Humanitarian Technology: Science, Systems and Global Impact 2015, HumTech2015

Up in the Air: Applying the Jacobs Crowd Formula to Drone Imagery

Austin Choi-Fitzpatrick[a], Tautvydas Juskauskas[b]*

[a]*Assistant Professor, School of Public Policy, Central European University, Budapest 1054, Hungary*
[b]*MPA student, School of Public Policy, Central European University, Budapest 1054, Hungary*

Abstract

The accurate estimation of event size is important for city planners, concert coordinators, social movements and anyone else interested in understanding how many people show up to an event. For social movements, the social theorists Charles Tilly long ago argued, large crowds signal worthiness, unity, numbers and commitment. Huge crowds on the street are a clear signal to authorities, the media, bystanders and the media itself. The same can be said for events with small turnouts. Event coordinators often have interests that lead to methods that inflate estimates. Critics and movement targets, on the other hand, are interested in minimizing the perceived size and scope of protests to their authority. Current efforts to estimate protest size have used on-the-ground methods (requiring many enumerators total control of the event area) or in-the-air methods (such as traditional aircraft, which are expensive and require advance notice and special permissions). Technological innovation involving Unmanned Aerial Vehicles (or "drones") provide an opportunity to more accurately and affordably estimate crowd size. In this brief methods article we introduce a novel adaptation of a standard estimation model (Jacobs Crowd Formula) for use on an entirely new platform.

1. Introduction

Recent technological innovation has facilitated an exponential increase in innovation and investment around Unmanned Aerial Vehicles (which we will refer to here as *drones*). Innovation has outpaced policymakers' ability to

————

* Corresponding author. Tel.: +36 1 327 3000/2366
  E-mail address: achoifitz@gmail.com

regulate its use. Regulation is a delicate matter involving numerous issues, not the least of which involve privacy and safety, which vary significantly by region. While the first author has addressed these issues elsewhere (Choi-Fitzpatrick 2014) [1], here we introduce a method for using camera-equipped drones to estimate the size of large gatherings of people using static imagery. There is every reason to believe that in time this approach could be extended to cover large gatherings of people using dynamic imagery, or to estimate the size of other gatherings.

2. A drone-based crowd estimation method

We propose the use a consumer-grade unmanned aerial vehicle (drone) to implement the Jacobs Crowd Formula (JCF, hereafter; see Jacobs 1967; McPhail, McCarthy 2004) [2,3]. While we suggest several modifications (listed below) they are simple extensions of the JCF. Thus, the main advantage of the proposed method is its ease of use. While more sophisticated crowd counting techniques are available, using complex mathematical modeling, these require special computing and financial resources and specific technical knowledge (and are usually optimized for CCTV footage). They are therefore difficult to replicate in places where these resources are difficult to obtain. The following method rests neither on sophisticated image processing nor on computer vision techniques. It rests instead on the JCF's underlying logic. With several modifications, it incorporates the following factors and actions:

1. Drone platform;
2. Digital image;
3. Area measurements;
4. Grid digitally applied to image;
5. Determine density levels in each grid;
6. Compile estimate of crowd size (add all grids, with corresponding density levels).

2.1. Drone platform

For the purpose of our testing, we decided to choose a drone that was affordable without compromising on quality and having in place all relevant technical specifications i.e. camera, GPS, heavy duty-load, better flight time etc. The chosen drone came with a 14 megapixel camera capable of producing 1080p HD video at 30 frames per second through the Vision app which also comes free with the DJI Vision and are compatible with both Android and iOS. The Vision also enables tilting the camera and thus producing smoother video. It has a 5200 mAh LiPo (lithium polymer) battery that provides 25 minutes of flight time on a single charge. It can transmit photos and videos in a maximum of 300m range connected through Wi-Fi and pre-installed GPS which also provides better control and stable hovering. Another advantage of the DJI Phantom 2 Vision was its compatibility with both iOS and Android operating systems, making it easy to attach the drone with existing phone or a tablet. When connected through GPS, the camera installed in the drone sends and stores its captured images and videos directly on the connected device – phone or tablet. Its auto-return mode ensures that the drone could return safely to the take-off point. This works as a safety mode when control of the drone could be lost at certain point and then the drone would return safely without causing any damage or revealing privacy. This mode, however, is dependent on availability of GPS connection.

2.2. Digital image

We made one important modification. Upon purchase, the camera did not point down at the 90 degree angle we determined was best for an accurate headcount. Jacobs' approach required estimates from multiple vantage points because images made at an angle from atop a building were often blocked by trees, vehicles, and other buildings. We modified the UAV frame to ensure the camera was angled perpendicular to the ground, effectively eliminating this requirement. This modification has the additional benefit of increasing privacy, as only the top of heads are visible. Early images were made using an ad-hoc modification, though we later accomplished the same objective with a 3D-printed gimbal. Many drone-mounted cameras create aspherical distortion, or a "fish-eye effect." In order to correct for this distortion we modified to the image using open-source software.

2.3. Area measurements

Accurate measurements of the area are paramount. While McPhail and McCarthy (2004) recommend the use of official maps, our method is designed for maximum replicability, including areas where the state is weak (no maps) or closed to the public (no detailed maps available). With addition to this, accurate measurement is important for the configuration of a regular square grid. For this reason we devised two methods for area measurement. In the first we laid a 10-meter marker onto the ground and used that as our reference point. This is ideal in cases where it is possible to scout a location prior to an event. In the second we identified key features (e.g., large vehicles, small buildings, parking stripes) visible from the air and used those as reference points. This approach is best in cases where advance scouting is impossible or dangerous. However, this method requires knowing the exact length of those reference points. Alternately, the measurement of an object can be done after the actual event. This is the practice we have been applying.

Once the exact length of the reference point or line has been determined, it must be translated into pixels as this is the unit of analysis for digital imagery. Note that the grid squares are also measured in pixels. The reference line has to be changed to 10m in order to ensure that the grid square in the picture is equal to a 100 m2 (10x10) area on the ground.

A series of actions are required in order to 1) translate a reference line into pixels and 2) to determine how many pixels is the line of 10m on the ground. The measuring tool can be accessed by entering the GIMP's main menu's section "Tools" and selecting "Measure." The measure tool is activated by clicking and holding the mouse button and then dragging it the length of the item to be measured. The information is displayed on the status bar or can be displayed in the Info window." It is unlikely that the reference line is exactly 10 meters long. The second step, therefore, requires a bit of simple math, which allows determining how many pixels correspond to a 10 meter reference line of the ground, described in below and depicted in Table 2.

Let's imagine that the real reference line visible in a picture is 4.5 m. long. It represents 125 pixels (px) in a picture. This equation has only one unknown (let's say y), which is the number of pixels that correspond to a 10 m reference line. The first action is cross-multiplication: 4.5 times y equals 10 times 125.

$$\frac{4.5\ (m)}{10\ (m)} = \frac{125\ (px)}{y\ (px)} \qquad (1)$$

$$4.5y = 1250 \qquad (2)$$

$$y = \frac{1250}{4.5} \qquad (3)$$

$$y = 278\ (px) \qquad (4)$$

This equation is used to determine how many pixels are equal to a 10 m. reference line in a given image. In such case 10 m. reference on the ground equals 278 pixels in the picture. This means that the dimensions of a grid would be 278 x 278 pixels. Table 2 shows a few dimension-sizes at three standard altitudes.

Table 1. Area Measurements and Crowd Estimation.

Altitude in meters (feet)	Photo dimensions in pixels after fish-eye correction	Fish-eye correction in GIMP software (main, edge)	Reference on ground in meters	10m on ground in pixels	10m x 10m on ground in pixels
50 (164)	4384x2466	-20, -20	10	533	533x533
100 (328)	4384x2466	-20, -20	10	270	270x270
150 (492)	4384x2466	-20, -20	10	174	174x174

2.4. Grid digitally applied to image

Placing a digital grid over the digital image allows for the rapid estimation of individual unit density and counting of total units. The grid should be composed of squares, whose dimensions should be - 10 m x 10 m on the ground (reference size, area = 100 square meters). After determining the number of pixels that correspond to the 10 m. reference line, a simple grid can be applied to the picture. A grid application is accomplished in two basic steps using the chosen GIMP 2.8. Grid can be applied according to the following menu functions: Image → Configure grid; View → Show grid.

The first step is to configure a grid. This requires selecting the "Image" option in the main menu and then selecting "Configure Grid." The grid dimensions are then entered into the section titled "Spacing." In this case 278px x 278px need to be entered into the fields of "width" and "height." The second step involves displaying the grid, this can be completed by choosing "View" on the main menu within software and selecting "Show grid."

2.5. Estimating the density levels of each grid

With the grid then applied, and with each grid measuring 10 meters between each gridlines, it is now possible to estimate the number of individuals within each grid. Using (Western) density levels established in the literature, we are able to base estimates on five density levels, effectively, where there are no people, where the crowd is very loose, relatively loose, relatively dense, and very dense. The five possible density levels are described in detail in the Appendix A, Table 2.

2.6. Compile estimate of a crowd size

The sixth and final step is counting how many squares of different density levels the grid has. The actual number of the crowd is summed up. This full method with its detailed instructions can be seen in Appendix A. This appendix clarifies the coding instructions given to the individual enumerators we used to establish the estimates used in this study. Our intention is that this article contains all information necessary for replicating this study and generating new estimates of one's own. In our other work (Choi-Fitzpatrick, Juskauskas and Sabur, nd) we have complemented this method with a Cohen's Kappa estimate of inter-coder reliability. Some users may choose to incorporate Cohen's Kappa as an optional 7th step in this estimation methodology. We find that this is the most conservative and robust approach to developing an auditable range estimate of crowd size. While single-person estimates might be preferred by advocates hoping to motivate supporters and authorities needing to produce an official statistic, it is not as statistically robust as the range provided by two coders using Cohen's Kappa.

3. Conclusion

This method improves on the status quo dramatically. Drone-based imagery captures crowds at a ninety-degree angle, thereby improving on pictures taken from the tops of buildings, since it is easier to estimate at this angle, and it is considerably easier to protect privacy and anonymity. The platform also improves on aircraft-mounted

solutions, which are expensive, and rely on governments to provide access to airspace. While these technical and economic benefits are important, actual implementation must adhere to appropriate ethical standards. We propose any future use adhere to the guidelines advanced in Choi-Fitzpatrick (2014). Specifically, any use of this new technology for civil society use (that is, nongovernmental and nonprofit) must prioritize six factors:

- **Subsidiarity** – The concept of subsidiarity suggests that decision-making and problem solving should occur at the lowest and least sophisticated level possible. The implication here is that a drone should only be used to address situations for which there is not a less sophisticated, invasive, or novel use.
- **Physical and material security** – This principle focuses on physical integrity issues related to the use of UAVs. Put bluntly, care must be taken so that these devices do not collide with people or with one another. Furthermore, they must not be weaponized in such a way that could cause physical harm to the public. How exactly this security is ensured is a matter of skill, which is determined by the operator, and situation, which is determined by weather and other environmental conditions. How it is defined is a matter of perspective.
- **Do no harm** – This principle draws inspiration from the UAViators' (uaviators.com) emphasis on a rights-based approach as found in the development and humanitarian aid communities. The focus is not on reducing physical and material security, but is instead on ensuring the public good (i.e., the harm in question is related to the public good rather than physical integrity).
- **Public interest** – This principle draws original inspiration from the concepts of newsworthiness and the public good, while recognizing that some seemingly insignificant or unpopular issues may be in the public's interest and for a public good without being considered newsworthy. Citizens and non-citizens should be protected from the prying eyes of the state and commerce, yet there is a need for a larger conversation about what level of privacy is to be expected when civil society actors have deployed drones for their own purposes.
- **Privacy** – Each principle must be held in balance with the others, and none more so than with respect to privacy. Citizens and non-citizens should be protected from the prying eyes of the state and commerce, yet there is a need for a larger conversation about what level of privacy is to be expected when civil society actors have deployed drones for their own purposes.
- **Data protection** – Data protection is critical. Filming a large event creates a digital record of participants. This data can be politicized in the wrong hands. As more UAVs gather more data, questions about how to handle big aerial data will emerge. Context-specific protocols must ensure the security of data, thereby protecting against physical or digital theft or corruption (Choi-Fitzpatrick 2014: 28-30).

Of course what remains is actually testing this method. We anticipate our approach is significantly robust, and will prove to have merit in the field. Efforts in this direction are under review (Choi-Fitzpatrick, Juskauskas, and Sabur, nd) [4] and available upon request. This use is in keeping with an ever-broadening conversation about how to best apply technology to the world's problems, especially with regard to humanitarianism. While our interest is in estimating the size of protest events, we look forward to seeing this method in the hands of other scholars and practitioners interested in refugee flows, concert attendance, public safety, wildlife and conservation population studies, and so forth. We look forward to future collaborations that will allow us to automate this process in ways anticipated by other methods for automating the assessment of visual data (e.g., Marana et al 1999; Ryan et al 2009; Ryan 2013; Kong, Gray and Tao 2005 and 2006) [5, 6, 7, 8, 9].

Acknowledgements

Appendix A. Instructions for crowd size estimation

Estimates can be augmented through the use of research assistants or collaborators. This requires training on each of these five possible density levels. We propose the following instructions as suitable for preparing qualified individuals to complete this task. We consider the average American college student to be a "qualified individual."

1. You have been provided with a printed photo of a crowd. This image has a grid laid over top. Your task is to examine each grid and determine which of five (5) density levels apply to it. In other words, how many people do you think are in each of the squares?
2. You will write your answer with a white pen (provided) directly onto the square in question. You will look at each square of the photo to estimate the number of people inside.
3. Your task is to determine density level of each grid, and then do the final calculation.
4. This task has two parts that are done at the same time: Estimating how many people are in each square (Part One: Determine Density Level) and estimating how much of that square is full (Part Two: Determine Percentage of Grid Filled). NOTE: *each grid will only have one Density Level.*
5. All of this can be tiring. So, feel free to take a break or two of five minutes. This will help you to relax your eyes and regain focus.

A.1. Three stages of the task of crowd size estimation

This task is accomplished in three stages: (1) determining the density level of each grid, (2) determining the percentage of each grid filled and (3) calculating the total size of a crowd size. This process is described below:

1. Part One: Determine Density Level Within Each Grid
    i. Density level X - If there simply isn't anyone in the grid, mark the grid with an X.
    ii. Density level 0 - If there are just a few people in the grid, fewer than 100 or so, then you should count each individual within that grid to come up with an exact count. You have been provided a laptop computer, with the same photo loaded in it, to help you with this count. NOTE: Be sure you're looking at the corresponding grid when you compare the paper and the screen! The laptop is not required for any other purposes.
    iii. Density level 1, 2, 3 – In Table 2 below you will see each of these density levels described.
        i. Density level 1 is a space you could walk through easily.
        ii. Density level 2 is a space you could walk through with some difficulty.
        iii. Density level 3 is a space in which it is impossible to move (like a fully packed subway car, for example).

2. Part Two: Determine Percentage of Each Grid Filled
    i. Not every grid is completely filled. You may find that a particular grid is Density Level 2 and half-full (people occupying only 50% of the space). If that is the case, you will write *both* the density level *and* the percentage filled.
    ii. The percentages are as follows:
        i. 25% (or ¼ full)
        ii. 50% (or ½ full)
        iii. 75% (or ¾ full)
        iv. 100% (or completely full)

     iii.   If you feel there are two different density levels present in one grid, then make your best guess about which one predominates.

3.   Part Three: Calculating the Total Size of a Crowd
     i.   Sum up all the results of the hard count of the Density Level 0.
     ii.   Count the number of grid squares that are ¼ full, ½ full, ¾ full and completely full within each density level.
     iii.   y is the number of grid squares; y x 0,25 – number of 25 % full grid squares, y x 0,5 – number of 50 % full grid squares, y x 0,75 – number of 75 % full squares, y x 1 – number of 100 % full grid squares.
     iv.   Apply the below formula separately to the Density Levels 1, 2, 3:

$$\big((y \times 0{,}25) + (y \times 0{,}5) + (y \times 0{,}75) + (y \times 1)\big) \times 109 - \text{Density Level 1} \qquad (1)$$
$$\big((y \times 0{,}25) + (y \times 0{,}5) + (y \times 0{,}75) + (y \times 1)\big) \times 238 - \text{Density Level 2} \qquad (2)$$
$$\big((y \times 0{,}25) + (y \times 0{,}5) + (y \times 0{,}75) + (y \times 1)\big) \times 435 - \text{Density Level 3} \qquad (3)$$

     v.   Sum up all the results and add up the hard count of the density level 0 and you will have the final estimation of a crowd.

$$(1) + (2) + (3) + (Density\ Level\ 0\ hardcount) = Total\ size\ of\ a\ crowd \qquad (4)$$

A.2. Description of Density Levels

Enumerators will find it helpful to have a guide to this estimation process. What follows is an intuitive rule-of-thumb guide that would help the average Westerner to make determinations about density based on mobility within familiar crowd settings. It can be easily modified for other cultural contexts.

Table 2. Density Levels.

Density level	Number of People	Description	
X	Zero People	DESCRIPTION: A rooftop, or any other empty space. ACTION: If you see no people in the grid, please mark the grid with an X.	a
DL 0	Up to 100	DESCRIPTION: Very loose crowd with a very low density level. You could ride your bike through this crowd easily. ACTION: If it looks like there are fewer than 100 people in this grid, then count them manually and write that number in the grid.	b
DL 1	~109	DESCRIPTION: Somewhat loose crowd with a pretty low density level. This is a crowd you could walk through easily without bumping into too many people (imagine about 1 person per square meter). On average, at this density level there are usually about 109 people in the grid or one person in 10 ft2 (0.93 m2). ACTION: If you believe a grid is full of DL1, mark it as **"1"**. If you believe it is a DL1 crowd in half of the image, write: **1 ½**"	c
DL 2	~238	DESCRIPTION: This is a dense crowd. You would have a hard time moving through this crowd, but it would be possible (imagine more than 2 people per square meter). On average, at this density level there are usually about 238 in the grid or one person in 4.5 ft2 (0.41 m2). ACTION: If you believe a grid is full of DL2, mark it as "**2**". If you believe it is a DL2 crowd in half of the image, write "**2 ½**."	d

DL 3	~435	DESCRIPTION: This is an extremely dense crowd. It would be nearly impossible to move your arms in this crowd (imagine more than 4 people per square meter!). This is the same as the very front of a concert, just in front of the stage. On average, at this density level there are usually about 435 people in the grid or one person in 2.5 ft2 (0.23 m2). *[NOTE: this density level rarely occurs]* ACTION: If you believe a grid is full of DL3, mark it as **"3"**. If you believe it is a DL3 crowd in half of the image, write **"3 ½."**	e

Fig. 1. (a) Density Level X; (b) Density Level 0; (c) Density Level 1; (d) Density Level 2; (e) Density Level 3

References

[1] Choi-Fitzpatrick, Austin. "Drones for Good: Technological Innovations, Social Movement and the State." *Journal of International Affairs* 68.1 (2014)

[2] Jacobs, Herbert A. "To Count a Crowd." *Columbia Journalism Review* 6 (1967): 36–40.

[3] McPhail, Clark, and John McCarthy. "Who Counts and How: Estimating the Size of Protests." *Contexts* 3, no. 3 (August 1, 2004): 12–18. doi:10.1525/ctx.2004.3.3.12.

[4] Choi-Fitzpatrick, Austin, Tautvydas Juskauskas, Muhammad Sabur. nd. "All the Protestors fit to Count: Using Unmanned Aerial Vehicles to Estimate Protest Crowd Size." *Under Review.*

[5] Marana, Aparecido Nilceu, Marcos Antonio Cavenaghi, Roberta Spolon Ulson, and F. L. Drumond. "Real-Time Crowd Density Estimation Using Images." In *Advances in Visual Computing*, 355–62. Springer, 2005. http://link.springer.com/chapter/10.1007/11595755_43.

[6] Ryan, David, Simon Denman, Clinton Fookes, and Sridha Sridharan. "Scene Invariant Multi Camera Crowd Counting." *Pattern Recognition Letters.* 44 (2014): 98–112.

[7] Ryan, David. 2013. *Crowd Monitoring Using Computer Vision*. Unpublished PhD Thesis. Queensland University of Technology.

[8] Kong, Dan, Doug Gray, and Hai Tao. 2005. "Counting pedestrians in crowds using viewpoint invariant training." In *British Machine Vision Conference Proceedings*, 63.1–63.10. BMVA Press.

[9] Kong, Dan, Doug Gray, and Hai Tao. 2006. "A viewpoint invariant approach for crowd counting." ICPR 2006. *18th International Conference on Pattern Recognition*, volume 3, pages 1187–90.