ATLANTIS PRESS

Advances in Engineering Research, volume 208

Proceedings of the International Conference on Innovation in Science and Technology (ICIST 2020)

Threshold Determination for BIM User Image Segmentation Using Fuzzy C-Means for Development of Adaptive BIM

Stefanus Santosa*
Politeknik Negeri Semarang
Semarang, Indonesia
*stefanus.santosa@polines.ac.id

Sudirman S. Pana
Universitas Ichsan Gorontalo
Gorontalo, Indonesia

De R. M. Setiadi
Dian Nuswantoro University
Semarang, Indonesia

Aryo Satito, Anwar S. Ardjo
Politeknik Negeri Semarang
Semarang, Indonesia

Yonathan P. Santosa
Soegijapranata Catholic University
Semarang, Indonesia

***Abstract*—The Building Information Modeling (BIM) Repository is required to adapt to its users who access BIM objects independently through the Common Data Environment (CDE). BIM objects are not always easy for all users to understand and misperceptions may occur due to users' varying abilities and learning styles. The solution can be done through the Intelligent Monitoring System (IMS) based learning technology to perform a visual analysis of BIM users. One important issue in the visual tracking system is the degradation of the model caused by inaccuracies in determining the segmentation threshold between the user's foreground and background images. Segmentation is still difficult when performed on complex images that have a lot of noise, inhomogeneity intensity, textures, or multi-phase structures. The focus of this research is to determine the threshold value using the Fuzzy C-Mean (FCM) approach which is compared with the performance of the Otsu method. Results showed that FCM has a smaller error rate than the Otsu method, 1.15E + 02 perframe compared to 5.11E + 02. FCM processing time is longer than Otsu, 3.6057 units of time perframe compared to 0.0331. We hope that this research can be used to development of an adaptive BIM Repository.**

Keywords*—image processing, computer vision, thresholding, Fuzzy C-Mean, otsu, computational education, building information modeling

I. Introduction

BIM is a representation of the physical and functional characteristics of all building elements in digital form [1]. This digital form not only 3 dimensional (3D) and 4D but 7D and xD [2]. It starts from Planning, Construction, to Demolition for Re-Engineering/ Ending of building.

BIM consists of aspects of management, learning, and mutual understanding of objects stored in the Common Data Environment (CDE) between users which includes conceptual processes, implementation, and maintenance. CDE which requires collaboration between organizations or construction industries can take place if each of them already has a BIM object library in an open access environment. Common frameworks used to create the documents and tools necessary to support an industry-wide: common understanding, common data exchange, common ways of working, basis for consistent up-skilling, training and education [3].

Many needs can be provided by BIM starting from data or information which in the area of briefing, regulation, databases, CAD software, visualization, modeling, simulation, specification, procurement, construction management, and building demolishment. BIM forces a person to be able to understand data or visual objects in the BIM open access network independently because they are not accompanied by the data or object maker. User capabilities vary widely, even though objects do not vary. So, BIM objects are not always easy for all users to understand and misperceptions may occur due to user’s varying abilities and learning styles.

BIM users consists of technician, construction and building manager, designer, overseer, workers, public, and government. Visual model understanding of BIM which explains stage of building and maintenance of a building is included in online learning (elearning or etutoring) which is nonformal (etraining) or informal (field explaination). The development of facilities can be done by using learning technology which based on monitoring technique and user control by using image processing and machine learning approaches.

Adaptive etutoring or etraining system for BIM which can interact naturally must have abilities to understand the user's behaviors. It must be capable to manage learning path with monitoring function to interpret and to conclude user's needs based on their behaviors [4,5]. It is known that user's activity such as movements and facial expression can shows level of concentration in understanding BIM object which exist in the repository. This kind of understanding or learning process needs a smart learning monitoring system.

In Intelligent Monitoring System (IMS), the key factor is detection and tracking while following construction or maintenance instructions or self-studying using BIM network. IMS can be used to do visual analysis for human activity [6,7]. One of important issue in visual tracking system is the model degradation which caused by inaccuracies between foreground and background image from the user's webcam or gadget. This issue is seldom talked about. A relevant method for solving this problem is active contour and image segmentation approaches [8].

II. Related Research

Image segmentation is one of the most basic and important task in image processing. Segmentation is used to divide image into each own partition which contains homogenic pixels which clustered based on pixel value characteristics, such as color intensity and texture. Many segmentation methods has been proposed, but segmentation tasks is still hard when applied to an image which complex with many noise, intensity inhomogeneity, texture, or multiphase structure [9,10].

Srinivas used Fuzzy C-Means approach for deciding initial threshold. The result shows that FCM is better compared to Otsu approach and previous methods. On the previous method, initial threshold was done iteratively by using mean of background pixel and foreground pixel which obtained by using random initial threshold [11].

Yang explained that to choose threshold value in Otsu method, intensity value of gray level image needs to be divided into two parts as the pixel values of foreground and the pixel values of background [12]. Then choose Occurrence Probability (ω) and Mean (μ) for each part. This values then can be used to calculate between-class variance by using this equation: $\sigma_B^2 = \omega_0(\mu_0 - \mu_T)^2 + \omega_1(\mu_1 - \mu_T)^2$, afterwards the threshold value can be obtained by finding maximum value from between-class variance. In this Otsu method, Yang improvised by changing mean and median parameters. The results show a high accuracy in threshold selection [12].

In regards to threshold for matting Basuki did an experiment for alpha matting by using Fuzzy C-Means for choosing adaptive threshold on the image condition. The purpose is to overcome trimap approach in matting process which for image with random pixel composition and object with many holes. In this experiment, Fuzzy C-Means has better MSE compared to Otsu approach [13].

Chen did a research in alpha matting which based on Nonlocal principal to extract simultaneously on multilayer image [14]. The implementation of Nonlocal principal is by using K-Nearest Neighbors (KNN) which finds neighbor similarities. This approach used Nonlocal principal to estimate alpha value and layer extraction for layer that overlaps by using Laplacian framework. The alpha value can do a good generalization for solving multilayer extraction problem [14].

Zheng in his research proposed two kinds of approaches in choosing alpha matte value which based on semi-supervised learning [15]. This approach includes local-based learning and global-based learning. In the process, training was done by using manifold structure of the image, for example surface intensity image. For global-based learning, alpha value estimated by choosing every pixel which is unknown into two parts from foreground pixel and background pixel which contains label. Both approaches considered easy to implement because only simple matrix operations are used [15].

Based on those state of the art above, then the selection of threshold value for alpha matting in this research will use Fuzzy C-Means clustering approach. This research is an early research which further research will be conducted to identify facial expression that includes eyes, mouth, brows, forehead, and gesture which will improve Adaptive BIM Repository.

III. Research Methods

A. Data Collection

The data that will be used for this experiment is video data that contains human interacting with elearning or etutoring. The location of the video data collections was done indoor with different lighting condition, wall texture, and different objects.

B. Object Identification

The experiment in this research is initialized by reading video data frame by frame. The starting frame is used for background initialization. Afterwards the absolute difference between starting frame and the next frame are calculated and used as the estimate foreground frame which subtracts between the frame that was initialized as the background with the current frame. Mathematically written as:

$$I_{diff} = \left| frame_{bgr} - frame_{curr} \right| \quad (1)$$

This differential operation is for identifying object existence in the current frame. The result is differential frame in RGB format. For choosing threshold value, the image was converted into pixel gray scale. Then clustered by FCM and Otsu approach as a comparison.

C. Experiment

This experiment used alpha matting approach. The process of alpha matting is initialized with initial alpha mate which is a continuous value in range of 0 until 1 which obtained by using binarization using the previously obtained threshold value from the resulting frame of absolute difference operation. This binarization process will results a binary level image. This process is done by doing selection of the pixel value of the gray level image using the following formula:

$$g(x,y) = \begin{cases} 1 & IF\ f(x,y) \geq T \\ 0 & IF\ f(x,y) < T \end{cases} \quad (2)$$

The alpha matte value will be used as the base to create foreground image (object) and background image from every frame in the video dataset. The segmentation is based on the following formula.

$$I = \alpha F + (1 - \alpha) B \quad (3)$$

Formula above explains that an image is a combination of foreground and background, with alpha value (α) is the opacity of a pixel used for combining foreground and background of an image. Thus, to get the foreground component, one can do a linear multiplication for every pixel with its alpha mate and to get the background component every pixel can be multiplied with the invers of the alpha matte. The formula to get the foreground and the background are described as follows.

$$Foreground = \alpha\ I \quad (4)$$

$$Background = (1 - \alpha)\ I \quad (5)$$

D. Evaluation

Evaluation of FCM approach was done with the process time and error rate of the segmentation. The results then will be compared to Otsu approach. The process time calculation in choosing threshold value was done with tic and toc function. The process results the time needed to execute all commands after tic function is called until toc function is called. The error rate measurement was done by calculating the MSE of the segmented object from the program output and compared with the manual segmentation using image editor software.

IV. Results and Discussion

A. Data Collection Results

Data collection from video collection of user image of etutoring or elearning is a video with pixel size of 640 x 480 and 320 x 240 with AVI file format. The dataset consists of 5 video files (Video A, B, C, D, and E) with different object which named with code A0xxx until E0xxx (Table 1).

As many as 30 frames from video was collected that the object position is always on the front of camera which used various distance but still in normal range when the subject is interacting with etutoring or elearning. Different room condition caused different lighting intensity when the video was taken. Background has different size between videos, so as the foreground.

TABLE I. Example of Data Collection

Video A	
Video B	
Video E	

The video duration ± 60 seconds. While the frame used as evaluation sample was taken randomly from video file. Based on the Co-Occurrence Matrix analysis, each video files has different intensity. It is shown with different contrast from each video data. Video with highest contrast are video A and B, while the lowest contrast is video C. The contrast difference was the results of different light intensity of the room when the video was taken. Frame sample which used as the experiment sample was used from the first frame and the following frame was taken randomly.

B. Object Identification Results

Absolute difference between first frame and the next frame which estimated as the foreground frame was running smoothly. When there are pixel changes in the pixel location which shown as changes in pixel values more than 0, will be considered as foreground. Pixel with 0 values (black) is the condition where there are no changes of pixel value or no object present (Table 2). This will make it easier for the following process to detect the object.

TABLE II. EXAMPLE OF DIFFERENTIAL FRAME

Frame	A0321
Background Frame	
Current Frame	
Differential Frame	

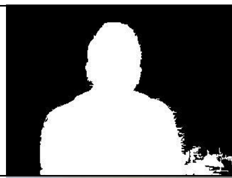
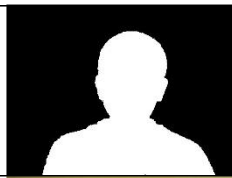
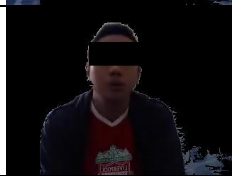
Even though this approach is lacking when the changed pixel is not an object. For instance, when the room lighting is changed or something which not the object was moving which increases the error. This process was done without scribble, but it will be difficult to do a scribble because the object is moving. The absolute difference as the differential frame was in RGB format. Before experiments, the image was converted to gray scale.

C. Experiment Results and Discussion

Choosing the threshold value was done based on every differential frame gray scale which retrieved from the previous stage using FCM. The threshold value obtained with FCM shows varying values for every frame compared with Otsu. This will impact greatly in the matting process and the accuracy of the result. The average threshold value for FCM is 0.1669 while Otsu is 0.9166.

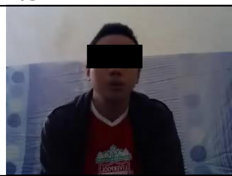
Based on the threshold with FCM approach and alpha matte calculation until obtained the matte image as shown in table 3. This image then used as the base to get the foreground (user object) and background image from every frame.

TABLE III. SEGMENTATION RESULT BY FUZZY C-MEAN THRESHOLD

Frame	A0321	E0307
Original		
Alpha Matte		
Background		
Foreground		

Threshold value from FCM is smaller than Otsu. Matte image is quite good thus the foreground and background appeared as in table 3. But these thresholds cannot solve pixel color because of reflections, thus many error happens in the face and hair area which many light reflections from and on the background which get reflections from the object.

TABLE IV. SEGMENTATION RESULT BY OTSU THRESHOLD

Frame	A0321	E0307
Original		
Alpha Matte		
Background		
Foreground		

As comparison with Otsu approach in table 4, shows segmentation with many pixels error on the object. This is because the threshold value with Otsu was too high, thus the pixel color on the object that close to black will be considered as background in the binarization process. The pixel value from the object to the background goes through color degradation. This degradation was caused by the low resolution of the image which produced by the camera. Compared with FCM the pixel error mostly happens in the object area was caused by the low pixel value when absolute difference process in the object area has high intensity light. The results with FCM and Otsu shows that the threshold will impact the alpha matte quality.

D. Evaluation Result and Discussion

Error rate between two approach was calculated with MSE. The evaluation of the FCM results shows a better result. FCM has lower error compared to Otsu. The MSE value from FCM was 3.46E+03 with average of 1.15E+02 for every frame while Otsu MSE value was 1.53E+03 with average of 5.11E+02 for every frame.

The time needed to find the threshold for FCM was 108.1721 time unit based on tic and toc function with the average of 3.6067 time unit. While Otsu needs faster time, which is 0.9930 time units with the average of 0.0331 time units per frame.

From this experiment, choosing threshold with FCM has advantages compared to Otsu. The results of the segmentation were better compared to Otsu, but the processing time is slower than Otsu. This is because FCM is an unsupervised algorithm. FCM decides optimum cluster center from the sample pixel by calculating Euclidean distance, the fuzzy cluster degree, and membership category which cannot be chosen. This process was done iteratively to choose the correct threshold value hence the longer processing time but more accurate.

The use of IMS as the adaptive BIM then it is needed as a tracking process for the BIM repository users. The object tracking was done by using regionprops function from tools which results a bounding box from object which segmented. The segmented foreground will impact the accuracy to do tracking to the object in a video. This process was done by giving boundary to the outer pixel of the object.

V. Conclusion

From this experiment, choosing threshold with FCM has advantages compared to Otsu. The results of the segmentation were better compared to Otsu, but the processing time is slower than Otsu. The MSE value from FCM was 3.46E+03 with average of 1.15E+02 for every frame while Otsu MSE value was 1.53E+03 with average of 5.11E+02 for every frame. The time needed to find the threshold for FCM was 108.1721 time unit based on tic and toc function with the average of 3.6067 time unit. While Otsu needs faster time, which is 0.9930 time units with the average of 0.0331 time units per frame.

Acknowledgment

The authors would like to thank the support from the Research Center and Community Service for the research under Grant SP DIPA-023.18.2.677603/2020 from The Featured Research Areas of Politeknik Negeri Semarang, Ministry of Education and Culture in Indonesia.

References

[1] C.M. Eastman, C. Eastman, P. Teicholz, R. Sacks, and K. Liston, BIM handbook: A guide to building information modeling for owners, managers, designers, engineers and contractors. New Jersey: John Wiley & Sons, 2011.

[2] B. Hardin and D. McCool, BIM and Construction Management: Proven Tools, Methods, and Workflows. 2nd ed. Indianapolis, Indiana: John Wiley & Sons, Inc., 2015.

[3] E. Bieńkowska, Handbook for the Introduction of Building Information Modelling by the European Public Sector. EU BIM Task Group, 2018.

[4] K. Chrysafiadi and M. Virvou, "Student modeling approaches: A literature review for the last decade," Expert Systems with Applications, vol. 40, no. (11), pp. 4715-4729, 2013.

[5] M.B. Ammar, M. Neji, A.M. Alimi, and G. Gouardères, "The affective tutoring system," Expert Systems with Applications, vol. 37, no. (4), pp. 3013-3023, 2010.

[6] M. Huang, G. Chen, G.F. Yang, and R. Cao, "An algorithm of the target detection and tracking of the video," Procedia Engineering, vol. 29, pp. 2567-2571, 2012.

[7] R. Zhang and J. Ding, "Object tracking and detecting based on adaptive background subtraction," Procedia Engineering, vol. 29, pp. 1351-1355, 2012.

[8] J. Fan, X. Shen, and Y. Wu, "Scribble tracker: a matting-based approach for robust tracking," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, no. (8), pp. 1633-1644, 2011.

[9] H.P. Narkhede, "Review of image segmentation techniques," International Journal of Science and Modern Engineering, vol. 1, no. (8), pp. 54-61, 2013.

[10] X.F. Wang, H. Min, L. Zou, and Y.G. Zhang, "A novel level set method for image segmentation by incorporating local statistical analysis and global similarity measurement," Pattern Recognition, vol. 48, no. (1), pp. 189-204, 2015.

[11] K. Srinivas and V. Srikanth, "Automatic Histogram Threshold with Fuzzy Measures using C-means," IJCSI International Journal of Computer Science Issues, vol. 8, no. (1), pp. 347-351, 2011.

[12] X. Yang, X. Shen, J. Long, and H. Chen, "An improved median-based Otsu image thresholding algorithm," Aasri Procedia, vol. 3, pp. 468-473, 2012.

[13] R.S. Basuki, M. Hariadi, and R.A. Pramunendar, "Fuzzy C-Means Algorithm for Adaptive Threshold on Alpha Matting," Proceedings of Computer Society Coference on Information Technology and Electrical Engineering, no. July, pp. 177-180, 2012.

[14] Q. Chen, D. Li, and C.K. Tang, "KNN matting," IEEE transactions on pattern analysis and machine intelligence, vol. 35, no. (9), pp. 2175-2188, 2013.

[15] Y. Zheng and C. Kambhamettu, "Learning based digital matting," In 2009 IEEE 12th international conference on computer vision, pp. 889-896, 2009. IEEE.