [8], [9]. Another approach is the use of acoustic sensors placed at known distances between them [10], [9], [2]. In [2], Sen *et al;* proposes an algorithm for estimating traffic speed based on Doppler effect of the sound originating from vehicular horns. The algorithm provides 70% accuracy, but incurs high computational costs [9]. In order to reduce on the computational cost of the algorithm , Sen *et al* [9] uses the RF link variations to estimate the congestion level (traffic queue length). The algorithm achieves a 90% accuracy in congestion level prediction but requires a lot of data for training[11] at the same time do not provide visual information about traffic scenes.

There have been several attempts to monitor traffic from images and videos of chaotic traffic scenes [11], [12], [13], [14], [15]. The work by Idé *et al* [12] and Jain *et al* [13] noise data from resolution camera to monitor traffic congestion. The research in [13] relies on the usage of image feeds from CCTV camera which are costly for the developing- world while work in [12] still lacks empirical evaluation of the accuracy of the algorithms yet. Sen *et al* [11] describes a method for evaluating a road traffic congestion prediction algorithm based on colored

 ISBN: 978-0-7972-1533-7

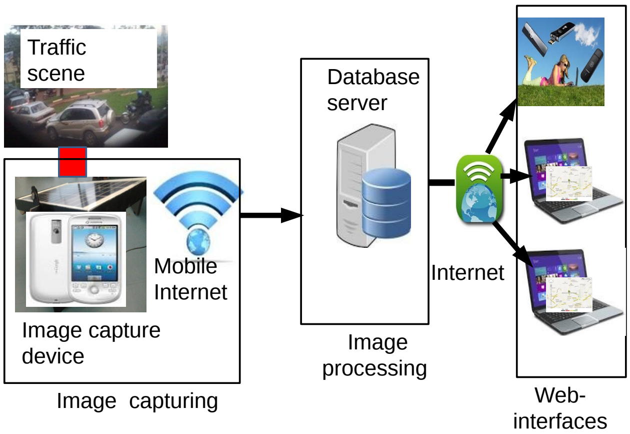


Fig. 1. Architecture of traffic monitoring system.

Fig. 2. Image capturing unit. Top left: internal components; top right: assembled unit; bottom left: wiring for external aerial; bottom right: deployed unit.

strips placed on the road side. The algorithm operate by calculating the percentage of the strip visible in the image through image segmentation which makes it not suitable for cluttered chaotic traffic where image segmentation is hard to archive at the same time employs high resolution expensive cameras.

Due to the limitations in current algorithms, in this paper we present a technique to measure speed based on video processing of chaotic traffic and its evaluations. The techniques does not perform any background subtraction/segmentation (which is often impractical on very crowded and cluttered road views), but rather the speed computations use feature correspondences between consecutive frames.

III. Prototype design, implementation and Testing

The prototype consists of three components that is the image capture, image processing and a set of web-interfaces (Fig 1).

A. *Image capturing Component*

The image capturing component is used to capture images from traffic scene. It consists of an Android Ideos U8150 mobile phone, a 7.2Ah battery pack, a 14W, 22V solar panel, a charging regulator and a steel box. The camera on the mobile phone is programmed to capture a predefined number of images at predefined time interval and upload them to the server through a wireless internet connection. The communication between the phone and the server is done through the Advanced Message Queuing Protocol(AMQP) which is an open standard application layer protocol for message-oriented middleware. By using this protocol we are able to send multiple image files at once and hence reducing on the bandwidth needed in uploading files. To test the feasibility of this mechanism we performed field experiments where monthly subscription for 350 MB of data at a cost of 15000 Uganda shilling ($6) was made at the start of the month of October 2012. This bandwidth was able to sustain the Internet connection for the phone for the whole month (between 1^{st} October and 31^{st} October 2012). The amount of data captured and sent to the server by the camera set to capture 5 images each of 1280 x 768 pixels at 2 minute interval for 24 hrs and upload interval of 2 in that month was 30GB compared to 350MB used in the upload.

The solar panel is used to charge the unit and its installed on the top of a steel box which encloses the mobile phone, the battery and the regulator. The solar panel tops up the battery pack via the charging regulator so that the unit has extra charge in case of several consecutive overcast. The mobile phone is charged by the battery pack. The arm extending from the solar panel allows the unit to be bolted to a wall or post, and the camera can be rotated through two axes. The steel box offers good protection to the mobile phone, the battery and the regulator from bad weather conditions and theft but it acts a Faraday cage for the mobile phone by cutting out reception for the mobile phone. In order to have continuous reception, we connect a wire from the mobile phone's internal antennae to a wire outside and we drill small holes in the bottom of the steel box to allow fresh air to circulate in the system so as the unit does not over heat. For a complete assembly of the various parts of the image capture unit see Fig 2.

B. *Image processing component*

The image processing component runs on the server, and comprises of several steps.

1) ***Road geometry estimation and camera calibration****:* Camera calibration is used to determine the projection equation between the world coordinate and image geometry. i.e to map a point $\left(x^{(im)}, y^{(im)}\right)$ in the image plane to a point $\left(x^{(w)}, y^{(w)}\right)$ in the world plane. In this work we perform an off-line manual calibration based on direct linear transformation(DLT) analysis as in [16]. A square grid measuring 1m x 1m is placed in the world coordinated system of a traffic scene with points at the border forming the set of control points denoted

 ISBN: 978-0-7972-1533-7

by $\left\{\left(x_i^{(w)}, y_i^{(w)}\right)|i = 1,\dots,4\right\}$ are obtained using GPS system and their corresponding image coordinates denoted by $\left\{\left(x_i^{(im)}, y_i^{(im)}\right)|i = 1,\dots,4\right\}$ are obtained from the captured images. Given this correspondence we compute the perspective transform **P** between the image plane and the World plane as portrayed in equation 1 through direct linear transformation analysis and we can infer two this transformation whenever we want to obtain the "real world" speed on the traffic

$$\begin{bmatrix} \lambda * x_i^w \\ \lambda * y_i^w \\ \lambda \end{bmatrix} = P \begin{bmatrix} x_i^{im} \\ y_i^{im} \\ 1 \end{bmatrix} \tag{1}$$

where P is a 3×3 matrix, $i = 1, 2, \dots, N$ with $N = 4$. This process is performed every time the camera changes position or location.

2) ***Specifying regions of interest (ROIs):*** We define a polygon R in the image enclosed by the image coordinates $\left\{\left(x_i^{(im)}, y_i^{(im)}\right)|i = 1,2,\dots,N\right\}$ where $N \geq 4$ specifying the region of interest we want to monitor and speed estimations are based on this region of interest. This is done manually at the beginning of the monitoring process.

3) ***Feature flow extraction:*** Informative features are extracted from the uploaded images using the method described in [14] Given an input of images taken at regular intervals; the first step is to take pairs of images and calculate the correspondence between the two, so that we can obtain flow vectors corresponding to every moving object. In this work we use the feature flow approach to extract the moving objects in that it is first and it can be incorporated into object recognition tasks. To extract important features or key point features, we use the scale and rotation-invariant detector and descriptor called Speeded-Up Robust Features (SURF) developed by Bay *et al* [17] in OpenCV, being faster than the popular SIFT descriptor [18]. From the extracted SURF features, we establish correspondences between correspondences between frames. SURF features are vectors which describe a visual feature in a representation invariant to rotation and scale. The SURF features are identified by applying difference of Gaussian convolutions of second order derivatives to an image at different scales. The descriptors are then created based on Sum of Haar Wavelet Responses around the interesting point. Then we use the SURF features to calculate correspondences between each frame, giving us a set of motion vectors /flow vector in the coordinates of the image as in [14].

We then infer road geometry in relation to the camera, to project those vectors into 'real-world' coordinates, allowing us to calculate speeds in km/h.

C. Web interfaces

The web-interfaces provides a mechanism for visualizing traffic information and for monitoring and updating the camera. Two kinds of web-interfaces are implemented.

1) ***Administrative web-interfaces:*** These interfaces provides a way for the administrator to monitor and update information about the installed cameras. Through these interfaces (s)he is able change camera setting which may include, the number of images to be captured , the image capture and upload interval.

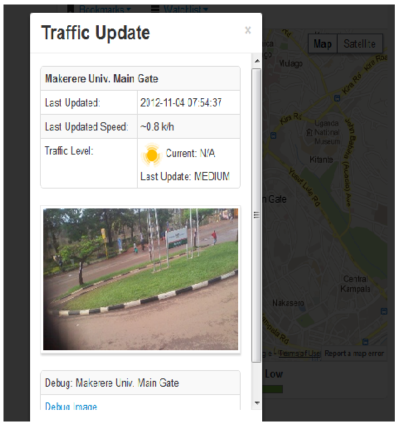


Fig. 3. Information visualization

2) ***Information display web-interface:*** This interface is used to display traffic information to the road user. Fig 3 shows a sample interface for displaying traffic information where the user can view current level of traffic flow at the same time visualize it from the image itself originating for the scene.

IV. Speed Estimation

During the estimation of traffic flow speed, we first specify the direction we want to monitor traffic. This is necessary for instance where there are two traffic flows moving at relatively the same speed in the opposite direction; without constraining the calculation to particular regions of interest and directions of flow may end up canceling one another and we observe zero flow. By specifying the direction of flow the application will consider only flow vectors moving in that particular direction in calculating the traffic speed within the region of interest.

A. Motion vector filtering by direction

Since the traffic flows we are handling can move in any direction with no defined lane markings, we first define a direction vector u(δx, δy) which specifies the direction in which we want to monitor traffic. Then for each flow vector $\{x_i^w|i = 1, 2, \dots, N\}$ computed, we compute the angle α between the flow vector as

$$\alpha = \arccos\left(\frac{u * x_i^w}{|u| * |x_i^w|}\right) \quad , \tag{2}$$

and set a threshold T such that a flow is said to be moving in the selected direction if $\alpha \leq T$. After obtaining all the flows moving in the specified direction, we apply a set of filtering rules to filter out all flows in the direction which may come as a result of mismatches.

1) *Filtering fast moving motion vectors:* When the selected motion vectors are mapped into the world coordinate system, some flows are moving very fast, so we filter out all the flows which are moving with speed more than $100km/h$ as they are most likely not vehicle flows; since most of the vehicles in the free flow tend to move with an average speed of between $60km/h$ and $70km/h$ as presented in [19].

 ISBN: 978-0-7972-1533-7

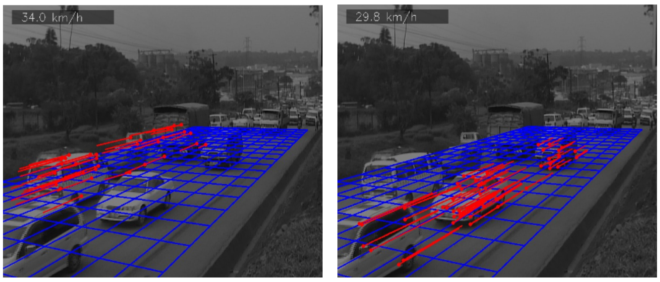


Fig. 4. Filtering motion vectors.

TABLE I. COST OF HARDWARE

Component	Cost ($)
Mobile phone	76.9
Battery pack	30.7
Regulator	7.7
Solar panel	19.2
frames,bolts and labor	19.2
Steel box	19.2
Installation	7.7
Total	**180.6**

2) Filtering motion based on (meter/pixel) ratio: Here we investigate the ratio ϱ defined by $\frac{L_{meter}}{L_{pixel}}$ where L_{meter} is the length of flow in meters and L_{pixel} is length of flow in pixels. When ϱ is high beyond a certain threshold T_{ratio}, we filter out the flow since this could indicate a mismatch between feature flows.

V. Performance Evaluation

The section presents results of field experiments carried out to access the accuracy our prototype in traffic flow speed estimation and costs for the hardware.

A. Hardware (Image capturing component)

*1) **Cost**:* The cost of the various components needed in the construction and installation of the image capturing unit is given in table I. The total cost of unit with installation costs inclusive is 180.6 USD compared to the cost of Traffic CCTV and its installation which is estimated between 9,000 and 19,000 USD [4], [5], while a machine vision sensor at an intersection is estimated between 16,000 and 25,500 USD per installation [3]; which makes it a more cost effective solution option for developing-world.

*2) **Robustness to changes in weather conditions**:* To test the robustness of the unit to the changes to weather condition the unit was installed in the field (Makerere main gate) for a period of 22 months from September 2012 to July 2014. The unit was exposed to various weather conditions and remained functional. Fig 5 shows the internal structure of the unit and solar after deployment.

*3) **Maintenance**:* As we tested the unit robustness to changes in weather conditions; we also tested how frequent the unit parts needed maintenance or replacement. It was observed throughout the period that all parts functioned as required and no maintenance or replacements were required, despite the unit being fully exposed to tropical weather throughout the testing period.

Fig. 5. Internal parts of the unit and solar panel after 22 months of continuous deployment.

B. Software

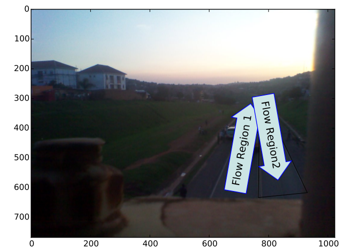
Software evaluation was done to assess how accurate it is in estimation of traffic flow speed. The system was tested with live traffic from the field for 3 hour. The system was set to capture five images every after 2 seconds delay and upload them at 2 second interval to the server. As the images arrived at the server the image processing software running at the server is triggered to extract the average flow speeds which from the uploaded images. Then the estimated speed is used in calculation of a speed moving averages as our long term is to estimate the average speed of traffic flow within a given time frame. A car with a GPS logger system was drove through the camera field of view. The speed estimates from the GPS logger was used to provide ground truth information on the accuracy of the software in speed estimation. The results in Fig 6 shows the speed moving average for both data obtained from the GPS logger and the software estimation with a window of 10 seconds. From the results it is observed that:

1) When the car is within the region of interest approximations from the GPS and software are similar. In figure 6(b)) the car started in the region of interest 1 and moved past the region; while in figure 6(c)) the car started away from the region and moved towards the region of interest and beyond. From 6(b)) we see the results from both the GPS and software are similar between the first 10 seconds the time when the car was in the region of interest and the same applies to Fig 6(c)) when the car approached the RIO between time interval of 50 and 60 seconds.
2) As the car moves out of the region of interest there is a lot of variations in both results. This is because, the status of traffic flow in the current location of the car may differ from the one in the region of interest. For example in the current location of the car the traffic flow may be free flow while in the region of interest, there is no flow (empty).
3) When the region of interest selected is large, the average flow speed becomes small see Fig 6(d). This because it includes a large set of objects/vehicles whose variations in speed becomes significant.

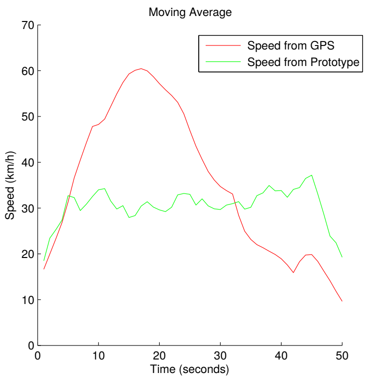
VI. Conclusion

The study has explored the evaluation of the performance of a software and hardware prototype for video-based traffic flow monitoring using the camera of a mobile phone as the basic unit of data capture and transmission. The evaluations

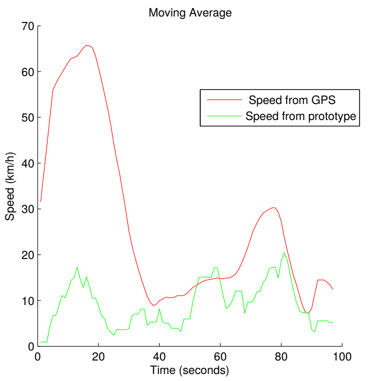
 ISBN: 978-0-7972-1533-7



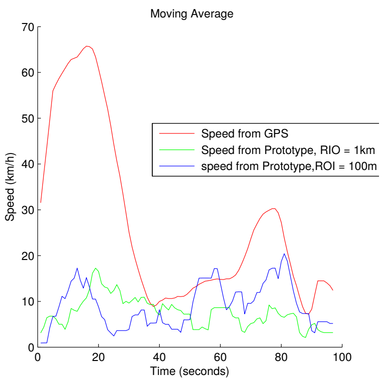
(a) Separate regions of interest within a single field of view.



(b) Speed in region 1



(c) Speed in region 2



(d) Average speed vs size of ROI

Fig. 6. Comparisons of inferred speed from vision system with speed recorded by GPS logger on board a vehicle.

shows that the prototype reduces the cost of deployment by 82% compared to video systems currently on the market at the same time providing public real time access to data from permanent installations of our prototype in Kampala as a resource for researchers of developing-world.

The initial results obtained shows that the performance of the software varied depending on the size/length of the region of interest but still suitable for crowded cities in that it makes no assumption of vehicles travelling in fixed lanes, absence of clutter or possibility of segmenting individual vehicles.

The main limitation in the software evaluation was the inability to get enough vehicle probes equipped with GPS in the regions of interest to provide enough ground truth information which could help in improving the evaluation results of the software. The other challenge faced was determining the appropriate interval for capturing images since with large time interval, a lot of information is lost especially during free flow while with short intervals lot of data is sent on the network which leads to congestion hence reducing the performance of the system. This calls for further investigation for better understanding of the appropriate interval and if possible setting dynamically depending on the level of traffic flow.

REFERENCES

[1] P. S. (POPSEC), "The state of uganda population report 2012; uganda at 50 years: Population and service delivery; challenges, opportunities and prospects," 2012. [Online]. Available: http://popsec.org/wp-content/uploads/2013/10/SUPRE-REPORT-2013.pdf

[2] R. Sen, B. Raman, and P. Sharma, "Horn-ok-please," in *Proceedings of the 8th international conference on Mobile systems, applications, and services*. ACM, 2010, pp. 137–150.

[3] G. Leduc, "Road traffic data: Collection methods and applications," *Working Papers on Energy, Transport and Climate Change*, vol. 1, p. 55, 2008.

[4] ServieMagic.co.uk, "How much does it cost to install cctv?" 2013. [Online]. Available: http://www.servicemagic.co.uk/tips-and-advice/how-much-does-it-cost-to-install-cctv.html

[5] ServieMagic, "Cctv cameras & cctv equipment price list?" 2014. [Online]. Available: http://www.cctv-centre.co.uk/cctv/cctvprices.htm

[6] M. Prashanth, N. P. Venkata, and R. Ramjee, "Trafficsense: Rich monitoring of road and traffic conditions using mobile smartphones," Microsoft Research, Tech. Rep. MSR-TR-2008-59, April 2008. [Online]. Available: http://research.microsoft.com/apps/pubs/default.aspx?id=70573

[7] P. Mohan, V. N. Padmanabhan, and R. Ramjee, "Nericell: rich monitoring of road and traffic conditions using mobile smartphones," in *Proceedings of the 6th ACM conference on Embedded network sensor systems*. ACM, 2008, pp. 323–336.

[8] P. Händel, J. Ohlsson, M. Ohlsson, I. Skog, and E. Nygren, "Smartphone based measurement systems for roadvehicle traffic monitoring and usage basedinsurance," *IEEE Systems Journal*, 2013.

[9] R. Sen, P. Siriah, and B. Raman, "Roadsoundsense: Acoustic sensing based road congestion monitoring in developing regions," in *Sensor, Mesh and Ad Hoc Communications and Networks (SECON), 2011 8th Annual IEEE Communications Society Conference on*. IEEE, 2011, pp. 125–133.

[10] R. Sen, A. Maurya, B. Raman, R. Mehta, R. Kalyanaraman, N. Vankadhara, S. Roy, and P. Sharma, "Kyun queue: a sensor network system to monitor road traffic queues," in *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*. ACM, 2012, pp. 127–140.

 ISBN: 978-0-7972-1533-7

[11] R. Sen, A. Cross, A. Vashistha, V. N. Padmanabhan, E. Cutrell, and W. Thies, "Accurate speed and density measurement for road traffic in india," in *Proceedings of the 3rd ACM Symposium on Computing for Development*. ACM, 2013, p. 14.

[12] T. Idé, T. Katsuki, T. Morimura, and R. Morris, "Monitoring entire-city traffic using low-resolution web cameras," in *Proceedings of the 20th ITS World Congress, Tokyo*, 2013.

[13] V. Jain, A. Sharma, and L. Subramanian, "Road traffic congestion in the developing world," in *Proceedings of the 2nd ACM Symposium on Computing for Development*. ACM, 2012, p. 11.

[14] J. A. Quinn and R. Nakibuule, "Traffic flow monitoring in crowded cities." in *AAAI Spring Symposium: Artificial Intelligence for Development*, 2010.

[15] B. Maurin, O. Masoud, and N. Papanikolopoulos, "Camera surveillance of crowded traffic scenes," in *Proc. of ITS America Twelfth Annual Meeting*, 2002, pp. 28–58.

[16] W. xi feng, T. Yong, Z. zhong zhe, and L. hai ying, "Feasibility of using digital photography for environmental monitoring of animals in an artificial reef," in *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, ser. Part B6b, vol. XXXVII, Beijing, 2008, pp. 339–342.

[17] H. Bay, A. Ess, T. Tuytelaars, and L. J. V. Gool, "Surf: Speeded up robust features," *Computer Vision and Image Understanding*, vol. 110, no. 3, p. 346359, 2008.

[18] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[19] R. Nakibuule, J. Ssenyange, and J. A. Quinn, "Low cost video-based traffic congestion monitoring using phones as sensors," in *Proceedings of the 3rd ACM Symposium on Computing for Development*. ACM, 2013, p. 52.

 ISBN: 978-0-7972-1533-7