The Implementation of Object Recognition using Deformable Part Model (DPM) with Latent SVM on Lumen Robot Friend

A Kurniawan[1], R Saputra[2], Marzuki[3], M S Febrianti[3], A S Prihatmanto[4]

1. Department of Informatics, Computer Science Faculty, Bandar Lampung University, Lampung, Indonesia. Ade.curniawan@gmail.com
2. Department of Informatics, Computer Science Faculty, Bandar Lampung University, Lampung, Indonesia. rahmatsptr96@gmail.com
3. School of Electrical Engineering and Informatics, Bandung Institute of Technology, Bandung, Indonesia.{rdmarzuki,suzanne_maria1986}@gmail.com
4. School of Electrical Engineering and Informatics, Bandung Institute of Technology, Bandung, Indonesia. asetijadi@lskk.ee.itb.ac.id

Abstract—Object recognition is part of image processing. It is used to recognize the surrounding objects based on their features, then they are processed further to obtain valid data information that can be used for other purposes. Currently, object recognition is mostly used by robot developers as one of the features in humanoid robot. One of the recent challenges occurring in humanoid robot is how the robot detects and localizes generic objects from categories such as human or car in static images using sensory visuals. It is a difficult problem since objects in such categories vary both in appearance and shape. For example, it is difficult to recognize an object that most of its shape is blocked by other objects. To solve the problem, researcher used Deformable Part Model and latent svm methods, where the data collection was performed through Library Research and Field Research approaches. The conclusion of this research is that recognition to objects using deformable part model provides a passable accuracy. After 3 experiments had been performed, system was able to recognize objects with highest reach by 88%. Keywords — Object Detection, Deformable Part Model, Object Recognition, pyramid HOG, LSVM.

1. Introduction

Deformable Part Model (DPM) is one of the feature extraction methods found by Prof. Felzenszwalb and his team in 2010. The advantage of this method is its ability to detect generic objects having various high complexities. DPM assumes that an object is constructed by parts that shape it. Therefore, the detector will first find an overall fit, and then use parts of its model to perfect the result. [4]. One of the challenges in humanoid robot is how does the robot detect and localize generic objects from categories such as human or car, in static images using sensory. It is a difficult problem since objects in such categories vary both in appearance and shape. Variations do not only come from the change in lighting and perspective, but it also due to non-grid deformation and intra-class variability in objects' shape and other visual properties. For example, someone who wears different clothes and performs various poses, someone who rides a horse so that only a few parts of his/her body are visible from the side, or someone who is blocked by objects around him/her so that only a few parts of his/her body are visible. In this paper, the researcher developed an object recognition application for the development of artificial intelligence (AI) in lumen robot friend. This application will be the basis for the development of Lumen's intelligence module in creating an AI that allows Lumen to recognize every single object around it.

2. Literature Review

2.1 Artificial Intelligence

According to Elaine Rich and Kevin Knight, Artificial Intelligence (AI) is part of computer science that enables the machine (computer) to do something as good as a human can do or even better. And what is meant by artificial intelligence is a combination of the ability to do:

a) Learning, all kinds of informal and formal learning through combination of experience, education and training. Learning shows a change in the system that allows the system to perform similar task more efficiently.
b) Pose Problems, to recognize problem situations and transform them to be more obvious.

2.2 Feature Extraxcy

Feature extraction is a process that accentuates features or characteristics of an image containing important information for classification or image data analysis. The purpose of feature extraction process is to improve the effectiveness and efficiency of analysis or classification process. Several computer vision techniques for extracting features are the contour, convex-hull and convexity defects detection techniques.

2.3 Computer Vision

Computer Vision is a field that aims to make an important decision about a real physical object and situation based on an image. Computer Vision is a combination of image processing and pattern recognition. The output of Computer Vision process is an image understanding. Development of this field is performed by adapting the ability of human vision in taking information. In the discipline, Computer Vision deals with a theory on AI that extracts information from images.

2.4 Humanoid Robot NAO

NAO is an autonomous, programmable, medium-sized humanoid robot developed by Aldebaran Robotics, a French startup company headquartered in Paris. The NAO project was launched in 2004. On August 15, Sony's dog robot, Aibo, was installed as a robot used in the RoboCup Standard Platform League (SPL), an international robotic competition. Currently, developers can use the NAOqi framework to stimulate the development process. NAOqi is a framework that has an application programming interface (API) that can be used to acquire data from either NAO sensors or NAO controlling actuators. API is now available in various programming languages such as MATLAB, Python, .NET, Java and C++.

2.5 AMQP and RabbitMQ

AMQP stands for advanced message queuing protocol or known as a protocol that controls queuing messages. AMPQ is a standard application layer protocol used by message-oriented middleware [5]. In AMQP, the delivery of messages between clients in a large quantity is possible to process. RabbitMQ is one of message-oriented middlewares implementing AMQP [6]. RabbitMQ is an open source program that can be operated on various operating systems. RabbitMQ also has application programming interface (API) that can be used for client program development. RabbitMQ API is now available in various programming languages such as C ++, Python, and .NET.

2.6 Deformable Part Model

Deformable Part Model (DPM) is one of the feature extraction methods found by Prof. Felzenszwalb and his team in 2010. The advantage of this method is its ability to detect generic objects that have various high complexities. Deformable Part Model has several steps, one of which is to build modeling that includes root and part, matching process and Mixture Model.

2.6.1 Modeling

Modeling of an object will be used as the target of detection. Here are several steps in creating a detection model are: Root is a process of capturing objects that are visible by the camera through the edge line that resembles the shape of object.

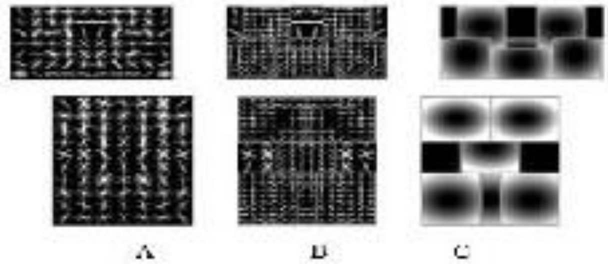

Figure 1. root filters coarse resolution, part filters finer resolution, deformation models

The process of dividing parts on the object features is performed by the following equation:

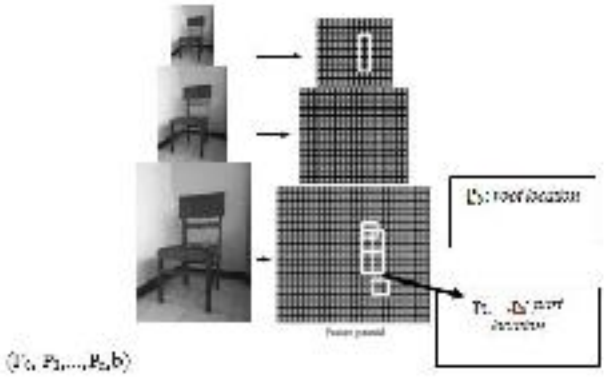

Figure 2. Pyramid Histogram of Oriented Gradient

Description:
F_0 = foot filter
P = Model for the part
b = real-valued biased term

Process is performed to calculate all root position scores according to the positioning of the possible part positions; it can be done by using the following equation:

$$score\,(\ _0,\ _0,\ _0) = \ _{0,0}(\ _0,\ _0) + \sum_{i=0}^{n} D_{i,l_0} - \lambda(Z(x_0,y_0)+v_i) + b \qquad (1)$$

Description:
$(\ _{0,0},\ _0)$ = Root Location
$\ _{0,0}$ = Model Filter in the Pyramid
$(\ _{0,0})$ = Part Position
$\sum_{i=0}^{n} D_{i,l_0} - \lambda(Z(x_0,y_0)+v_i)$ = Optimal Placements

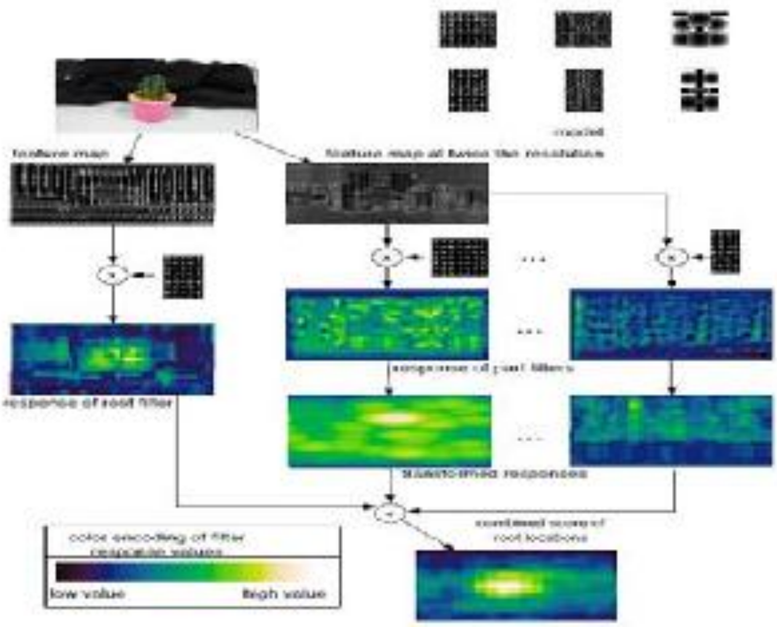

Figure 3. root filters coarse resolution, part filters finer resolution, deformation models

Object detection by *Mixture Model* uses matching algorithms to locate root positions independently on each detected *part*.

$$M = (M_1, \ldots M_m) \qquad (2)$$

Where M = the value collected based on the root location in a part

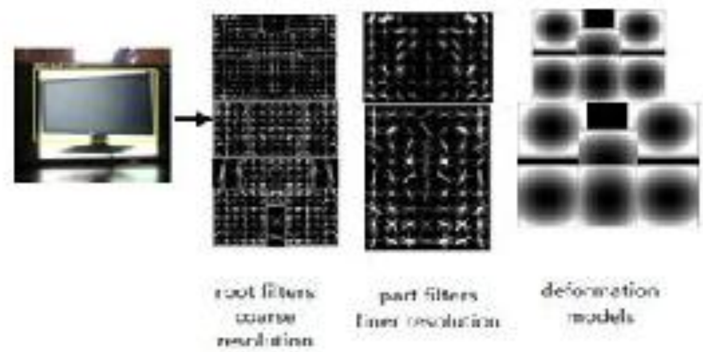

Figure 4. root filters coarse resolution, part filters finer resolution, deformation models

2.7 Latent SVM

Latent SVM or latent Support Vector Machine is a method used as a feature classifier

2.7.1 *Semiconvexity*

Semiconvexity is performed to obtain initial value on lsvm processing. It can be performed by using the following equation:

$$L_D(\beta) = \frac{1}{2}\|\beta\|^2 + C\sum_{i=1}^{n}\max\left(0, 1 - y_i f_\beta(x_i)\right) \qquad (3)$$

Where Fβ as the maximum *convex* in the *convex* function for negative samples, the function will not exist if there is a possibility of *latent* score in each positive sample: fβ is *linear*

2.7.2 Optimization

The *optimization* value can be obtained by using the following equation:

Re-label Positive sample:

Optimize *LD*(*β*, *Zp*) over *Zp*, i.e. select

i = argmax *β*.Φ(*xi* , z)

z Z(xi)

Optimize:

β: Stochastic gradient descent

2.7.3 Stochastic Gradient Descent

Stochastic Gradient Descent can be determined by using the following equatiogn:

$$L_D(\beta) = \frac{1}{2}\|\beta\|^2 - C\sum_{i=1}^{n}\max\left(0, 1 - y_i f_\beta(x_i)\right)$$

$$\nabla L_D(\beta) = \beta \quad C\sum_{i=1}^{n} h(\beta, x_i, y_i) \qquad \text{(subgradient)}$$

$$h(\beta, x_i, y_i) = \begin{cases} 0 \text{ if } y_i f_\beta(x_i) \geq 1 \\ -y_i \Phi(x_i, z_i(\beta)) \text{ otherwise} \end{cases} \qquad (4)$$

2.7.3 Data-Mining Hard Example, LSVM Version

Hard Example on LSVM can be determined by using the following equation:

Feature cache F: set of (i, v) with $1 \leq i \leq n$ and $v = \phi(x_i, z)$ with

$z \in Z(x_i)$ (one for positive examples)

$I(F)$ examples indexed by F

$$L_F(\beta) = \frac{1}{2}\|\beta\|^2 + C\sum_{i \in I(F)} \max(0, 1 - y_i(\max_{(i,v) \in F} \beta . v)) \qquad (5)$$

3. Design And Implementation

In this section, the design and implementation of the object detection system on the robot is presented.

3.1 System Architecture

The general working scheme program in this *paper* can be seen in the figure 1, From the architecture chart, it can be seen that the process of introduction to objects is performed by using the NAO robot as a tool to implement the algorithms. In the initial stage, the acquired data is in the form of a video sent to the data storage server, Network Attached Storage (NAS), using the image server service application. Subsequently, the data successfully sent to the NAS is traggered automatically through the RabbiMQ messaging server, and then it is preceded to the image processing and objects recognition stages. After the introduction to objects has been done, the final step is to send the results of recognition to NAS, so that the recognized data can be further processed for other purposes.

3.2 Lumen AI Integration

The general working scheme program in this *paper* can be seen in the following:

Lumen AI Integration is the integration stage of algorithms on lumen robot friend. Here are figures showing the results of Lumen AI Integration implementation where figure (a) shows the result of object detection by Lumen in resting position and (b) shows the result of object detection in standing position.

Figure 5. root filters coarse resolution, part filters finer resolution, deformation models

4. Test And Result

Based on the test that had been performed, it was found the following results

Table I. Precision Record.

		Real value	
		True	False
Prediction Value	True	TP (True Positive)	FP (False Positive)
	False	FN (False Negative)	FN (False Negative)

$$precision = \frac{TP}{TP + FP}$$

$$recall = \frac{TP}{TP + FN}$$

$$accuracy = \frac{TP + FN}{TP + TN + FP + FN}$$

The average scores were obtained after several experiments had been performed. The highest score of precision occurred on the chair was 85%.The highest score of recall occurred on the sofa object was

77%. And the highest score of accuracy occurred on the chair object was 87.6%. The following is graph showing the average score of various experiments:

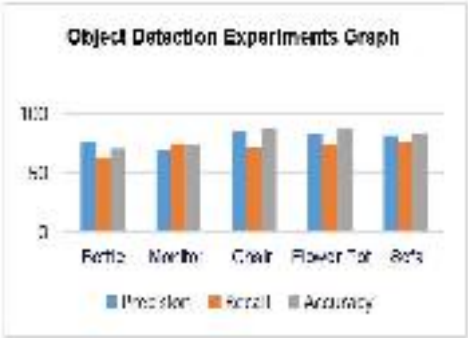

Figure 6. root filters coarse resolution, part filters finer resolution, deformation models

Figure 7. root filters coarse resolution, part filters finer resolution, deformation models

4.1 Recognition Result

The following figures are the data sample with the highest score in the detection process, in which objects are recognizable from various positions and perspectives. In some of the figures, false positive is depicted from various categories such as human, flower pot, bottle, chair, monitor, sofa and cat. The number of false positive (as in the figure of human example) shows that the bounding box emerges from more diverse part criteria.

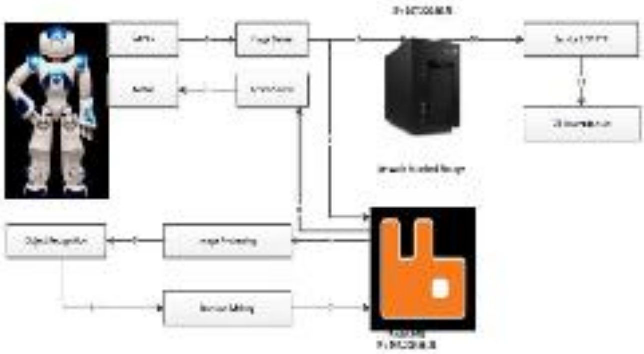

Figure 8. System Architecture

5. Conclusion

Based on the research results and discussion above, the conclusions are as follows:

1. From the result of object detection using DPM algorithm, it is obtained the *precision* score of 77.3% for bottle object, 69% for monitor object, 83% for flower pot object, 82.3% for sofa object and 85% for chair object. Therefore, the level of precision of this application is good.
2. The *recall* score of the bottle object is 63.6%, 75.3% for monitor object, 75% for flower pot object, 77% for sofa object and 72% for chair object. Therefore, the level of success of this application is good.
3. The *accuracy* score of the bottle object is 71.3%, 74.3% for monitor object, 87% for flower pot object, 84.3% for sofa object and 87.6% for chair object. Therefore, the level of accuracy of this application is good.
4. The NAO position will affect the accuracy of detection.

6. Suggestions

This researcher suggests the following:

1. It is advisable to take a more diverse dataset (set of positive and negative data) to obtain a better detection result.
2. It is advisable to pay attention to the ambient lighting to maximize the performance of detection systems on NAO robot's intelligent systems.

References

[1] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, *"Object detection with discriminatively trained part-based models,"* Pattern Analysis and Machine Intelligence, IEEE Transactions on, **vol. 32**, no. 9, pp. 1627–1645, 2010.

[2] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, *"Object detection with discriminatively trained part-based models,"* Pattern Analysis and Machine Intelligence, IEEE Transactions on, **vol. 32**, no. 9, pp. 1627–1645, 2010.

[3] N. Dalal, B. Triggs, *"Histograms of oriented gradients for human detection"*, Computer Vision and Pattern Recognition 2005. CVPR 2005. IEEE Computer Society Conference on, **vol. 1**, pp. 886-893, 2005.

[4] R. Ranjan, V. M. Patel, and R. Chellappa, *"A Deep Pyramid Deformable Part Model for Face Detection A Deep Pyramid Deformable Part Model for Face Detection"*, International Conference on Intelligent Robots and Systems (IROS), 2015 IEEE. pp. 4880-4887, 2015.

[5] S. Mittal, T. Prasad, and S. Saurabh, *"Pedestrian detection and tracking using deformable part models and Kalman filtering,"* Int. SoC Des. Conf., 2012 IEEE, pp. 324–327, 2012.

[6] G. Jie, Z. Honggang, C. Daiwu, and Z. Nannan, *"Object Detection Algorithm Based on Deformable Part Models"*, Proceedings of IC-NIDC2014, 2014 IEEE, pp. 90-94, 2014.

[7] L. C. Leon and R. Hirata, *"Vehicle detection using mixture of deformable parts models: Static and dynamic camera,"* in Brazilian Symposium of Computer Graphic and Image Processing, 2012 IEEE, pp. 237-244, 2012**.**

[8] Kurniawan, A., *"Implementasi Pengenalan Objek Menggunakan Deformable Part Model(DPM) Pada Lumen Robot Friend"*, Informatics Final Project, Universitas Bandar Lampung, 2017.

[9] Syarif, P. Nhirun, and S. Sholata, *Pengembangan Lumen Sebagai Robot Pemandu Pameran*, Studi Kasus Electrical Engineering Days 2015, B300 Engineering Documents, Bandung Institute of Technology, 2015.

[10] Cygane. B, 2013. *"Object Detection And Recognation In Digital Images"*. Poland : University of Science and Technology.

[11] W. Pratt, 2007. *"Digital Image Processing"*, New York : PIKS Scientific Inside. Hoboken, NJ : John Wiley & Sons, Inc.

[12] Marzuki, A. Sukoco, and M. S. Febrianti, *"Visual-based machine understanding framework for decision making on social robot,"* 2015 4th Int. Conf. Interact. Digit. Media, 2015 IEEE, pp. 1–6, 2015.