Data annotation system for intelligent energy conservator in smart building

Bens Pardamean1,4, Tjeng Wawan Cenggoro2,4, Bloomest Jansen Chandra3, Reza Rahutomo4

1 Computer Science Department, BINUS Graduate Program - Master of Computer Science Program, Bina Nusantara University, Jakarta, Indonesia 11480
2 Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480
3 Computer Science Department, College of Letters and Science, University of Wisconsin, 53706 Madison, WI, USA
4 Bioinformatics and Data Science Research Center, Bina Nusantara University, Jakarta, Indonesia 11480

Email: wcenggoro@binus.edu

Abstract. The concept of smart building includes the optimization of energy usage in a building. One of the possible solutions for this is to adaptively adjust appliances utilization according to activity level in the building. Thus, an intelligent activity estimation system needs to be developed. However, massive annotated dataset is necessary to train the system. Therefore, we propose a system that enables rapid data annotation to collect the massive dataset. With the proposed system, an image can be annotated within 4.8 seconds in average.

1. Introduction

Smart building is an extension of smart city in a smaller scale, that is within a single building. The adaptation of smart city concept as smart building is natural as many of the application can directly be implemented for the case of building management [1]. One of the cases is automatic monitoring of activity level detection in a room. The application of this solution can optimize the use of energy usage by the electrical devices in the room.

The most practical implementation of the activity level detection system is using CCTV as the input sensor. The utilization of CCTV avoids additional cost to procure more sensors, as CCTV is usually already available in the building rooms. However, this strategy presents another challenge, that is to build an intelligent system that can detect activity level from the CCTV video.

Developing such an intelligent system nowadays is not impossible with the recent advancement of Artificial Intelligence (AI) in computer vision researches. With the utilization of deep learning, many previously difficult problems in computer vision have been solved, for instance, image classification [2–6] and object detection [7–10]. However, a massive scale of dataset is needed for deep learning to deliver a well enough performance to be implemented in real environment.

Retrieving this massive dataset is not a simple task. One of the hurdles that needs to be addressed is how to annotate the data with such scale in a quick time. Without a proper strategy, this annotation process will take years to be done. The challenge of big data annotation is usually addressed by using
crowdsourcing strategy [11–13] combined with a properly designed information system that facilitates the annotation process [14, 15]. One of the design aspects in the annotation system that needs to be considered is the user interface. A proper interface design is crucial to cut the time needed for overall annotation process. In this paper, we present a system with a user interface design that enables rapid data annotation for smart building system. It is specifically designed to annotate data for training intelligent activity level estimator system.

2. Related Works
With the requirement of huge training data for deep learning, researches in collecting such data is inseparable from the actual development of the algorithm. For instance, WalmartLabs developed Chimera [14], an information system that generates annotation by combining crowd sourcing strategy with classification models. Chimera helped WalmartLabs to accomplish a large-scale classification task.

On the other hand, Cenggoro et al. [15] noticed that object counting datasets previously did not use crowd sourcing strategy in collecting data annotation. As a result, datasets in object counting tends to be much smaller compared to datasets in other computer vision tasks [12, 13, 16]. Therefore, to further progress researches in object counting with deep learning, they proposed an information system that is capable in gathering data annotation for object counting with crowd sourcing strategy.

3. Development Workflow for Intelligent Activity Level Estimator
For a better understanding in the role of the proposed annotation system, we provide an illustration of the overall processes to develop an intelligent system for room activity level detection in figure 1. The development comprises six separate processes that cover starting from the data collection tasks to the deployment of the intelligent system.

The proposed system is implemented as part of the annotation system in process 2 in the figure 1. The mechanism of this annotation system can be elaborated by the use case diagram in figure 2. This system is used by three entities: the user that provides annotation, the CCTV that supplies the raw video data, and the researcher that acquire the final dataset from the system. The user interface, which is the core of the quick annotation, serves only the user for data annotation process in “Annotate video frames” case in figure 2.

![Figure 1: Intelligent system development workflow for room activity level detection](image-url)
4. User Interface Design

As depicted in figure 1, the source data are streamed as video files from CCTV. Afterward, the system is expected to provide annotations in the pair of image and people count. Therefore, we identify four challenges that needs to be solved for a rapid data annotation in our case:

1. The annotation system should ease users in navigating within video;
2. The annotation should be saved in a format that allows researcher to extract images from annotated frames in the video;
3. Users should be able to identify which video needs to be annotated next. Afterwards, they should be able to quickly navigate to that video;
4. Users should be able to track their previous annotations at glance and delete incorrect annotations.

To solve those four challenges, the system utilizes a user interface as depicted in figure 3. This interface is designed with 4 main panels. The top left panel allows users to interactively navigate within a video with a simple video streaming interface. Users can stop the video at this panel at the frame they want to annotate, then provide the annotation in top right panel.
Figure 4: Implementation of the user interface design

Meanwhile, the bottom left panel enables users to see which video needs to be annotated. This panel provides list of videos along with their duration and total annotations that have been provided. With this information, users can easily detect if a video have not been annotated or the annotations are still insufficient. To navigate to the video, users can simply click at the filename in the list. At the same time, users can see the list of annotations they have provided in the selected video at bottom right panel. By clicking at a single row in the list, users can navigate to the corresponding frame that they have annotated. With this, users can quickly check whether the previous annotation is correct or not. If it is incorrect, users can simply click a delete link at the row corresponds to the annotation in bottom right panel. The implementation of the proposed design is provided in figure 4.

5. Results and Discussion

After implemented, we test the system to a user. This test is carried out within 2 hours. The user is asked to annotate 34 videos within that interval. He annotated all 34 videos within 1 hour 44 minutes and 58 seconds, resulting 896 images. Therefore, the average time needed to annotate an image is only about 4.8 seconds. The result of this test shows that our proposed system is able to successfully accelerate the annotation process. Figure 5 shows the time spend to annotate each video used in the test. The video id in the x axis is sorted chronologically. From the plot, we can see as well that the time spent tends to decrease as the user annotates more images. This suggests that the interface allows the user to quickly adapt to the system.

Additionally, this system happens to also encourages users to annotate the videos one by one. We can see from figure 5 that only video id 4 that is annotated in more than one time-interval (denoted as 4a and 4b in the plot). Other videos are annotated within only one time-interval. This behavior is favored for the speed of annotation process, as the users can focus on a video at a time.

The system also successfully encourages the user to annotate more images given longer video duration. This feature is favored for annotation process since it allows more annotation to be collected, resulting in bigger dataset as the outcome. Figure 6 shows the plot of number of annotated images versus the corresponding video duration. The figure also plots the fitted linear model on the data. The plot shows a visible correlation between number of annotated images and video duration. This fact is backed up with the linear model R-squared value of 0.5751. This correlation is also statistically significant with p-value of 7.94 x 10^-8.
Results and Discussion

Data annotation plays a vital role for an intelligent system development in smart building. For this reason, we propose a system that facilitates rapid annotation of CCTV footages for an intelligent activity level estimator. The system successfully enables quick annotation process with the average of 4.8 seconds spent per annotated images. In addition, the interface is easy for users to adapt, which also contributes to the quick process. Not to mention that the interface also encourages the users to provide more annotations given longer video duration.

For further development, the dataset collected by this research can be utilized to train an AI model for room activity level detection. The AI model can be implemented as an intelligent system as a part of integrated smart building system.

Figure 5: Plot of average seconds spend per annotation versus the corresponding video id

Figure 6: Plot of number of annotated images versus video duration
References

[1] Minoli D, Sohraby K and Occhiogrosso B 2017 IEEE Internet of Things Journal 4 269–283
[2] Krizhevsky A, Sutskever I and Hinton G E 2012 Advances in neural information processing systems pp 1097–1105
[3] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V and Rabinovich A 2015 Proceedings of the IEEE conference on computer vision and pattern recognition pp 1–9
[4] He K, Zhang X, Ren S and Sun J 2016 Proceedings of the IEEE conference on computer vision and pattern recognition pp 770–778
[5] Huang G, Liu Z, Van Der Maaten L and Weinberger K Q 2017 Proceedings of the IEEE conference on computer vision and pattern recognition pp 4700–4708
[6] Hu J, Shen L and Sun G 2018 Proceedings of the IEEE conference on computer vision and pattern recognition pp 7132–7141
[7] Redmon J, Divvala S, Girshick R and Farhadi A 2016 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pp 779–788 ISBN 978-1-4673-8851-1 ISSN 01689002
[8] Ren S, He K, Girshick R and Sun J 2017 IEEE Transactions on Pattern Analysis and Machine Intelligence 39 1137–1149 ISSN 01628828
[9] Lin T Y, Dollár P, Girshick R, He K, Hariharan B and Belongie S 2017 Proceedings of the IEEE conference on computer vision and pattern recognition pp 2117–2125
[10] Lin T Y, Goyal P, Girshick R, He K and Dollár P 2017 Proceedings of the IEEE international conference on computer vision pp 2980–2988
[11] Kovashka A, Russakovsky O, Fei-Fei L, Grauman K et al. 2016 Foundations and Trends in computer graphics and Vision 10 177–243
[12] Deng J, Dong W, Socher R, Li L J, Li K and Fei-Fei L 2009 2009 IEEE conference on computer vision and pattern recognition (Ieee) pp 248–255
[13] Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S, Huang Z, Karpathy A, Khosla A, Bernstein M et al. 2015 International journal of computer vision 115 211–252
[14] Sun C, Rampalli N, Yang F and Doan A 2014 Proceedings of the VLDB Endowment 7 1529–1540
[15] Cenggoro T W, Tanzil F, Aslamiah A H, Karuppiah E K and Pardamean B 2018 IOP Conference Series: Earth and Environmental Science vol 195 (IOP Publishing) p 012063
[16] Lin T Y, Maire M, Belongie S, Hays J, Perona P, Ramanan D, Dollár P and Zitnick C L 2014 European conference on computer vision (Springer) pp 740–755