Human Tracking in Video Surveillance Using Particle Filter

Abdul-Lateef Yussiff, Suet-Peng Yong, Baharum B. Baharudin
Department of Computer and Information Sciences
Universiti Teknologi PETRONAS
Bandar Seri Iskandar, Tronoh
Perak, Malaysia.
MALAYSIA
ayussiff@gmail.com, yongsuetpeng@petronas.com.my, baharbh@petronas.com.my

***Abstract*—Automated human tracking is a task that has a wide area of applications and has become more important nowadays. This research proposes to investigate the use of Bayesian inference technique specifically particle filter for tracking human in video surveillance. Kalman filter which has been the de facto technique for real world tracking performs poorly for most of the problems because, the real world applications are often non-linear and non Gaussian. The particle filter on the other hand is a tool for estimating the posterior probability density of state of a dynamic model that includes non-linear and non-Gaussian real world applications. The filter uses random sample to estimate the possible location of the tracked object in the next immediate frame even in the presence of occlusion. In order to initialize the tracking process, humans are first detected using a pretrained human detection model in video. The detector utilize model fusing method which is the combination of histogram of oriented gradient based human detector model and Haar feature based upper body detector to locate position of moving person in video. The technique performed excellently well when evaluated on the publicly available CAVIAR dataset and outperformed the Kalman filter algorithm.**

***Index Terms*—Particle filter, Object tracking, Human Tracking, Probabilistic inference, Surveillance video**

I. Introduction

Moving object detection and tracking are the basic preliminary steps for video analysis and surveillance contents. According to [1], several video applications requires object of interest to be detected and tracked in a scene for the purpose of extracting semantic information prior to activity modeling and understanding.

Object tracking has been a well studied problem with a lot of attention in the computer vision and robotics communities for several decades. In spite of many contributions by researchers, there has not been a major breakthrough in the real time applications [2]. Addressing this tracking problem will instantly benefit many applications, such as human computer interaction, robotics and surveillances, and will provide relevant information with high degree of confidence for tasks that require higher-level reasoning, such as activity or event recognition, and behavior analysis. Many approaches have been proposed, each having advantages in a particular context and unique problem associated with it [3]. This paper concentrates on using probabilistic inference technique to model dynamic systems, and the focus will be on the usage of particle filter to estimate the positions of an unknown number of moving persons under complex circumstances from a frame of surveillance video to the other. Particle filter is applicable under any general hypothesis and can cope with presence of dense clutter and also very easy to implement. Human motion is non-linear and non-Gaussian making Kalman filter to fail and can ot be applied to track this king of motion [4]. The estimated states (object position and other relevant information) will help reduce the search area and computational time during tracking. Probabilistic inference or estimation theory is an umbrella theories used to estimate the state of a dynamic system by combining, usually all available reliable knowledge of the system including measurements and theoretical models with a statistical approach. Examples of a probabilistic inference technique includes but not limited to Kalman filter [5] and particle filter [6], [7].

Kalman filter has been the favorite technique for state estimation, which yield an optimal solution to the class of linear-Gaussian problem, however, there has not been an efficient and universally accepted algorithm for dealing with class of non-linear and non-Gaussian problems which is mostly common to many real world tracking applications [8].

Particle filter consists of propagating a weighted set of particles which approximate the posterior probability density of the state conditioned on the observations based on the Monte Carlo integration principles [9]. The tracking process using particle filter is an inference problem that can be summarized as given an initial state density $p(x_0)$, transition density $p(x_t|x_{t-1})$, and observation or measurement likelihood $p(z_t|x_t)$, the goal is to deduce the actual state of the object at a specify time t, i.e posterior probability density $p(x_t|Z_t)$ [10]. The particle filter will be discussed further in Sec. III-B1.

Sec. II addresses the related research work, Sec. III discusses the methodology of the research. Experiments and Results are discussed in Secs. IV and V respectively, and finally the conclusion of the paper is addressed in Sec. VI

978-1-4799-7896-0/15/$31.00 ©2015 IEEE

II. Previous Works

Significant amount of work has already been reported in the field with each system employing its own unique tracking technique. Among the proposed techniques are: Background subtraction, Mean-shift, Optical flow, Feature matching, Particle filters, Kalman filter, etc. [3]. Each proposed algorithm has its weakness as well as its strength. Background subtraction technique and Mean Shift method are commonly used in the literatures because of their simplicity. However, their simplicity nature also exposes their vulnerability.

The particle filter, also known as a Recursive Bayesian estimation predicts the posterior probability of the current state using the propagation rule of state density, gained attention for object tracking due to its robustness in tracking visual object. Isard and Blake [7] modeled tracking as a non-linear system, then applied sequential Monte Carlo technique to estimate an optimum state while tracking a single object in video. Our proposed system is similar to that, it applies probabilistic filters for tracking person, however differ in the number of objects being tracked as well as the manner of obtaining measurement (people detection) value. The proposed technique detects and tracks people in the scene, and also preserves their unique identification during tracking.

Gustafson et al. [11], [12] developed the framework for positioning, navigating and tracking using particle filter in Airborne hunting and collision avoidance vehicle. Their source of data was primarily coming from Global Positioning System (GPS). A variant of particle filter called marginalized particle filter framework which is the combination of Kalman with particle filter was employed for state estimation.

Hue et al. [9] proposed a modified particle filter to estimate the trajectories of multiple-targets from the noisy bearings measurements. Gibbs sampler was added to the standard particle filter to cater for multi-estimation targets. Ali and Dailey [13] proposed algorithm for tracking people in high density crowd. In their work, human were detected using Viola Jones' detection cascade and particle filter for tracking while color histogram was used for appearance modeling.

Khan et al. assumed moving object often interacts with each other therefore employ particle filter to track multiple interacting targets [14]. They developed a joint tracker by incorporating Markov random field into the particle filter to deal with complicated interacting targets. Medeiros et al. [15] show the implementation of a histogram-based particle filter for object tracking in parallel on smart cameras based on single instruction multiple data processors. [16] employs particle filter for target tracking based on wavelet features in real time.

The work by Xu et al. [17] is close to our approach in that, histogram of oriented gradient (HOG) and particle filter was combined to track moving person in video. Our work is different from this work because, they first isolate moving foreground object using the background subtraction techniques prior to the extraction of HOG from the detected person. Our technique combines both the HOG and Haar-feature and does not need any background subtraction to detect moving person in the video.

III. Methodology

A. Detection of moving Person

Prior to the process of tracking, moving object (human) needs to be detected and key salient points need to be extracted. These salient points are unique and will make it easy to monitor the moving objects from one frame to the other.

In this paper, a robust hybrid technique which combines two state of the art pretrained models to detect human in the video. First, Histogram of Oriented Gradient (HOG) [18] features are used to train a classifier that detect fully visible persons in the video. Secondly, Haar-like wavelets [19] are used as features while training a classifier to detect human's upper body in the video. The fusion of the two models produce a robust human detector that can detect even occluded person in the [20], [21]. HOG has been successfully used in the pedestrian detection in still images. The main idea of HOG is to use gradient orientation histogram of small grid to describe the image. HOG features describes the human body shape very well and it is invariant to small transformation. However HOG assumes humans are fully visible in the image and this sometimes not possible due to occlusion. The Haar-like wavelets have been successfully employed to detect human faces in still image [19]. They two pretrained models mutually compensate for each others deficiencies to detect human in video even in the occluded scene.

When human is located in the video, the intrinsic properties such as the position or location on the video frame where the human is found and the rectangular enclosed box(es) which is/are made of width and the height of the person are extracted.

Having detected the moving objects in the scene, we need a technique to associate each image in the current frame to the previously detected object in the immediate past frame of the video. Maintaining list of the tracked objects is very crucial in this application. Some members of the tracked list may be matched and some will not have assignment and in the similar case, the newly detected measurements may be unassigned to any of the tracked object(s). The assigned tracks are updated using the corresponding detections. The unassigned tracks are marked as 'Missing'. Finally, an unassigned detection measurement triggers the activation of a new tracking that requires monitoring. There is an attribute in each tracked object that keeps record of number of consecutive frames it has stayed unassigned, this is necessary in order to determine when the object under investigation has left the scenes and hence needs to be deleted from list of active objects. The unassigned objects are deleted from the tracked list after some number of consecutive missing count.

B. Tracking System Modeling

Designing an efficient tracking models is an art that requires domain knowledge of the application in question. The tracking system is modeled as a state space model. Human is assumed to be walking with constant velocity so the human's current state vector needs to be predicted from the previous state (location). Salient properties (Bounding boxes) are extracted from each video frame and association of detection result to the tracked human in two consecutive frames are done using Hungarian technique [22] of assignment problem. Finally, the associated data serves as measurement input to the particle filter for state estimation.

Once the people are located on the frame, and the rectangular bounding box enclosing each detected person is extracted. Each bounding box has some intrinsic properties associated with it. Example of such properties are: spatial position x, y in the $2-dimensional$ Cartesian coordinates and the size of bounding box in terms of height, h and width, w of the box. Thus intrinsic properties associated with every single person detected in a given frame, t is represented as:

$$s_j^t = (x_j^t, y_j^t, w_j^t, h_j^t) \tag{1}$$

There are several other possibilities to define these salient properties, however, this representation was found to be good enough to represent each person during tracking process.

The subsequent goal is to design a tracking algorithm to hypothesize that human's position in two consecutive frames belong to the same person. To achieve this, a prediction of the future location from the current is used. Hence, particle filter is used to predict people's position in the current frame, given their previous location in the preceding frame.

1) Particle Filter: Particle filter (a.k.a bootstrap filter [6], condensation algorithm [7]) consists of propagating a weighted set of particles which approximate the probability density of the state conditioned on the observations based on the Monte Carlo integration principles [9]. Particle filter is applicable under any general hypothesis and can cope with presence of dense clutter and also very easy to implement. Clutter is generally considered as a model describing false positive with statistical properties different from those of the moving object. The main idea is to represent the posterior probability density function (pdf) as a set of random particles instead of using function over the state space. The dynamic process represented by the state equation is given as

$$x_k = f_{k-1}(x_{k-1}, w_k) \tag{2}$$

This can be written in the equivalent probability form as $x_k \sim p(x_k|x_{k-1})$. where, $x_k \in \Re^n$ is the stochastic process', state vector, $f_k \in R^n$ is the system transition function, w_k, is a white noise, zero mean with known pdf not necessarily Gaussian, k is the time sequence. $p(.|.)$ is the conditional probability of the state which obeys Markovian properties. The measurement is related to the state vector through the equation given as

$$z_k = h_k(x_k, v_k) \tag{3}$$

This is written in probability form as $z_k \sim p(z_k|x_k)$ where $z_k \in \Re^m$ is the sequentially arriving observed and measured bounding box's value. h_k is the measurement function, v_k is white, zero mean noise with known pdf but not necessarily Gaussian. Given a known initial pdf $p(x_0)$ and all the measurement so far including time k. The goal is to recursively estimate the posterior pdf $p(x_k|z_k)$. The prior pdf of the state at any time interval k is given as

$$p(x_k|z_{k-1}) = \int p(x_k|x_{k-1})p(x_{k-1}|z_{k-1})dx_{k-1} \tag{4}$$

but the probability model for $p(x_k|x_{k-1})$ is defined in term of state dynamic equation and the error statistics of w_{k-1}

$$p(x_k|x_{k-1}) = \int p(x_k|x_{k-1}, w_{k-1})p(w_{k-1}|x_{k-1})dw_{k-1} \tag{5}$$

Also by assumption of independency, $p(w_{k-1}|x_{k-1}) = p(w_{k-1})$. Therefore, after substitution this known probabilities in equation 5 , we have

$$p(x_k|x_{k-1}) = \int \delta(x_k - f_{k-1}(x_{k-1}, w_{k-1})) \times p(w_{k-1})dw_{k-1} \tag{6}$$

$\delta(.)$ is called Dirac delta function due to the fact that x_{k-1} and w_{k-1} is determinable and known. At any time, measurement z_k becomes available, the prior can be update to get the posterior probability according to the Bayes rule

$$p(x_k|Z_k) = \frac{p(z_k|x_k)p(x_k|Z_{k-1})}{\int p(z_k|x_k)p(x_k|Z_{k-1})dx_k} \tag{7}$$

The denominator integral is just for normalization so that the posterior will be less than or equal to 1 and $Z_k = \{z_i\}_{i=1}^k$. Therefore $p(z_k|x_k)$ is defined in terms of the measurement model and v_k which is a known statistics.

$$p(z_k|x_k) = \int \delta(z_k - h_k(x_k, v_k))p(v_k)dv_k \tag{8}$$

If the Gaussian and linearity assumptions do not hold, then evaluating the integration becomes intractable [10], hence sampling techniques such as particle filter can be very advantageous to solve this kind of problems.

2) PF Algorithm:

1) Generate N random samples from the initial distribution $p(x_0)$; $\{x_0^i\}_{i=1}^N$
2) Calculate the weights $w_k^i = p(z_k|x_k^i)$
3) Normalize the weights such that sum of all weights equal unity; $\hat{w}_k^i = \frac{w_k^i}{\sum_{j=1}^N w_k^j}$
4) Output the desired statistics of the particles such as MMSE, RMSE, etc.
5) Resample N new set of particles with replacement by generating $\{x_k^i\}_{i=1}^N$ such that
$$Pr(x_k^i = x_k^j) = \hat{w}_k^j$$
6) Predict new particles by using the state dynamic equation
$$x_{k+1}^i = f(x_k^i, w_k), i = 1, ..., N$$
7) Increment the frame number, k, then go to step number 2.

IV. Experiments

Experimental results is presented in this section. The performance of the particle filter outlined in Sec III-B1 is evaluated on CAVIAR (Context Aware Vision using Image-based Active Recognition) dataset [23], one of the popular dataset in people tracking and crowd modeling domain. The CAVIAR project was funded by EC's Information Society Technology programme to capture indoor video surveillance under different lighting conditions with variable degree of occlusions and scaling with the goal of analyzing the video stream to detect unusual events in the video sequence. The activities captured in the dataset includes shop entering and exiting, walking, meeting, fighting, window shopping, etc.

Particle filter with sequential importance resampling is adopted for the people tracking. The tracking process starts when an object is detected and has been confirmed for some number of conservatives frames of the video. The particle filter algorithm is as indicated in the Sec. III-B1. The number of sample used for the experiment is 500. The experiment was run 5 times on the *ThreePastShop2cor* and the root mean square of the state estimation is calculated for each trial of the experiment.

Fig. 1 shows the people tracking result on the *ThreePastShop2cor* dataset obtained from CAVIAR [[23]].

V. Result and Discussion

Fig 2 shows the path or trajectory of the tracked person in the scenes. The measurement (the detected position of the person) is shown in on the graph as the circle while the solid line is the estimate of the position of the person using the particle filter algorithm. The trajectory as can be seen from the graph shows non-linearity nature of the data obtained from the detection. The particle filter estimated the position of tracked person with high accuracy which can be deduce from its to lower estimation error. The experiment was conducted several times using the same video dataset and for each experiment, the Root Mean Square Error (RMSE) is calculated which is shown in the Table I.

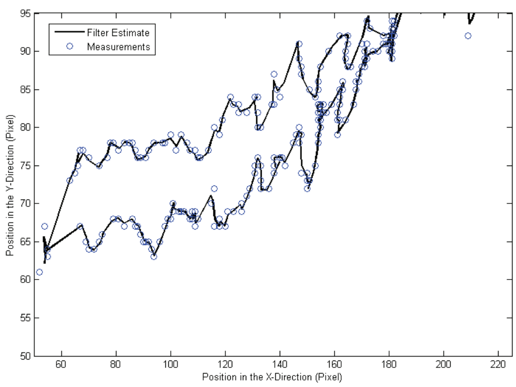

Fig. 2: Plot of measurements and filter estimates trajectory of the tracked person based on the pixel location on the image frame

TABLE I: The Root Mean Square Error (RMSE) for each dataset of using Particle filter (PF) algorithm and Kalman filter (KF)

Experimental dataset	PF RMSE	KF RMSE
EnterExitCrossingPaths1cor	0.265	1.568
EnterExitCrossingPaths2cor	0.263	2.125
ThreePastShop2cor	2.5220	14.5639
TwoEnterShop2cor	2.1929	16.3470
Average	1.2516	8.6501

Fig. 3 is the graph showing the pixel location of the tracked person in the video image frame of *ThreePastShop2cor* (CAVIAR) dataset. The tracked person denoted as *Person1* moves along a path in a particular direction. Another person denoted as *Person2*, starts moving in the opposite direction at the time when *Person1* disappear from the scenes.

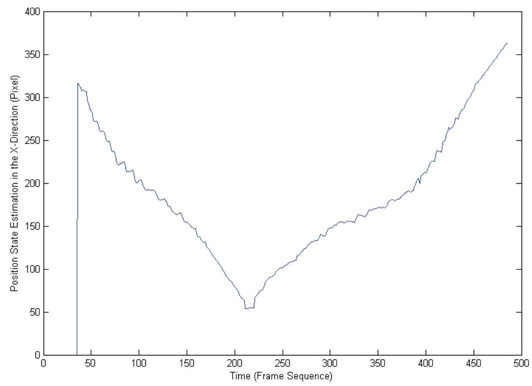

Fig. 3: Position of the tracked person in the X-direction in unit of pixel. This is to show the location of the moving person in one of the vector components with respect to time.

Both Figs. 4 and 5, show the state estimation errors for each time sequence in the X-axis and Y-axis respectively. The errors which is randomly distributed about the horizontal axis supports the validity and correctness of the particle filter model.

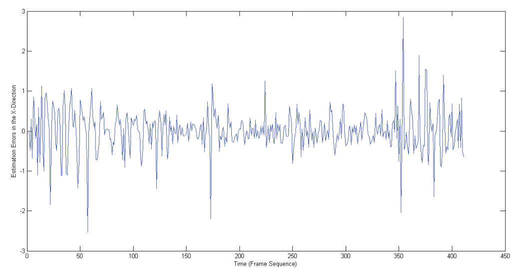

Fig. 4: Position estimation error plots in the X Coordinate axis. This shows that the errors are randomly distributed around the zero error. This is to validate the normalcy of the data in the x-direction.

(a) Original Frames (b) Particle Filtered Frames (c) Kalman Filtered Frames

Fig. 1: The figures show people tracking at work on some selected frames. (a) The first column shows the original frames (b) The second column shows Particle filtered frames (c) The Kalman filtered frames. Analysis of the bounding boxes placement during tracking shows the Particle filtered tracking is better.

VI. Conclusion

The use of Bayesian inference technique for object tracking is investigated. Due to the lower performance of the Kalman filter on most real world tracking applications which often violates linear - Gaussian assumptions, a generic and robust method of estimating the state of a dynamic system is applied to tracking human in both indoor and outdoor surveillance video. The proposed technique uses a combination of two pretrained human detection models to locate the position of human in the video. Then particle filter is use to track human from one frame to the other. The adopted particle filter addresses the sample impoverishment which is a common problem to most particle filter and did not assume linear-Gaussian assumption in which the Kalman filter relies on. The

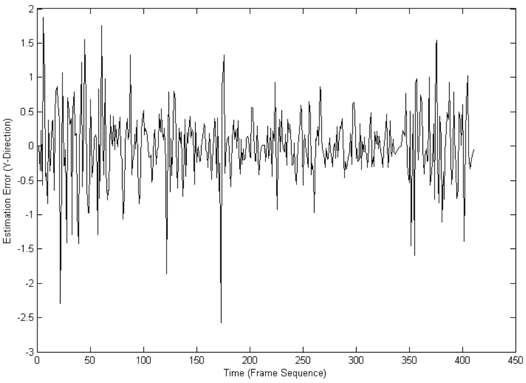

Fig. 5: Pixel position estimation error plots in the Y Coordinate axis. This graph is to show the normalcy of the data used for the tracking.

evaluation of the algorithm on the CAVIAR dataset shows a good result that yield lower RMSE of 1.252 on the average when compared with Kalman filter algorithm's RMSE of 8.6501.

References

[1] E. Corvee, F. Bremond *et al.*, "Haar like and lbp based features for face, head and people detection in video sequences," in *International Workshop on Behaviour Analysis and Video Understanding (ICVS 2011)*, 2011.

[2] A. W. Smeulders, D. M. Chu, R. Cucchiara, S. Calderara, A. Dehghan, and M. Shah, "Visual tracking: An experimental survey," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 36, no. 7, pp. 1442–1468, 2014.

[3] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *Acm Computing Surveys (CSUR)*, vol. 38, no. 4, p. 13, 2006.

[4] L. Jin, J. Cheng, and H. Huang, "Human tracking in the complicated background by particle filter using color-histogram and hog," in *Intelligent Signal Processing and Communication Systems (ISPACS), 2010 International Symposium on*. IEEE, 2010, pp. 1–4.

[5] M. S. Grewal and A. P. Andrews, "Kalman filtering: theory and practice using matlab," *John Wiley & Sons, Baltimore, MD*, vol. 2, p. 35, 2001.

[6] N. J. Gordon, D. J. Salmond, and A. F. Smith, "Novel approach to nonlinear/non-gaussian bayesian state estimation," in *IEE Proceedings F (Radar and Signal Processing)*, vol. 140, no. 2. IET, 1993, pp. 107–113.

[7] M. Isard and A. Blake, "Contour tracking by stochastic propagation of conditional density," in *Computer VisionECCV'96*. Springer, 1996, pp. 343–356.

[8] R. Chen and J. S. Liu, "Mixture kalman filters," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 62, no. 3, pp. 493–508, 2000.

[9] C. Hue, J.-P. Le Cadre, and P. Pérez, "Tracking multiple objects with particle filtering," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 38, no. 3, pp. 791–812, 2002.

[10] X. Zhang, W. Hu, S. Maybank, X. Li, and M. Zhu, "Sequential particle swarm optimization for visual tracking," in *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE, 2008, pp. 1–8.

[11] F. Gustafsson, F. Gunnarsson, N. Bergman, U. Forssell, J. Jansson, R. Karlsson, and P.-J. Nordlund, "Particle filters for positioning, navigation, and tracking," *Signal Processing, IEEE Transactions on*, vol. 50, no. 2, pp. 425–437, 2002.

[12] F. Gustafsson, "Particle filter theory and practice with positioning applications," *Aerospace and Electronic Systems Magazine, IEEE*, vol. 25, no. 7, pp. 53–82, 2010.

[13] I. Ali and M. N. Dailey, "Multiple human tracking in high-density crowds," *Image and vision computing*, vol. 30, no. 12, pp. 966–977, 2012.

[14] Z. Khan, T. Balch, and F. Dellaert, "An mcmc-based particle filter for tracking multiple interacting targets," in *Computer Vision-ECCV 2004*. Springer, 2004, pp. 279–290.

[15] H. Medeiros, G. Holguín, P. J. Shin, and J. Park, "A parallel histogram-based particle filter for object tracking on simd-based smart cameras," *Computer Vision and Image Understanding*, vol. 114, no. 11, pp. 1264–1272, 2010.

[16] T. Rui, Q. Zhang, Y. Zhou, and J. Xing, "Object tracking using particle filter in the wavelet subspace," *Neurocomputing*, vol. 119, pp. 125–130, 2013.

[17] J. Xu, A. Beaugendre, and S. Goto, "Real-time human tracking by detection based on hog and particle filter," in *Computer Sciences and Convergence Information Technology (ICCIT), 2011 6th International Conference on*. IEEE, 2011, pp. 193–198.

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

[19] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1. IEEE, 2001, pp. I–511.

[20] A.-L. Yussiff, S.-P. Yong, and B. B. Baharudin, "People detection enrichment for abnormal human activity detection," *Australian Journal of Basic and Applied Sciences*, vol. 7, no. 8, pp. 632–640, 2013.

[21] A.-L. Yussiff, S.-P. Yong, and B. B. Baharudin, "Detecting people using histogram of oriented gradients: a step towards abnormal human activity detection," in *Advanced in Computer Science and its Applications*. Springer, 2014, pp. 1145–1150.

[22] H. W. Kuhn, "The hungarian method for the assignment problem," *Naval research logistics quarterly*, vol. 2, no. 1-2, pp. 83–97, 1955.

[23] Caviar: Context aware vision using image-based active recognition. [Online]. Available: http://homepages.inf.ed.ac.uk/rbf/CAVIAR/