Internet of Video Things in 2030: A World with Many Cameras

Anup Mohan, Kent Gauen, Yung-Hsiang Lu
Purdue University, West Lafayette, Indiana
{gauenk, yunglu}@purdue.edu

Wei Wayne Li and Xuemin Chen
Texas Southern University
{liww,chenxm}@tsu.edu

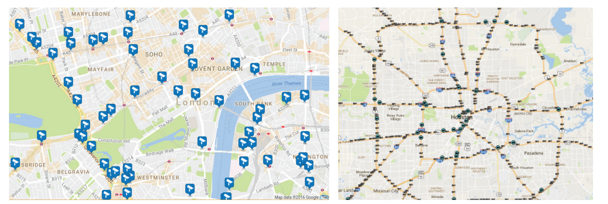
***Abstract*—The Internet of Things (IoT) is the internetworking of a variety of devices, including sensors. Among all sensors, visual sensors (i.e. cameras) are special because they can provide rich and versatile information. The world already has more than one billion cameras on mobile phones. We define the internetworking of visual sensors as the Internet of Video Things. This article estimates the number of cameras the world will see in 2030 and the implications of a large number of cameras. Transmitting, storing, and analyzing the data from cameras could impose significant challenges to existing technological infrastructures. This paper surveys recent progress in relevant technologies and suggests directions for future research.**

I. A World of Many Cameras

The Internet of Things comprises of many sensors connected to the Internet and can provide continuous real-time data [1]. The Internet of Video Things comprises of visual sensors (e.g. cameras) connected to the Internet. Cameras are unique in their ability to capture rich contextual and behavioral information. Many different types of information can be captured by cameras: crowds in a city, animals in natural reserves, congested highways, or damaged buildings after an earthquake. The same image can provide rich information. For example, a single image taken in a flooded city may simultaneously show the level of flood water, the number of people on a roof waiting for rescuers, and a driver trapped in a disabled vehicle.

There are already more than one billion cameras on mobile phones. In 2015 alone, nearly 1.5 billion smartphones are sold [2]. It is estimated that nearly 2.6 billion smartphones will be in use in 2017 [2]. There are approximately 1.8 billion registered vehicles [3]. Many vehicles have front-facing cameras (also called dashcams). New vehicles must have rear-facing cameras starting from 2018 [4] . Some car manufacturers consider to replace side mirrors with cameras and project the views on dashboards because the cameras promise to eliminate blind spots. If every vehicle has two cameras (front and back), there will be over three billion more cameras.

How many more cameras have been deployed and will be deployed? Will the world see a trillion cameras? Let's attempt to answer this question in the not-so-distant-future year 2030. How many cameras can the world possibly have? How would the data from these cameras be used? Where will the data be stored? What are the implications for networking?



(a) London (b) Houston

Fig. 1. Locations of traffic cameras in London and Houston.

To answer these questions, this paper first examines where the cameras may be deployed. The following is a list of locations where cameras are likely deployed. (a) Mobile phones. It is projected that the world population will exceed 8 billion by 2030 [5]. If one third population has mobile phones and each mobile phone has 2 cameras, there will be 5 billion cameras. (b) Motor vehicles. Today, many vehicles are idle at park lots. The automobile industry is going through significant changes. If fully autonomous vehicles become widely adopted in 2030, it is possible there are fewer vehicles at higher utilization. This paper assumes that there are one billion vehicles and each has four cameras, together with many other types of sensors. (c) Civil Infrastructures: Highways, Bridges, Traffic Intersections. Figure 1 shows the locations of traffic cameras in London and Houston as an example. Based on the World Fact Book [6], the world has approximately 19 million kilometers of paved roads. If one camera is deployed every 100 meters, there will be 190 million cameras. (d) Buildings (such as airports, offices, classrooms, hospitals, residential houses). We did not find a single source that provides the number of buildings and the amount of floorspace. Instead, we use the numbers from the USA as references. The USA has 5.6 million commercial buildings with 8.1 billion square meters [7]. In 2005, there were approximately 111 million houses in the USA [8]. The population in the USA is approximately 300 million. Extrapolating to 8 billion people, there will be 210 billion square meters of commercial space and 2.9 billion houses. Such extrapolation may be wrong, due to the vast differences in lifestyles in different countries. Moreover, urbanization is a global trend, and more than half of world population will live in cities [9]. Nevertheless, this extrapolation can serve as the "first-order" approximation. If there is a camera every

978-1-4673-6853-7/17/$31.00 ©2017 IEEE

thousand square meters (at building entrance, for example), there will be 210 million cameras in commercial buildings. If each house has one camera looking at the front doors, there will be 2.9 billion cameras. (e) The number of cameras for any other purpose (such as observing wildlife and beaches) will likely be only a small fraction of the numbers for commercial and residential buildings.

Adding all the categories together, there are approximately 13 billion cameras. This is only ten times higher than the number of cameras in existing mobile phones. Modifying some of the assumptions is possible. For example, if there is a camera for every kilometer, there will be 19 million traffic cameras. If each vehicle has 10 cameras, instead of 4, there will be 10 billion cameras. It is likely that the number of cameras in the world would be somewhere between 10 and 100 billion in 2030. These estimates suggest that there will *not* be a trillion cameras in the world. Please be aware that this estimation is the number of deployed cameras in 2030, not the number of cameras to be sold from now to 2030. In the coming years, many people will replace their mobile phones and vehicles. Over these years, many cameras will be manufactured, sold, and disposed of.

II. Amount of Data from Cameras

How much data would these cameras produce? To answer this question, we consider three factors: (1) How many pixels would each camera have? (2) What is the frame rate? (3) How many photographs and how many hours of video do these cameras produce?

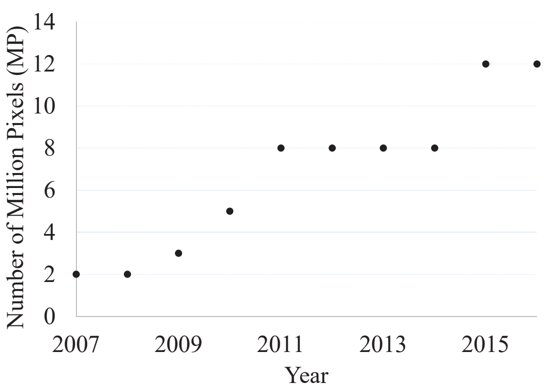


Fig. 2. Number of million pixels (MP) in iPhones.

To answer the first question, we use the numbers of pixels in iPhones as references. Figure 2 shows that in 9 years, the numbers increase by 6 times from 2MP to 12MP. This is 22% annual growth rate. If this growth rate continues to 2030, a mobile phone's camera will have 200MP. A billion-pixel camera is not too far-fetched as the European Space Agency is preparing a billion-pixel camera [10]. We do not expect every camera to have more than 200 million pixels. Replacing cameras deployed in buildings or along highways can be expensive. It would be difficult to determine the average number of pixels in the billions of cameras. Perhaps 100MP would be a reasonable number for many cameras.

Cinema movies usually use 24 frames per second; videos at 30 frames per second are common. Two recent movies experiment higher frame rates. The Hobbit uses 48 frames per second, and Billy Lynn uses 120 frames per second. It is yet too early, and there is not enough evidence whether the higher frames would become main streams. For now, we assume that most future cameras will provide at most 30 frames per second. A single HD (1MP) network camera can generate more than one Mbps data [11]. In 2030, if 100MP is the average number of pixels among all cameras; then each camera can generate 100 Mbps data. This is quite modest.

To answer the third question, we consider smartphones and other cameras like network cameras separately. The number of photos uploaded to the Internet by smartphones have increased rapidly over the past decade, and more than 1 billion photos were uploaded per day in 2014 [12]. We expect this trend to continue in 2030 due to the technological advancements in mobile camera, more time spent by the user on digital media [13], and increasing number of mobile applications using images and videos. By 2030, we will see more videos being captured and uploaded to the Internet as the mobile video usage is growing stronger every year [13]. However, we do not expect the mobile users to record videos for long durations. With the number of network cameras deployed increasing every year and a strong estimated growth in Internet surveillance video traffic [14], we expect strong growth in the amount of videos recorded using network cameras. Su et al. [15] estimates more than 16PB of data can be saved per day from 10 million network cameras at a rate of 0.15Mbps. If we consider 10 million cameras, with an average resolution of 10MP, streaming data at 30 frames per second, we can estimate the data generated by these cameras to be greater than 1EB per day by 2030.

III. Network

The vast amounts of video data generated by the cameras flow through the network to be analyzed at servers (e.g., cloud). The Cisco Visual Networking Index [14] (*Cisco VNI*) predicts that the IP video traffic will be 82% of the consumer Internet traffic by 2020 from 70% in 2015. Following this trend, we assume that the IP video traffic will be more than 95% of the Internet traffic by 2030. Satyanarayanan et al. [16] suggest that the current Metropolitan Area Networks (MAN) with a bandwidth of 100Gbps, can support a maximum of 12,000 users uploading 1080p video over the Internet. Even though the bandwidth of the MAN will increase by 2030, the data produced by the cameras and number of users analyzing the data will be sufficiently higher and create bandwidth issues. When the network cannot support the video traffic, the quality of the data (e.g., bit rate) is reduced, and this may impact the accuracy of many analyses to be performed on the video. Since energy is the primary constraint for edge devices, performing all the analysis at the edge may not be feasible as video analyses are computationally intensive.

As the video traffic on the Internet is expected to be more than 1 ZB per year in 2030, the case for distributed cloud architecture in networks becomes more important. The key concept of such methods is to have intermediate machines (e.g., cloudlets) in the network between the camera and the cloud servers [17]. The cloudlets will analyze the video data and send only the required information over the Internet thus reducing traffic. The increase in usage of video data for emergency response situations calls for intelligent routing mechanism in the Internet [18]. For example, in the case of a natural disaster (e.g., tornado, flood), the emergency response team should be able to retrieve and analyze video streams from the affected areas in real-time. The data requests from the response team should receive higher priority as the bandwidth is limited especially due to the network links being damaged by the disaster. Hence more studies should be taken on the concepts of OpenFlow and Software Defined Networking (SDN) to implement intelligent routing over the network.

IV. Processing

Recent advancements in the design of CPU, GPU, and FPGA have made it possible to analyze large amounts of visual data in real-time. Frameworks like TensorFlow [19] make it easier to execute applications on heterogeneous hardware. Image processing using deep learning is a promising field, and the implication of the current success appears vast. Because of the popularity and success demonstrated using image data for deep learning, and the growth of Artificial Intelligence applications, it is reasonable to assume large quantities of image data will be collected and analyzed.

As explained earlier, due to bandwidth constraints, it is important to analyze data at the edge devices and/or intermediate stages (e.g., cloudlets). Deep learning models suggest a potential reduction in quantity of information needed to be saved and transferred across a network. This implies a need for implementing learning and analysis on-board along-side cameras.

In order to reduce the amount of data, a camera may only need to report the analysis result of a given image rather than the image itself. For instance, sociologists find people to be of interest in the analysis of behavior [20]. A camera which reports the behavior of people, instead of the original image, will likely need fewer bits than the original image. The pixels will be transformed into a vector of behavioral indicator values. This analysis may be done at the cameras. However, it is worth noting that local computation cannot solve all the potential issues. Sometimes it is useful to use stereo images to perform data analysis, and multi-camera object tracking obviously needs to share data beyond single cameras. Local image processing cannot be used to reduce that kind of network traffic.

In the future of Internet of Video Things, we expect to see analysis being performed at different stages. The edge devices can perform light computations, whereas the intermediate stages (e.g., cloudlets) can perform heavier computations. Cloud servers can be used for the heaviest computations as well as to save the results. Many research questions arise under this scenario: (i) Can the application be executed at different stages? (ii) How to partition the application? (iii) How to select the partition and stages to execute in order to optimize the network bandwidth usage, overall cost, and performance?

V. Storage

Su et al. [15] estimates the number of hard drives and the cost to save video data from 10 million network cameras for a year to be 15 million disks and $2B respectively. Even though this is expensive, it is still feasible. With the number of video sources growing rapidly, we take a closer look at the amount of storage available and the size of the video data being generated. To estimate the amount of storage available, we use the total size of the hard disk drives shipped by the two major vendors Western Digital and Seagate [21], and predict the storage size for 2030. To estimate the size of the video data, we only consider surveillance videos as it is common to save the surveillance video data. We use the Cisco VNI [14] prediction for Internet video surveillance traffic for 2020 and estimate the data size for 2030.

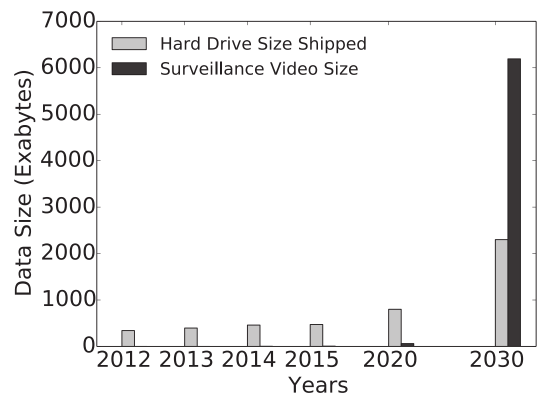


Fig. 3. The total size of hard drives shipped by the vendors (Western Digital and Seagate) and the Internet video surveillance traffic estimates using Cisco VNI for different years. The surveillance video data expects to be more than twice the total storage size available by 2030.

Figure 3 shows the total size of hard drives shipped by the vendors and the Internet video surveillance traffic for different years. Cisco VNI [14] estimates the surveillance video traffic to increase 10 times from 2015 - 2020. Keeping the same trend, we estimate the video traffic to increase 100 times from 2020 - 2030. The hard drive size has increased linearly from 2012 - 2015 [21]. Assuming that more hard drives will be shipped due to the significant growth of video data over the Internet by 2030, we estimate the total hard drive size for 2020 and 2030 with an exponential increase in size. Even with this aggressive estimation, Figure 3 shows that the surveillance video data expects to be more than twice the total storage size available. This suggests that at the expected growth rate, it will be infeasible to store all the surveillance video data in 2030. Also it should be noted that the surveillance video is only 4% of the Internet video traffic [14]. The total Internet video traffic is expected to be more than 1ZB per year in 2030. This exceeds the estimated amount of storage. Based on this data, we can infer the following:

- There is a need to reduce the amount of video data stored by saving final/intermediate results of analysis (e.g., number of people) and the metadata information (e.g., time, date, and location) instead of the whole video.
- There is a strong need to increase the amount of storage available to meet the requirements.

VI. Discussion

Visual data from network cameras are used for a variety of applications ranging from security to leisure, and can be important in a "smart city". The usage of visual data to aid emergency response in case of natural or man-made disasters is growing [18, 22, 23] and the trend is expected to continue. Self-sustaining, low-power cameras cleverly placed throughout a city in undisclosed locations with the ability to report crime automatically creates an profound, and perhaps formidable, society. People who need help the most, are incapable of calling for help. It is feasible to think of a system to detect criminal behavior and take appropriate action, for instance alerting the police of the crime and location. The image processing to determine if a call should be made can be performed on the camera itself. As the capabilities of technology increase, ethical questions which use to be theoretical now become real. Much of the future of this technology is bound to legal and ethical question about privacy and security [24].

VII. Conclusion

The number of cameras connected to the Internet in the form of network cameras and smartphones are increasing rapidly forming the Internet of Video Things. They provide real-time information which can be used for important applications like improving emergency response and enhancing public safety. This paper estimates the number of cameras the world will see in 2030 to be approximately 13 billion cameras. This paper also examines the impact of these cameras and massive amount of information generated by them on the networking, storage, and processing infrastructure and suggests some directions of future research. We expect the computations to be distributed and move towards the edge devices to reduce bandwidth and storage requirements in the case of large scale video analysis.

References

[1] Telecommunicaton Standardization Sector of ITU. Over of the Internet of things Series Y: Global Information Infrastructure Internet Protocol Aspects and Next-Generation Networks. *International Telecommunication Union*, 2.2060(3):1–3, 2012.

[2] Statista. Statistics and facts about smartphones. https://www.statista.com/topics/840/smartphones/.

[3] World Health Organization. Registered vehicles data by country. http://apps.who.int/gho/data/node.main.A995.

[4] National Highway Traffic Safety Administration. Final Rule Requiring Rear Visibility Technology. http://www.nhtsa.gov/About+NHTSA/Press+Releases/2014/NHTSA+Announces+Final+Rule+Requiring+Rear+Visibility+Technology, 2014.

[5] UN DESA. World population projected to reach 9.7 billion by 2050. http://www.un.org/en/development/desa/news/population/2015-report.html, July 29 2015.

[6] Central Intelligence Agency. The World FactBook: Roadways. https://www.cia.gov/library/publications/the-world-factbook/fields/2085.html.

[7] US Energy Information Administration. Commercial Building Energy Consumption Survey. http://www.eia.gov/consumption/commercial/reports/, March 4 2015.

[8] US Energy Information Administration. Number of U.S. Housing Units. https://www.eia.gov/emeu/efficiency/recs_2_table.htm, July 2009.

[9] Urbanization. http://www.unfpa.org/urbanization.

[10] Eye of Gaia: Billion-Pixel Camer to Map Milky Way. http://www.esa.int/Our_Activities/Space_Engineering_Technology/Eye_of_Gaia_billion-pixel_camera_to_map_Milky_Way.

[11] Axis Communications. Bandwidth and storage considerations. http://www.axis.com/us/en/learning/web-articles/technical-guide-to-network-video/bandwidth-considerations.

[12] Mary Meeker. Internet trends 2014–code conference. *Retrieved May*, 28:2014, 2014.

[13] Mary Meeker. Internet trends 2015-code conference. *Glokalde*, 1(3), 2015.

[14] Cisco. Cisco visual networking index: Forecast and methodology, 2015–2020. *CISCO White paper*, 2016.

[15] Wei-Tsung Su et al. Harvest the information from multimedia big data in global camera networks. In *IEEE International Conference on Multimedia Big Data*, 2015.

[16] M Satyanarayanan et al. Edge analytics in the internet of things. *IEEE Pervasive Computing*, 14(2):24–31, 2015.

[17] P Simoens et al. Scalable crowd-sourcing of video from mobile devices. In *International conference on Mobile systems, applications, and services*, pages 139–152. ACM, 2013.

[18] Rasha Gargees et al. Incident-supporting visual cloud computing utilizing software-defined networking. *IEEE Transactions on Circuits and Systems for Video Technology*, 27(1):182–197, 2017.

[19] Martın Abadi et al. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*, 2016.

[20] Binnur Ercem. Strategies for success course guide: Introduction to sociology, 2008.

[21] Market Views: Hard Drive shipments. http:www.anandtech.com/show/10098/market-views-2015-hard-drive-shipments.

[22] Sung Ye Kim et al. Visual analytics on mobile devices for emergency response. In *IEEE Symposium on Visual Analytics Science and Technology.*, pages 35–42, 2007.

[23] Youngsol Koh et al. Improve safety using public network cameras. In *IEEE Symposium on Technologies for Homeland Security (HST)*, pages 1–5, 2016.

[24] Aaron Doyle et al. *Eyes Everywhere: The global growth of camera surveillance*. Routledge, 2013.