CROWD VOLUME ESTIMATION USING PHOTOGRAMMETRIC TECHNIQUES

H. Attya, **A. Habib**, **I. Detchev**, **A. Rawabdeh**
Geomatics Engineering Department, University of Calgary
2500 University Dr. NW, T2N 1N4
htattya, ahabib, i.detchev, amalrawa@ucalgary.ca

ABSTRACT

Incidents caused by human crowds could occur in various venues and under different circumstances, the most common being sport events, festivals, and religious events. Starting as early as the nineteenth century, research efforts have been geared towards crowd behaviour monitoring strategies especially in the fields of emergency and safety management. Attempts have been made also as recent as the mid and late nineties of the last century to use computer graphics in crowd modeling and simulation. Although crowd simulation is widely applied in several fields, research related to the derivation of quantitative crowd information is quite limited and is mainly focused on crowd volume using image processing and computer vision techniques. There were some attempts in the last two decades of the twentieth century to employ real time close-range photogrammetry in pedestrian detection and counting. However, these methods are hardly being used in high-density crowd monitoring because of the extreme difficulty in individual detection and tracking under these conditions. This paper provides a conceptual framework for the utilization of close-range photogrammetry to estimate crowd volume using a low-cost digital camera. The framework starts by developing a three dimensional model of the site in question prior to its observation in the presence of a crowd. This model will be used later to geo-reference the collected images from a dynamic camera system. The 3D model together with the geo-referencing parameters of the collected imagery will be finally used to derive crowd volume parameters. Preliminary results of the developed system will be illustrated together with the plans for the implementation of the proposed framework.

KEYWORDS: Crowd volume, close-range photogrammetry, low-cost camera, geo-referencing, 3D models

INTRODUCTION

Human crowd could be defined as the assembly of many people in a common place and sharing the same attention and goals, and interacting with each other in different ways (Macionis, 2005). Starting as early as the nineteenth century, human crowd studies caught the attention of researchers not only in the fields of emergency and safety but also in psychology and sociology. This interest is mainly geared toward understanding the behavior of the crowd in order to ensure safety and avoid tragedies, which could happen during the assembly of thousands of people in one place. Religious venues, sports events, and festivals are sometimes ideal places for tripping tragedies. Tragic examples happened in several places in the world such as in Iraq where more than 1000 people were killed and other 2000 were injured in tripping during a religious event in Baghdad in 2005. In November 2010, in Cambodia, more than 300 people were killed and hundreds were injured in a stampede during a water festival.

Recently, the interest in the crowd studies extends to include wider spectrum of disciplines such as urban planning, environmental engineering, and public health. However, the goal is unified toward providing – as best as possible – comfortable conditions for the people involved in the crowd in terms of place, transportation, and public health.

Image processing and computer vision techniques are widely applied in the fields of crowd simulation and volume estimation. There were some attempts also to apply real time close-range photogrammetry for pedestrian body detection and counting in non-crowded venues. The situation will be different in crowded scenes because of the difficulties in individual detection and tracking. A conceptual framework is proposed in this paper to apply close–range photogrammetry in the estimation of the crowd parameters using a low-cost digital camera.

This paper consists of five sections including this brief introduction which is followed by a concise explanation of the computer vision techniques utilized in the crowd volume estimation. The proposed approach is explained in the third section, while the fourth and fifth sections include the experimental results and conclusions.

COMPUTER VISION TECHNIQUES

Crowd control and management is of major interest especially for agencies which deal with community safety and security. Therefore, significant research in image processing and computer vision was geared to computer aided surveillance and automation of crowd analysis process to assist rapid and accurate reaction whenever needed. However, applications of computer vision techniques in crowd simulation are as recent as mid and late nineties of the last century (Thalmann & Musse, 2007). Jacques Jr. et al. in their survey for the computer vision techniques in crowd analysis provided three main categories for crowd estimation (Jacques Jr., 2010):

- **Pixel-Level Analysis**
  This analysis is used to estimate the density of people in an image. Individual pixels are analysed by background subtraction or edge detection. This approach is challenged by errors due to the perspective projection of the camera and the strong occlusion, which usually happens in dense crowds.
- **Texture Analysis**
  Instead of individual pixel analysis, this approach adopts modeling of textures in different images to estimate the number of people. Variations in people clothing, pose, illumination, scale, and resolution make the use of this approach quite challenging (Arandjelović, 2008).
- **Object-Level Analysis**
  In this approach, the identification of individual people in the image is attempted for direct people counting. However, this approach could be possible in situation of low density or non-crowded scenes where no occlusion is affecting the analysis.

A detailed discussion of the different computer vision techniques used in crowd analysis is provided by Jacques Jr. et al. (2010), and Zhan et al. (2008). All of those techniques are aiming at extracting the crowd volume either in terms of people density or direct people counting using image parameters. Although these techniques are automated and therefore beneficial in certain applications such as security surveillance, they have common problems such as clutter, occlusion, camera perspective geometry, image resolution, and differences in illumination and people clothing.

PROPOSED APPROACH

The proposed approach assumes that photogrammetry could be a powerful tool in the subject of crowd volume estimation through reconstruction of three dimensional model of the object space, which will overcome the problem of the perspective geometry of the camera. The approach has three steps:

- **3D Model Reconstruction**
  In this step, a 3D model is developed for the site in question. Bundle adjustment procedure is implemented to come up with the model. This step should be established prior to the crowd event in order to ensure best inter-visibility between the camera and the objects. This model will be used later to provide the necessary control for geo-referencing the images captured during the event, i.e. the images with the crowd.
- **Geo-referencing of Crowd Images**
  Several images could be captured during the event. These images could be geo-referenced using control from the previously derived 3D model through single photo resection procedure. After geo-referencing, the crowd images could be used to extract the three-dimensional geometrical dimensions of the crowd which help in calculating the crowd volume.
- **Calculating the Crowd Volume**
  Crowd volume could be calculated from the geo-referenced crowd images by monoplotting a polygon surrounding the crowd clusters in the images. From the coordinates of the polygons' vertices, the areas of the polygons could be calculated. Then, in order to calculate the crowd volume, rational crowd density within each polygon could be assumed. In this work, we approached the density assumptions through exploring the concept of "proxemics" which was developed by Edward Hall in 1959. Proxemics is an invisible comfort area surrounding an individual's body (Hall, 1959). In his book "The Silence Language," Hall argued that individuals are trying to maintain comfort zones during their communications with others. These zones could be as small as 0.5 meters or as large as more than 3.5 meters. The concept of proxemics has been used before in crowd analysis research such as the research of Jacques Jr. et al., 2007).

EXPERIMENTAL RESULTS

A dynamic camera system has been used to capture the required images to reconstruct the 3D model of the Stampede site in Calgary and then to capture several images for the crowd event in the site. Thousands of people gather in this place annually to celebrate the western heritage of the city. In order to reconstruct the 3D model of the site, a multi-sensor advanced triangulation program (Habib et al., 2007) has been used to run a bundle adjustment procedure on the images shown in Figure (1), which were captured prior the crowd event.

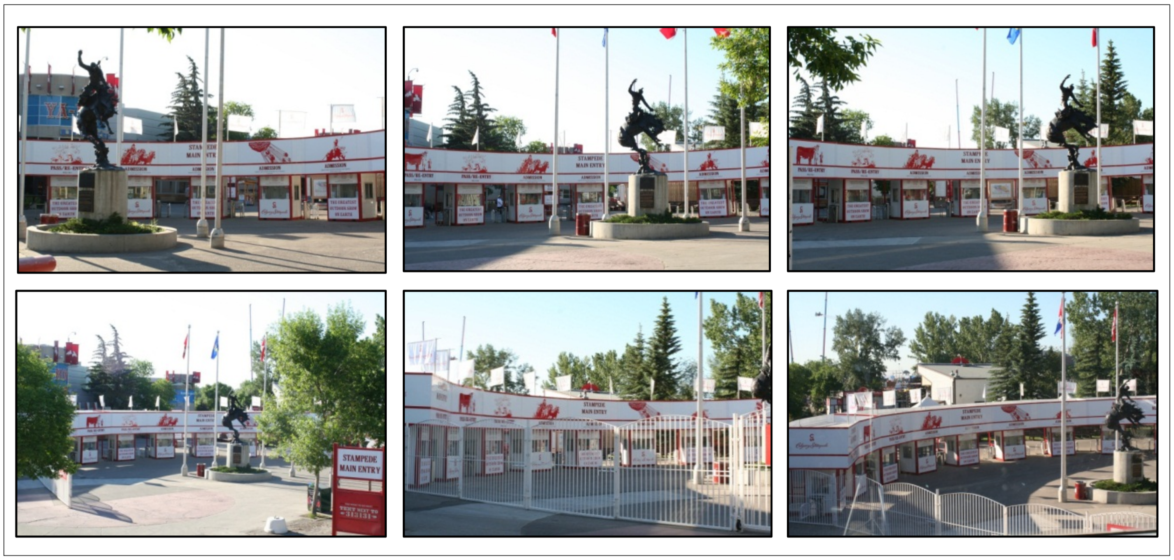

Figure 1. The images used to reconstruct the 3D model of the Stampede site in Calgary.

Captured images during the crowd event show that people are clustered around the entry and exit gates. Therefore, we draw polygons around each cluster in the image to achieve best accuracy as shown in Figures (2) and (3). After geo-referincing the images shown in Figures (2) and (3) using the parameters of the afortmentioned 3D model of the site, the areas of the drawn polygons have been calculated using a monoplotting procedure. Using the areas of those polygons in computations will certainly be more accurate than taking the total area of the site into consideration. Three different scenarios for crowd density have been tested in the experiment. Keeping the proxemic concept in mind, we chose as minimum as 0.2 meters as a distance to be maintained between individuals in the crowd. We assumed that this is the minimum distance an andividual can maintain between himself and others in order to avoid being crushed. Because usually during the incidents of

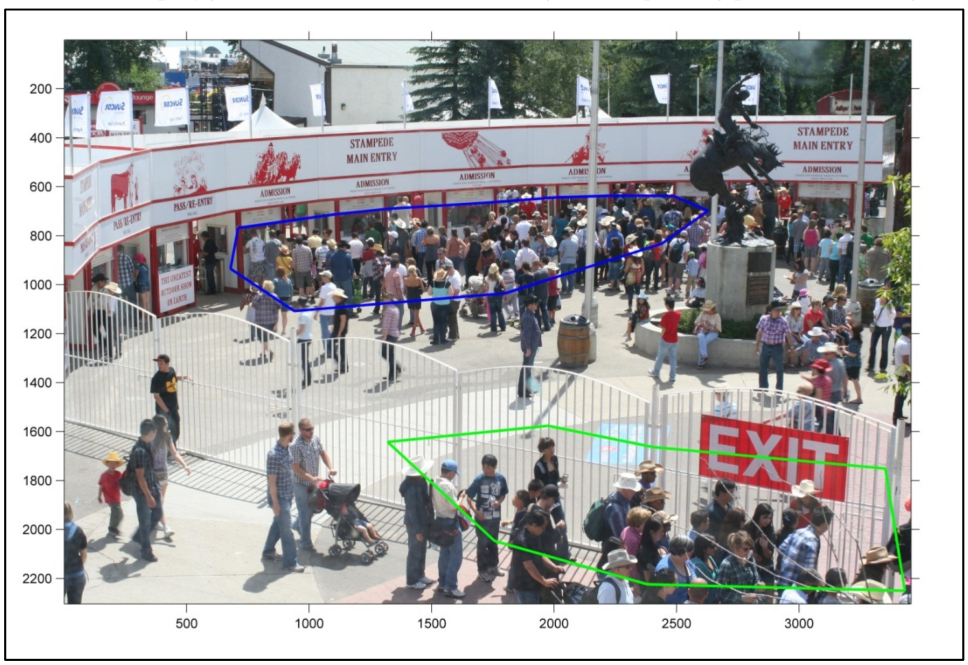

Figure 2. First image tested in the experiment.

dense crowd, people cannot maintain their comfort zones. Then, 0.5 meters has been selected as it is the minimum proxemic distance suggested by Hall, which is abviously very diffecult to be maintained during a dense crowd. We also chose to test 1.2 meters as the maximum distance in the experment because it is the social proxemic distance, which we believe that it should be very easy for individuals to maintain their personal comfort in non-crowded situations.

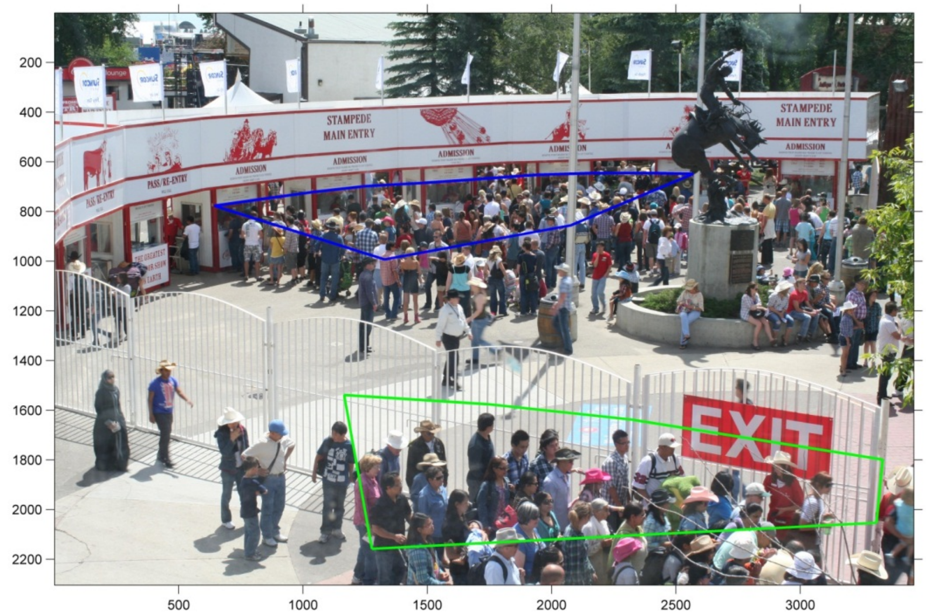

Figure (3). Second image tested in the experiment.

Table (1) shows the computation results of the two images in the experiment shown in Figures (2) & (3). In order to analyze the accuracy of the approach, we did a manual counting of people from the images. The results show that in the case of the studied crowd and by comparing the computed crowd and the counted crowd, the adopted distance between individuals could be around 0.5 meters. This is giving an approximate density of 1.3 persons per square meter, which seems to be reasonable in the studied case of the crowded site at that time. The results show better accuracy in the small polygons because of the nearly uniform clusters of people with no or very few gaps. However, the accuracy in the larger polygons is still relatively good with a range of around 70%.

Table 1. Result analysis showing the computed vs. counted crowd volume

Image	Polygon	Area (m^2)	Distance (m)	Density (person/m^2)	Computed Crowd (persons)	Counted Crowd (persons)
Figure (2)	1	106	0.2	8	844	110
			0.5	1.3	135	
			1.2	0.22	24	
	2	19.7	0.2	8	157	40
			0.5	1.3	25	
			1.2	0.22	4	
Figure (3)	1	132.7	0.2	8	1056	105
			0.5	1.3	169	
			1.2	0.22	29	
	2	16.3	0.2	8	123	33
			0.5	1.3	21	
			1.2	0.22	4	

CONCLUSIONS

Our work shows that there is a strong possibility for the photogrammetric techniques to be used in crowd volume estimation, especially when integrated with image processing and computer vision methods. The suggested approach could be more efficient in the cases of the dense crowds where no or very few gaps between people exist. However, the accuracy of the approach could be better enhanced by drawing smaller polygons to decrease the possibility of gaps between individuals. Because of its geometric basis and simplicity when compared with computer vision methods, this approach could be more interesting in the urban planning fields where comfortable space is the main concern.

ACKNOWLEDGEMENTS

This work was supported by the Canadian GEOID NCE Network (IV-17) and the National Science and Engineering Council of Canada (Discovery Grant).

REFERENCES

Arandjelović, O., 2008. Crowd Detection from Still Images**.** *Electronic Proceedings of the 19th British Machine Vision Conference*, 1-4 Sep., URL: http://www.comp.leeds.ac.uk/bmvc2008/proceedings/index.html, Leeds (last date accessed 9 January 2012).

Habib, A., Shin, S., Kim, K., Kim, C., Bang, K., Kim, E., Lee, D., 2007. Comprehensive Analysis of Sensor Modeling Alternatives for High Resolution Imaging Satellites. *Photogrammetric Engineering and Remote Sensing*, 2007, 73 (11):1241-1251.

Hall, Edward T., 1959. *The Silent Language*, Doubleday.

Macionis, John J., 2005. *Sociology*, Prentice Hall.

Jacques Jr., J. C. S., S. R. Musse, and C. R. Jung, 2010. Crowd Analysis Using Computer Vision Techniques, *IEEE Signal Processing Magazine*, 27(5), 66-77.

Jacques Jr., J. C. S., A. Braun, J. Soldera, S. R. Musse, and C. R. Jung, 2007, Understanding People Motion in Video Sequences Using Voronoi Diagrams, *Pattern Analysis Application*, 10(4), 321–332.

Thalmann D., S. R. Musse, 2007, *Crowd Simulation*, Springer.

Zhan B., D. N. Monekosso, P. Remagnino, S. A. Velastin, and Xu L, 2008, Crowd Analysis: A Survey, *Machine Vision and Applications,* (2008) 19:345–357, Springer-Verlag.