

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 1, January 2015

Regression Based Crowd Density Estimation

AR.Arunachalam, Dr. Robert Masilamani

Assistant Professor ,Department of Computer Science, Bharath University, Chennai, India

Professor , Department of Computer Science, Bharath University, Chennai, India

ABSTRACT: We present a privacy-preserving system for estimating the size of inhomogeneous group, composed of pedestrians that travel in different ways, without using explicit component segmentation or tracking. First, the group is segmented into objects of homogeneous motion, using the compound of dynamic textures motion model. Second, a set of important feature is extracted from every segmented area, and the related features and the number of people per segment is learned with Bayesian Process regression. In this paper two Bayesian regression models are examined. The first model is a combination of Gaussian process regression with a compound kernel, which associates for both the global and local trends of the count mapping but is limited by the real-valued outputs that do not match the discrete counts. This limitation is addressed with a second model, which is depends on a Bayesian treatment of Poisson regression that introduces a prior distribution on the linear weights of the model. The two regression-based crowd counting methods are evaluated on a big pedestrian data set, containing very distinct camera angles, pedestrian traffic, and odds, such as bikes or skateboarder. Research results show that regression-based counts are accurate regardless of the group size, outperforming the count calculates produced by state-of-the-art pedestrian detectors. Finally, result of the system running on 2 h of video. It demonstrates the efficiency and robustness of the regression-based crowd size estimation over long periods of time.

KEYWORDS: Bayesian regression, crowd analysis, Gaussian processes, Poisson regression, surveillance.

I. INTRODUCTION

Recent work on crowd counting using low-level feature regression has shown promise in computer vision. One advantage with these methods is that they bypass intermediate processing stages, such as people detection or tracking, that may be susceptible to problems when the crowd is dense. In , the scene is segmented into crowds moving in different directions and various low-level features are extracted from each crowd segment (*e.g.* information on the shape, edges and texture of the segment). The crowd count in each segment is then estimated with a Gaussian process (GP) regression function that maps feature vectors to the crowd size. Experiments in indicate that the counting algorithm is capable of producing accurate counts, for a wide range of crowd densities.

One problem with the system of is that it uses GP regression, which models *continuous real-valued* functions, to predict *discrete counting numbers*. Because of this mismatch, regression may not be taking full advantage of Bayesian inference. For example, rounding of the realvalued predictions is not handled in a principled way, *e.g.* by reducing the confidence when the prediction is far from an integer. In addition, the confidence levels are currently measured in standard-deviations, which provides little intuition on the reliability of the estimates. A confidence measure based on posterior probabilities seems more intuitive for counting numbers. Finally, negative outputs of the GP must be truncated to zero, and it is unclear how this affects the optimality of the predictive distribution.

One common method of regression for counting numbers is Poisson regression , which models the noisy output of a counting function as a Poisson random variable, with a log-mean parameter that is a linear function of the input vector.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 1, January 2015

This is analogous to standard linear regression, except that the mean is modeled as the exponential of a linear function to ensure non-negative values, and that the noise model is Poisson because the outputs are counting numbers. One way of extending Poisson regression to the Bayesian setting is to adopt a hierarchical model, where the log-mean function is modeled with a standard Gaussian process. These solutions, however, have two disadvantages. First, because of the lack of conjugacy between the Poisson and the GP, must approximate inference with Markov-chain Monte Carlo (MCMC), which limits these algorithms to small datasets. Second, the hierarchical model contains two redundant noise sources: 1) the Poisson-distributed observation noise, and 2) the Gaussian noise of the GP in the log-mean function. These two noise terms model essentially the same thing: the noise in observing the count. A more parsimonious representation would include only the observation noise, while modeling the mean as a deterministic function.

The contributions of this paper are threefold, spanning open questions in computer vision and machine learning. First, a “crowd-centric” methodology for estimating the sizes of crowds is moving in different directions, which does not depend on object detection or feature tracking, is presented. Second, a Bayesian regression procedure is derived for the estimation of counts, which is a Bayesian extension of Poisson regression. A closed-form approximation to the predictive distribution, which can be kernelized to handle nonlinearities, is derived, together with an approximate

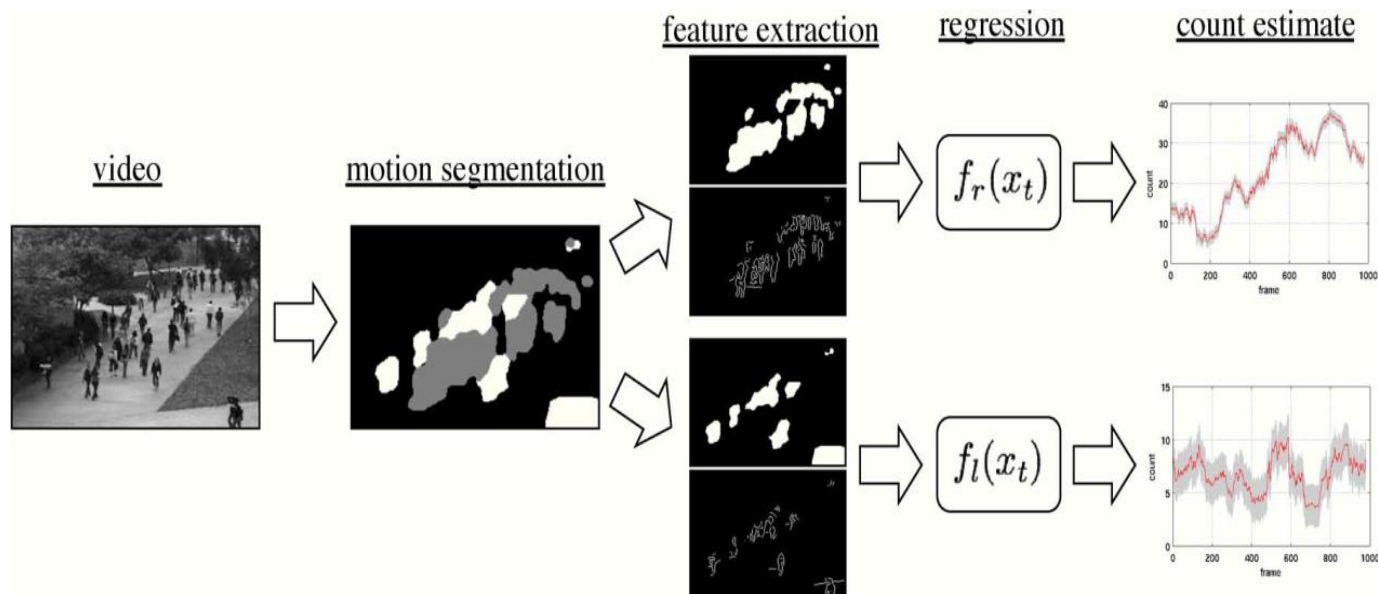


Fig.1 Crowd counting from low-level features. The scene is segmented into crowds moving in different directions. Features are extracted from each segment and normalized to account for perspective. The number of people in each segment is estimated with Bayesian regression.

procedure for optimizing the hyper parameters of the kernel function, under the Type-II maximum marginal likelihood criteria. It is also shown that the proposed approximation to BPR is related to a GPR with a specific noise term. Third, the proposed crowd counting approach is validated on two large data sets of pedestrian imagery, and its robustness demonstrated through results on 2 hours of video. To our knowledge, this is the first pedestrian counting system that accounts for multiple pedestrian flows and successfully operates continuously in an outdoor unconstrained environment for such periods of time.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 1, January 2015

II. METHODOLOGY

The goal of the proposed system is to estimate the number of people moving in each direction, in a privacy-preserving manner[1]-[3]. An outline of the crowd counting system appears in Figure 1. The video is segmented into crowd regions moving in different directions, using a mixture of dynamic textures. For each crowd segment, various features are extracted, while applying a perspective map to weight each image location according to its approximate size in the real scene. Finally, the number of people per segment is estimated with Bayesian regression

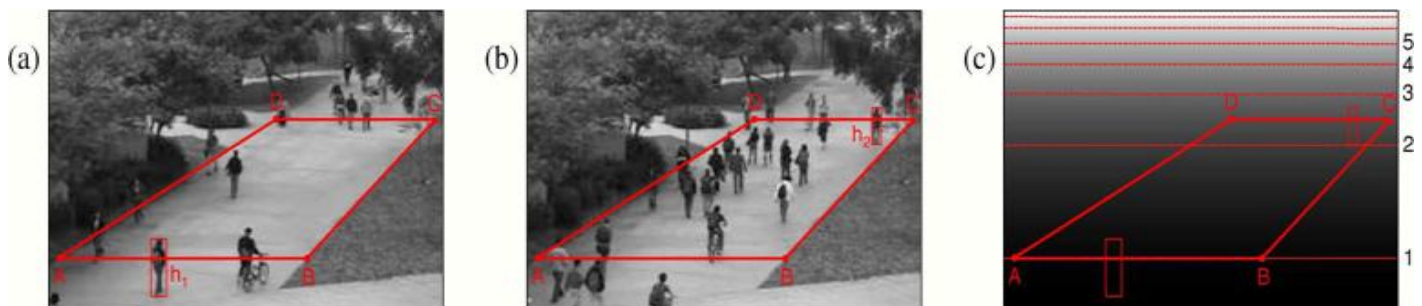
A. Crowd segmentation

We adopt the *mixture of dynamic textures* [18] to segment the crowds moving in different directions. The video is represented as collection of spatio-temporal patches ($7 \times 7 \times 20$ patches in all experiments reported in the paper), which are modeled as independent samples from a mixture of dynamic texture models [19]. The mixture model is learned with the expectation-maximization (EM) algorithm [18]. Video locations are then scanned sequentially; a patch is extracted at each location, and assigned to the mixture component of largest posterior probability. The location is declared to belong to the segmentation region associated with that component. For long sequences, where characteristic motions are not expected to change significantly, the computational cost of the segmentation can be reduced by learning the mixture model from a subset of the video (*e.g.* a representative clip). The remaining video can then be segmented by computing the posterior assignments as before. This procedure tends to work well in practice, and was used in this paper to segment a full hour of video[4][5].

B. Perspective normalization

Before extracting features from the video segments, it is important to consider the effects of perspective. Because objects closer to the camera appear larger, any feature extracted from a foreground object will account for a smaller portion of the object than one extracted from an object farther away[6]-[9]. This makes it important to normalize the features for perspective. One possibility is to weight each pixel according to a perspective normalization map. The pixel weight is based on the expected depth of the object which generated the pixel, with larger weights given to far objects

In this work, we approximate the perspective map by linearly interpolating between the two extremes of the scene. A ground plane is first marked, as in Figure 2a, and the distances \overline{ab} and \overline{cd} are measured. Next, a reference pedestrian is selected, and the heights h_1 and h_2 are measured when the center of the person is on \overline{ab} and \overline{cd} (see Figures 2a and 2b). The pixels on \overline{ab} are given a weight of 1, and the pixels on \overline{cd} a weight of $\frac{h_1|\overline{ab}|}{h_2|\overline{cd}|}$. Finally, the remaining pixel weights are computed by interpolating linearly between the two lines.



International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 1, January 2015

Figure 2. Perspective map: a) reference person at the front of walkway, and b) at the end; c) the perspective map, which scales pixels by their relative size in the true 3d scene.

Figure 4c shows the perspective map of the scene using the above procedure. In this case, objects on the front-line \overline{ab} are approximately 2.4 times bigger than objects on the back-line \overline{cd} . Finally, for features based on area (e.g. segmentation area), the weights are applied to each pixel. For features based on edges (e.g. edge histogram), the square-roots of the weights are used [10].

C. Feature extraction

Ideally, features such as segmentation area or number of edges should vary linearly with the number of people in the scene [14, 11]. To model the non-linearities, we extract an additional 28 features from each crowd segment.

Segment features: These features capture segment shape and size.

- *Area* – total number of pixels in the segment.
- *Perimeter* – total number of pixels on the segment perimeter, computed with morphological operators.
- *Perimeter edge orientation* – orientation histogram of the segment perimeter. The orientations are quantized into 6 bins, and opposite orientations (180° apart) are considered equal. The orientation of each edge pixel is computed by finding the maximum response to a set of oriented Gaussian filters at that point.
- *Perimeter-area ratio* – ratio between the segment perimeter and area, which measures the complexity of the segment shape. Segments of high ratio contain bumps in their perimeter, which may be indicative of the number of people contained within.

Internal edge features: The edges contained in a crowd segment are a strong clue about the number of people in the segment [15, 14]. A Canny edge detector [20] is applied to the entire image, the edge image is masked by the crowd segmentation, and the following features are computed:

- *Total edge pixels* – total number of edge pixels contained in the segment.
- *Edge orientation* – histogram of the edge orientations in the segment, generated in the same way as the perimeter edge histogram.
- *Minkowski dimension* – the Minkowski fractal dimension of the edges in the segment, which estimates their degree of “space-filling”

Texture features: Texture features, based on the gray-level co-occurrence matrix (GLCM), were used in [16] to classify image patches into 5 classes of crowd density. In this work, we adopt a similar set of measurements for *counting* the number of pedestrians in each segment. The image is quantized into 8 gray levels, and the 2nd-order joint conditional probability density function $f(i, j/d, \theta)$ is estimated

for distance $d = 1$ and angles $\theta = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$.

The following texture properties are computed:

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 1, January 2015

- Homogeneity: $S_h(d, \varphi) = \sum_{i,j} \frac{f(i,j|d, \varphi)}{1+(i-j)^2}$
- Energy: $S_g(d, \theta) = \sum_{i,j} f(i,j|d, \theta)^2$
- Entropy: $S_p(d, \theta) = \sum_{i,j} f(i,j|d, \theta) \log f(i,j|d, \theta)$

resulting in a total of 12 texture features. .

D. Gaussian process regression:

A Gaussian process (GP) is used to regress feature vectors to the number of people per segment. The GP defines a distribution over functions, which is “pinned down” at the training points. The classes of functions that the GP can model is dependent on the kernel function used. For the task of pedestrian counting, we note that the dominant trend of many of the features is linear (e.g. segment area), with local non-linearities. To capture both trends, we combine the linear and the squared-exponential (RBF) kernels, i.e.

$$k(x_p, x_q) = \alpha_1 (x_p^T x_q + 1) + \alpha_2 e^{-\frac{\|x_p - x_q\|^2}{\alpha_3}} + \alpha_4 \delta(p, q)$$

with hyper parameters $\alpha = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$. The first and second terms of the kernel are the linear and RBF components, while the third term models observation noise.

Note that the linear component of the kernel captures the dominant trend, while the RBF component models local non-linearities. Finally, while the same feature set is used throughout the system, a different regressor is learned for each direction of crowd motion because the appearance changes with the traveling direction[12][13].

E. Bayesian Poisson Regression:

While GPR is a Bayesian framework for regression problems with real-valued output variables, it is not a natural regression formulation when the outputs are nonnegative integers, i.e., $Y \in Z_+ = \{0, 1, 2, 3 \dots\}$ as is the case for counts. A typical solution is to model the output variable as Poisson or negative binomial (NB), with an arrival-rate parameter that is a function of the input variables, resulting in the standard Poisson regression or NB regression. Although both these methods model counts, they do not support Bayesian inference, i.e., do not consider the weight β vector as a random variable. This limits their generalization from small training samples and prevents a principled probabilistic approach to learning hyper parameters in a kernel formulation[16]-[18].

The posterior distribution of β , given a training sample $\{x, y\}$, is given by Bayes’ rule as follows:

$$p(\beta | X, y) = \frac{p(y|X, \beta)p(\beta)}{\int p(y|X, \beta)p(\beta) d\beta}$$

Due to the lack of conjugacy between the Poisson likelihood and the Gaussian prior does not have a closed-form expression; therefore, an approximation is necessary

Approximate Posterior Distribution

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 1, January 2015

We first derive a closed-form approximation to the posterior distribution

$$p(\mathbf{y}|\mathbf{X}, \boldsymbol{\beta}) = \prod_{i=1}^N \frac{1}{\mathbf{y}_i!} e^{v(x_i)\mathbf{y}_i} e^{-v(x_i)}$$

$$= \prod_{i=1}^N \left[\frac{e^{v(x_i)(\mathbf{y}_i+c)} e^{-e^{v(x_i)}}}{\Gamma(\mathbf{y}_i + c)} \right] e^{-cv(x_i)} \frac{\Gamma(\mathbf{y}_i + c)}{\mathbf{y}_i!}$$

Where $c>0$ is a constant. The approximation is based on two facts. First, the term in the square brackets is the likelihood of the data under a log-gamma distribution of parameters $(\mathbf{y}+c, 1)$

Finally, by completing the square, the posterior distribution is approximately Gaussian, i.e.,

$$p(\boldsymbol{\beta}|\mathbf{X}, \mathbf{y}) \approx N(\boldsymbol{\beta}|\hat{\boldsymbol{\mu}}_{\boldsymbol{\beta}} \sum \boldsymbol{\beta})$$

with mean and variance

$$\hat{\boldsymbol{\mu}}_{\boldsymbol{\beta}} = (\Phi \Sigma_y^{-1} \Phi^T + \Sigma_p^{-1})^{-1} \Phi \Sigma_y^{-1} \mathbf{t}$$

$$\hat{\Sigma}_{\boldsymbol{\beta}} = (\Phi \Sigma_y^{-1} \Phi^T + \Sigma_p^{-1})^{-1}.$$

Note that setting $c=0$ will yield the original posterior approximation in . Constant acts as a parameter that controls the smoothness of the approximation around $\mathbf{y}=0$, avoiding the logarithm of or division by zero. In the experiments, we set this parameter to $c=1$.

Bayesian Prediction

Given a novel observation \mathbf{x}_* , we start by considering the predicted log-arrival rate $\mathbf{v}_* = \boldsymbol{\phi}(\mathbf{x}_*)^T \boldsymbol{\beta}$. It follows from that the posterior distribution of \mathbf{v}_* is approximately Gaussian:

$$p(\mathbf{v}_*|\mathbf{x}_*, \mathbf{X}, \mathbf{y}) \approx N(\mathbf{v}_*|\hat{\boldsymbol{\mu}}_{\mathbf{v}_*}, \hat{\sigma}_{\mathbf{v}_*}^2)$$

The predictive distribution for \mathbf{y}_* is

$$p(\mathbf{y}_*|\mathbf{x}_*, \mathbf{X}, \mathbf{y}) = \int p(\mathbf{y}_*|\mathbf{v}_*)p(\mathbf{v}_*|\mathbf{x}_*, \mathbf{X}, \mathbf{y})d\mathbf{v}_*.$$

In summary, the predictive distribution of can be approximated by an NB as follows:

$$\mathbf{y}_*|\mathbf{x}_*, \mathbf{X}, \mathbf{y} \sim \text{NegBin}(e^{\hat{\boldsymbol{\mu}}_{\mathbf{v}_*}}, \hat{\sigma}_{\mathbf{v}_*}^2)$$

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 1, January 2015

of mean and scale

$$\hat{\mu}_\nu = \phi(\mathbf{x}_*)^T (\Phi \Sigma_y^{-1} \Phi^T + \Sigma_p^{-1})^{-1} \Phi \Sigma_y^{-1} \mathbf{t}$$

$$\hat{\sigma}_\nu^2 = \phi(\mathbf{x}_*)^T (\Phi \Sigma_y^{-1} \Phi^T + \Sigma_p^{-1})^{-1} \phi(\mathbf{x}_*).$$

Finally, the mode of \mathbf{y}_* is adjusted downward depending on the amount of over dispersion, i.e., $\text{mode}(\mathbf{y}) = \begin{cases} 1 - \sigma_\nu^2 \\ 0 \end{cases}$, where $\lfloor \cdot \rfloor$ the floor function.

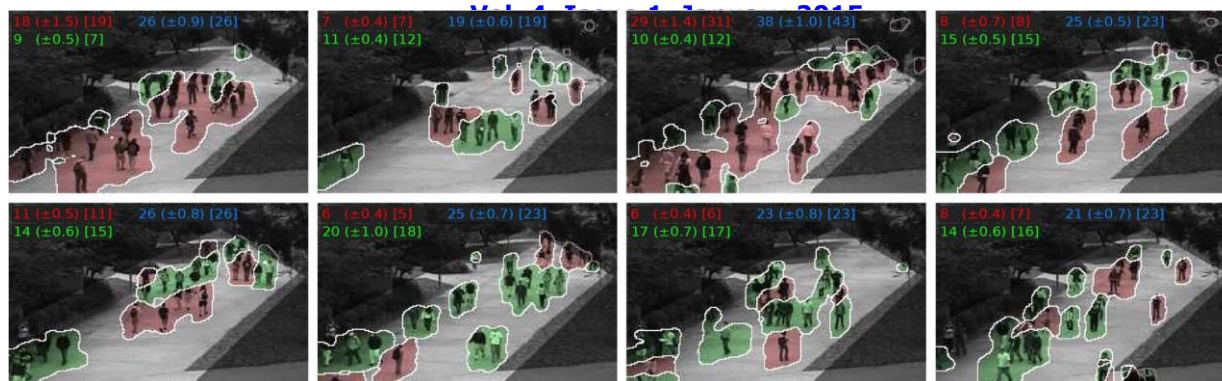
E. Result Analysis

The crowd counting system was trained on the 800 training frames, and tested on the remaining 1200. The GP output was rounded to the nearest integer to produce a crowd count, and both the mean-squared-error (MSE) and the absolute error between this estimate and the ground-truth were recorded.[14-16] For comparison, we also trained the system with different subsets of the features: only segment area, the segment features, and the segment and edge features. Finally, we compared performance against the feature sets of [17] (segment size histogram and edge orientation histogram) and [18] (segment area and total edges).

Finally, using all features (*i.e.* adding the texture features) performs best, with an MSE of 4.181/1.291. This demonstrates the informativeness of the different feature subsets: the segment features provide a coarse linear estimate, which is refined by the edge and texture features accounting for various non-linearities. The full feature set also performs better than those of [19][20]. Finally, using features that do not accounting for perspective, the system performance drops significantly (MSE 6.869/2.541), indicating that normalization is indeed required. [21-23]

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)



Crowd counting examples: The red and green segments are the “away” and “towards” components of the crowd. The estimated crowd count for each segment is shown in the top left, with the (uncertainty) and the (ground truth). The prediction for the “scene” class, which is count of the whole scene, is shown in the top right. The ROI is also highlighted

Finally, we applied Bayesian regression (BPR or GPR) on the detector counts (HOG or DPM), in order to remove any systematic bias in the count prediction[24]. Using the training set, a Bayesian regression function was learned to map the detector count to the ground-truth count. The counting accuracy on the test set was then computed using the regression function. The (best) results are presented in the bottom half of Table V. There is not a significant improvement compared with the raw counts, suggesting that there is no systematic warping between the detector counts and the actual counts.[25]

III. CONCLUSIONS

In this paper, we have proposed the use of Bayesian regression to estimate the size of inhomogeneous crowds, which are composed of pedestrians traveling in different directions, without using intermediate vision operations, such as object detection or feature tracking. Two solutions were presented, based on the GPR and BPR. The intractability of the latter was addressed through the derivation of closed-form approximations to the predictive distribution. It was shown that the BPR model can be kernelized, to represent nonlinear log-arrival rates, and that the hyperparameters of the kernel can be estimated by approximate maximum marginal likelihood. Regression-based counting was validated on two large data sets and was shown to provide robust count estimates regardless of the crowd size.

REFERENCES

1. J. Vaidya, M. Zhu, and C. W. Clifton, *Privacy Preserving Data Mining*. Springer, 2006.
2. Udayakumar R., Khanaa V., Kaliyamurthie K.P., "High data rate for coherent optical wired communication using DSP", *Indian Journal of Science and Technology*, ISSN : 0974-6846, 6(S6) (2013) 4772-4776.
3. V. Verykios, E. Bertino, I. Fovino, L. Provenza, Y. Saygin, and Y. Theodoridis, "State-of-the-art in privacy preserving data mining," *ACM SIGMOD Record*, vol. 33, no. 1, pp. 50–57, March 2004.
4. Hariharan V.S., Nandlal B., Srilatha K.T., "Efficacy of various root canal irrigants on removal of smear layer in the primary root canals after hand instrumentation: A scanning electron microscopy study", *Journal of Indian Society of Pedodontics and Preventive Dentistry*, ISSN : 0970-4388, 28(4) (2010) pp.271-277.
5. P. Viola, M. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," *IJCV*, vol. 63(2), pp. 153–61, 2005.
6. Udayakumar R., Khanaa V., Kaliyamurthie K.P., "Optical ring architecture performance evaluation using ordinary receiver", *Indian Journal of Science and Technology*, ISSN : 0974-6846, 6(S6) (2013) pp. 4742-4747.
7. T. Zhao and R. Nevatia, "Bayesian human segmentation in crowded situations," in *CVPR*, vol. 2, 2003, pp. 459–66.
8. Kanniga E., Selvaramarathnam K., Sundararajan M., "Embedded control using mems sensor with voice command and CCTV camera", *Indian Journal of Science and Technology*, ISSN : 0974-6846, 6(S6) (2013) pp.4794-4796.
9. B. Leibe, E. Seemann, and B. Schiele, "Pedestrian detection in crowded scenes," in *CVPR*, vol. 1, 2005, pp. 875–85.
10. Udayakumar R., Khanaa V., Kaliyamurthie K.P., "Performance analysis of resilient fth architecture with protection mechanism", *Indian Journal of Science and Technology*, ISSN : 0974-6846, 6(S6) (2013) pp. 4737-4741
11. B. Wu and R. Nevatia, "Detection of multiple, partially occluded humans in a single image by bayesian combination of edgelet part detectors," in *ICCV*, vol. 1, 2005, pp. 90–7.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 1, January 2015

12. S.-F. Lin, J.-Y. Chen, and H.-X. Chao, "Estimation of number of people in crowded scenes using perspective transformation," IEEE Trans. System, Man, and Cybernetics, vol. 31, no. 6, 2001.
13. A.Geetha, Face Recognition Using OPENCL, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering , ISSN (Print) : 2320 – 3765, pp- 7148-7151, Vol. 3, Issue 2, Febuary 2014.
14. A.Geetha, Universal Asynchronous Receiver / Transmitter (UART) Design for Hand Held Mobile Devices, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, ISSN: 2231-5381, pp 25-26, Volume 3 Issue 1 No1 – January 2012.
15. B.Mehala, Mrs.Anitha Sampath Kumar, Design and Implementation of Resonant Circuit Based On Half-Bridge Boost Rectifier with Output Voltage Balance Control, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, ISSN (Online): 2278 – 8875,pp 9370-9378, Vol. 3, Issue 5, May 2014
16. B.Vamsi Krishna, Starting Inrush Current Control of Three-Phase Induction Motors for Dispersed Generating Systems, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, ISSN (Online): 2278 – 8875,pp 6411-6422, Vol. 2, Issue 12, December 2013
17. B.Vamsi Krishna, A New Technique for Elimination of Harmonics Using Three Phase Shunt Active Filter, International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, ISSN (Online): 2278 – 8875,pp 5596-5602, Vol. 2, Issue 11, November 2013
18. V. Rabaud and S. J. Belongie, "Counting crowded moving objects," in CVPR, 2006.
19. G. J. Brostow and R. Cipolla, "Unsupervised bayesian detection of independent motion in crowds," in CVPR, vol. 1, 2006, pp. 594–601.
20. B. Leibe, K. Schindler, and L. Van Gool, "Coupled detection and trajectory estimation for multi-object tracking," in ICCV, 2007.
21. N. Paragios and V. Ramesh, "A mrf-based approach for real-time subway monitoring," in CVPR, vol. 1, 2001, pp. 1034–40.
22. S.-Y. Cho, T. W. S. Chow, and C.-T. Leung, "A neural-based crowd estimation by hybrid global learning algorithm," IEEE Trans. Syst, Man, Cybern., vol. 29, pp. 535–41, 1999.
23. C. S. Regazzoni and A. Tesei, "Distributed data fusion for real-time crowding estimation," Signal Process., vol. 53, pp. 47–63, 1996.
24. A. C. Davies, J. H. Yin, and S. A. Velastin, "Crowd monitoring using image processing," Electron. Comm. Eng. J., vol. 7, pp. 37–47, 1995.
25. D. Kong, D. Gray, and H. Tao, "Counting pedestrians in crowds using viewpoint invariant training," in British Machine Vision Conf., 2005.