

Two Person Interaction Recognition System Using SVM

Vaibhav Gawande, Yogesh Swami, Pratik Kakde, Yuvraj Lokhande, Prof. Rakesh Badodekar

BE Student, Department of Information Technology, Sinhgad Institute of Technology, Savitribai Phule Pune
University, Pune, India.

Professor, Department of Computer Engineering, Sinhgad Institute of Technology, Savitribai Phule Pune University,
Pune, India.

ABSTRACT- Human activity recognition has potential to impact a wide range of applications from surveillance to human computer interfaces to content based video retrieval. Recently, the rapid development of inexpensive depth sensors (e.g. Microsoft Kinect) provides adequate accuracy for real-time full-body human tracking for activity recognition applications. In this paper, we create a complex human activity dataset depicting two person interactions, including synchronized video, depth and motion capture data. Moreover, we use our dataset to evaluate various features typically used for indexing and retrieval of motion capture data, in the context of real-time detection of interaction activities via Support Vector Machines (SVMs). Experimentally, we find that the geometric relational features based on distance between all pairs of joints outperform other feature choices. For whole sequence classification, we also explore techniques related to Multiple Instance Learning (MIL) in the sequence is represented by a bag of body-pose features. We find that the MIL based classifier outperforms SVMs when the sequences extend temporally around the interaction of interest.

KEYWORDS: Body-Pose Features, SVM, MIL

I. INTRODUCTION

Human activity recognition is an important field for applications such as surveillance, human-computer interface, content-based video retrieval, etc. Early attempts at human action recognition used the tracks of a person's body parts as input features. However, most recent research moves from the high-level representation of the human body (e.g. skeleton) to the collection of low-level features (e.g. local features) since full-body tracking from videos is still a challenging problem. Recently, the rapid development of depth sensors (e.g. Microsoft Kinect) provides adequate accuracy of real-time full-body tracking with low cost. This enables us to once again explore the feasibility of skeleton based features for activity recognition. Past research proposed algorithms to classify short videos of simple periodic actions performed by a single person (e.g. 'walking' and 'waiving'). In real-world applications, actions and activities are seldom periodic and are often performed by multiple persons (e.g. 'pushing and hand shaking'). Recognition of complex non-periodic activities, especially interactions between multiple persons, will be necessary for a number of applications (e.g. automatic detection of violent activities in smart surveillance systems). In contrast to simple periodic actions, the study of causal relationships between two people, where one person moves, and the other reacts, could help extend our understanding of human motion. In this work, we recognize interactions performed by two people using RGBD (i.e. color plus depth) sensor. Recent work has suggested that human activity recognition accuracy can be improved when using both color images and depth maps. On the other hand, it is known that a human joint sequence is an effective representation for structured motion. Hence we only utilize a sequence of tracked human joints inferred from RGBD images as a feature. It is interesting to evaluate body-pose features motivated from motion capture data using tracked skeletons from a single depth sensor. Since full-body tracking of humans from a single depth sensor

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 3, March 2017

contains incorrect tracking and noise, this problem is somewhat different from scenarios with clean motion capture data. In this paper, we create a new dataset for two-person interactions using an inexpensive RGBD sensor. We collect eight interactions: approaching, departing, pushing, kicking, punching, exchanging objects, hugging, and shaking hands from seven participants and 21 pairs of two-actor sets. In our dataset, color-depth video and motion capture data have been synchronized and annotated with action label for each frame. Real time human activity detection has multiple uses from human computer interaction systems, to surveillance, to gaming. However, non-periodic actions do not always have a clearly defined beginning and ending frame. Since recorded sequences are manually segmented and labeled in training data, a segmented sequence might contain irrelevant actions or sub-actions. To overcome this problem, we use the idea of Multiple Instance Learning (MIL) to tackle irrelevant actions in whole sequence classification. We find that classifiers based on Multiple Instance Learning, have much higher classification accuracy when the training sequences contain irrelevant actions than Support Machine (SVM) classifiers.

II. LITERATURE SURVEY

1. N. Ikizler-Cinbis and S. Sclaroff. Object, scene and actions: Combining multiple features for human action recognition. In ECCV, 2010.

In many cases, human actions can be identified not only by the singular observation of the human body in motion, but also properties of the surrounding scene and the related objects. In this paper, we look into this problem and propose an approach for human action recognition that integrates multiple feature channels from several entities such as objects, scenes and people. We formulate the problem in a multiple instance learning (MIL) framework, based on multiple feature channels. By using a discriminative approach, we join multiple feature channels embedded to the MIL space.

2. S. Ali and M. Shah. Human action recognition in videos using kinematic features and multiple instance learning. IEEE TPAMI, 32(2):288303, Feb. 2010.

We propose a set of kinematic features that are derived from the optical flow for human action recognition in videos. The set of kinematic features includes divergence, vorticity, symmetric and anti-symmetric flow fields, second and third principal invariants of flow gradient and rate of strain tensor, and third principal invariant of rate of rotation tensor. Each kinematic feature, when computed from the optical flow of a sequence of images, gives rise to a spatiotemporal pattern. It is then assumed that the representative dynamics of the optical flow are captured by these spatiotemporal patterns in the form of dominant kinematic trends or kinematic modes.

3. A. Patron-Perez, M. Marszalek, A. Zisserman, and I. Reid. High five: Recognising human interactions in tv shows. In BMVC, 2010.

In this paper we address the problem of recognising interactions between two people in realistic scenarios for video retrieval purposes. We develop a per-person descriptor that uses attention (head orientation) and the local spatial and temporal context in a neighbourhood of each detected person. Using head orientation mitigates camera view ambiguities, while the local context, comprised of histograms of gradients and motion, aims to capture cues such as hand and arm movement. We also employ structured learning to capture spatial relationships between interacting individuals.

4. J. Aggarwal and M. Ryoo. Human activity analysis: A review. In ACM Computing Surveys, 2011.

This article provides a detailed overview of various state-of-the-art research papers on human activity recognition. We discuss both the methodologies developed for simple human actions and those for high-level activities. Approach based taxonomy is chosen that compares the advantages and limitations of each approach. Recognition methodologies for an analysis of the simple actions of a single person are first presented in the article.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 3, March 2017

5. M. Alcoverro, A. Lopez-Mendez, M. Pardas, and J. Casas. Connected operators on 3D data for human body analysis. In CVPR Workshops, 2011.

This paper presents a novel method for filtering and extraction of human body features from 3D data, either from multi-view images or range sensors. The proposed algorithm consists in processing the geodesic distances on a 3D surface representing the human body in order to find prominent maxima representing salient points of the human body. We introduce a 3D surface graph representation and filtering strategies to enhance robustness to noise and artifacts present in this kind of data. We conduct several experiments on different datasets involving 2 multi-view setups and 2 range data sensors: Kinect and Mesa SR4000. In all of them, the proposed algorithm shows a promising performance towards human body analysis with 3D data.

6. G. Guerra-Filho and A. Biswas. A human motion database: The cognitive and parametric sampling of human motion. In FG, 2011.

In this paper, we describe the construction of our Human Motion Database using controlled sampling methods (parametric and cognitive sampling) to obtain the structure necessary for the quantitative evaluation of several motion-based research problems. The Human Motion Database is organized into several components: the praxicon dataset, the cross-validation dataset, the generalization dataset, the compositionality dataset, and the interaction dataset.

7. H. Kuehne, H. Jhuang, E. Garrote, T. Poggio, and T. Serre. HMDB: A large video database for human motion recognition. In ICCV, 2011.

With nearly one billion online videos viewed every day, an emerging new frontier in computer vision research is recognition and search in video. While much effort has been devoted to the collection and annotation of large scalable static image datasets containing thousands of image categories, human action datasets lag far behind. Current action recognition databases contain on the order of ten different action categories collected under fairly controlled conditions.

8. T. Leung, Y. Song, and J. Zhang. Handling label noise in video classification via multiple instance learning. In ICCV, 2011.

In this paper, we propose the use of MILBoost in a large-scale video taxonomic classification system comprised of hundreds of binary classifiers to handle noisy training data. We test on data with both artificial and real-world noise and compare against the state-of-the-art classifiers based on AdaBoost. We also explore the effect of different bag sizes on different levels of noise on the final classifier performance.

9. Y. Liu, C. Stoll, J. Gall, H.-P. Seidel, and C. Theobalt. Markerless motion capture of interacting characters using multiview image segmentation. In CVPR, 2011.

We therefore propose combined image segmentation and tracking approach to overcome these difficulties. A new probabilistic shape and appearance model is employed to segment the input images and to assign each pixel uniquely to one person. Thereafter, a single-person markerless motion and surface capture approach can be applied to each individual, either one-by-one or in parallel, even under strong occlusions.

10. J. Sung, C. Ponce, B. Selman, and A. Saxena. Human activity detection from RGBD images. In AAAI workshop on Pattern, Activity and Intent Recognition (PAIR), 2011

In this paper, we use a RGBD sensor (Microsoft Kinect) as the input sensor, and present learning algorithms to infer the activities. Our algorithm is based on a hierarchical maximum entropy Markov model (MEMM). It considers a person's activity as composed of a set of sub activities, and infers the two-layered graph structure using a dynamic programming approach. We test our algorithm on detecting and recognizing twelve different activities performed by four people in different environments, such as a kitchen, a living room, an office, etc., and achieve an average performance of 84.3% when the person was seen before in the training set.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 3, March 2017

III. TWO-PERSON INTERACTION MECHANISM

We collect two person interactions using the Microsoft Kinect sensor. We choose eight types of two-person interactions, motivated by the activity classes from, including: Approaching, departing, pushing, kicking, punching, exchanging objects, hugging, and shaking hands. Note that all of these action categories have interactions between actors that differ from the categories performed by a single actor independently. These action categories are challenging because they are not only non-periodic actions, but also have very similar body movements. For instance, ‘exchanging object’ and ‘shaking hands’ contain common body movements, where both actors extend and then withdraw arms. Similarly, ‘pushing’ might be confused with ‘punching’. All videos are recorded in the same laboratory environment. Seven participants performed activities and the dataset is composed 21 sets, where each set contains videos of a pair of different persons performing all eight interactions. Note that in most interactions, one person is acting and the other person is reacting. Each set contains one or two sequences per action category. The entire dataset has a total of 300 interactions approximately. Both color image and depth map are 640 * 480 pixels.



Figure 1: Visualization of our interaction dataset. Each row per interaction contains a color image, a depth map, and extracted skeletons at the first, 25%, 50%, 75%, and the last frame of the entire sequence for each interaction: approaching, departing, kicking, punching, pushing, hugging, shaking hands, and exchanging. A red skeleton indicates the person who is acting, and a blue skeleton indicates the person who is reacting.

The dataset apart from an image and a depth map also contains 3-dimensional coordinates of 15 joints from each person at each frame. The articulated skeletons for each person are automatically extracted by OpenNI with NITE middleware provided by PrimeSense. The frame rate is 15 frames per second (FPS). The dataset is composed of manually segmented videos for each interaction, but each video roughly starts from a standing pose before acting and ends with a standing pose after acting. Our dataset also contains ground truth labels with each segmented video labeled as one action category. Ground truth label also contains identification of “active” actor (e.g. the person who is punching), and “inactive” actor (e.g. the person being punched). Figure shows example snapshot images of our dataset.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

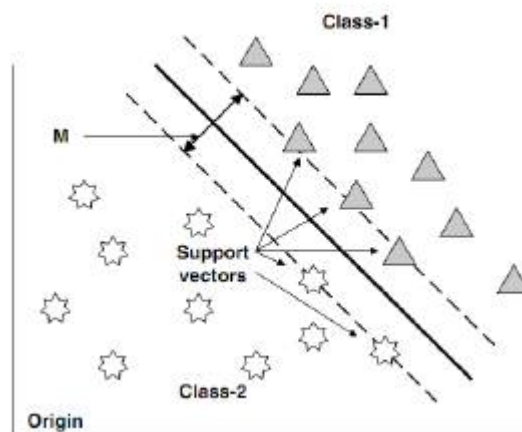
Website: www.ijirset.com

Vol. 6, Issue 3, March 2017

Although the skeleton extraction from depth maps provides a rather accurate articulated human body, it contains noisy and incorrect tracking. Especially, since the full-body tracking by NITE middleware is less stable on fast and complex motions, and occlusions, there often exist tracking failures in our dataset. For example, the position of an arm is stuck in Figure 1e and Figure 1a. The overall tracking is sometimes bad when a large amount of body parts of two persons overlap. More examples can be found in the supplementary material.

IV. SUPPORT VECTOR MACHINE

SVM is a binary classifier to analyze the data and recognize the pattern for classification. The main goal is to design a hyperplane that classifies all the training vectors in different classes. The aim is to determine a function $F(x)$ which obtains the hyperplane. A hyperplane separates two classes of data sets. The linear classifier is defined as the optimal separating hyperplane. The data sets can be separated in two ways: linearly separated or nonlinearly separated.



The vectors are said to be optimally separated if they are separated without error and the distance between the two closest vector points is maximum.

For linear separable data sets, training vectors of different class of pairs (a_m, b_m)

Where $m = 1, 2, 3, 4, \dots, t$

$$a_m \in R_n$$

$$b_m \in \{+1, -1\}$$

The decision boundary is placed using maximal margin between the closest points. w being a vector perpendicular median to the street. a be the unknown of to be placed in particular class according to the decision boundary.

And hyper plane $(w \cdot a) + c = 0$ with c as constant

For classification

$$(w \cdot a_m) + c_0 \geq 1, \quad \forall b_m = +ve \text{ samples}$$

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 3, March 2017

$$(w \cdot a_m) + c_0 \leq -1, \quad \forall b_m = -ve \text{ samples}$$

where $(w \cdot a_m)$ shows dot product of w and a_m .

The inequalities if added i.e multiplying eqⁿ (1) and (2) with +1, -1, and b_m .

suppose b_m such that $b_m = 1$ for +ve samples

$$b_m = -1 \text{ for -ve samples}$$

We get,

$$b_m [(w \cdot a_m) + c_0] \geq 1$$

$$b_m [(w \cdot a_m) + c_0] \geq 1$$

Therefore rearranging the above equations

$$b_m (w \cdot a_m) + c_0 - 1 \geq 0$$

for points into dataset to in the gutter i.e on the decision boundary

$$b_m (w \cdot a_m) + c_0 - 1 = 0$$

The SVM classifier places the decision boundary by using maximal margin among all possible hyper planes. To maximize the M the margin between the hyper plane. $\|w\|$ width of the street i.e hyperplane has to minimize. the unit vector for width is given as

$$\text{Width} = (a_+ - a_-) \cdot \frac{w}{\|w\|}$$

$\|w\|$ is magnitude, a_+, a_- are the coordinates in the hyper plane

Dot product is scalar. To maximize, for coordinates (1,-1) the width = $\frac{2}{\|w\|}$.

We can conclude, we need to max($\frac{1}{\|w\|}$) i.e min $\|w\|$ i.e min $\frac{\|w\|^2}{2}$ done for mathematical convenience.

When we use extremist i.e min and max with constraints like constant b , the problem is called quadratic optimization problem. Hence to solve it we need langrange function. Let the langrange multiplier be α_k ,

a vector defining the boundary is w , is the input data vector points. C is scalar threshold value. α_k be the obtained when function L is minimized w.r.t w, c and to get maximize value, we need to differentiate function L w.r.t w, c and equating it with zero shown in equation below:

$$L = \frac{1}{2} \|w\|^2 - \sum \alpha_k [b_m (w \cdot a_m) + c_0 - 1]$$

$$\frac{\partial L}{\partial w} = w - \sum \alpha_k a_m b_m = 0$$

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 3, March 2017

Rearranging the above equation,

$$w = \sum \alpha_k a_m b_m$$

i.e the linear sum of all the samples.

Here, there is an advantage of support vector machine over ANN with back propagation. In SVM there is one min solution only. Hence, in SVM local minima problem is solved. On other hand in back propagation ANN, there occur this local minima problem.

$$\frac{\partial L}{\partial c} = - \sum \alpha_k b_m = 0$$

$$\sum \alpha_k b_m = 0$$

L function is changed to the dual Langrangian (∂L) should be maximized with respect to non-negative. α_k ∂L function is given as below

$$\max \partial L = - \frac{1}{2} \sum_m \sum_k \alpha_k \alpha_m a_m b_m a_k b_k$$

Equating the value of w in langrange function L, we get

$$L = \sum \alpha_k - \frac{1}{2} \sum_m \sum_k \alpha_k \alpha_m a_m b_m a_k b_k$$

The above equation shows standard quadratic optimization problem. For dual optimization problem, the most suitable solution is with parameters w and c of hyperplanew.r.t α_k . So the optimized hyperplane solution is given as:

$$F(x) = \sum_{m=1}^t \alpha_m b_m (a_m \cdot a) + c \geq 0$$

V. EXPERIMENTS

1. Experiments for Realtime interaction detection

The features defined in Section 4.1 can be divided into three groups: two joint features, two plane features, and two velocity features. We evaluate which group of features are the most appropriate for real-time two-person interaction detection. 30 joints (i.e. 15 joints of each person) are used for joint features. Both plane features and velocity features are much higher dimension vectors. For this reason, we choose ten markers Blue rectangles indicate true positive instances and red rectangles indicate true negative instances. If a positive bag has at least one instance is positive (i.e. actual kicking action), while a negative bag does not have any of kicking action. The MILBoost classifier outputs an action label given a test sequence.

‘head’, ‘elbows’, ‘hands’, ‘knees’ and ‘feet’) for a target joint (i.e. i in Equation 1 and 2), while selecting only six important markers (i.e. ‘torso’, ‘head’, ‘hands’ and ‘feet’) for reference planes or vectors. By doing so, we create a lower dimension feature without losing meaningful information. However, both plane features and velocity features have

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 3, March 2017

very high dimension (i.e. $800 \times W$ for plane and velocity, even higher for normal velocity). All body-pose features are computed within $W=3$ (0.2 seconds). To classify eight action categories, we train SVMs in a one-vs-all fashion, and evaluation is done by 5-fold cross validation, i.e. 4 folds are used for training, and 1 for testing. The dataset is composed by 21 sets of two actors. We randomly split the dataset into 5 folds of 4-5 two-actor sets each. The partitioning of the dataset into folds is performed so that each two-actor set is guaranteed to appear only in training or only in testing. Table 1 shows the real-time activity detection accuracy of eight complex human-human interactions from our dataset. The results are averaged over the 5 permutations and the parameter selection of SVMs is done by nested cross validation with cost $C \in \{0.01, 0.1, 1, 10, 100\}$. We also evaluated with a non-linear kernel for the SVM, but it yielded very little, if any, improvement for most of features since our body-pose features are high dimensional. The result shows joint features result in higher detection accuracy than plane features and velocity features. We conclude that the geometric relational body-pose feature based on Euclidean distance between joints captures the temporal and dynamic information of body movement for complex human activities in the real-time action detection scenario. Also, it is more stable with noisy full-body tracking than velocity features. Yao et al. pointed out velocity features have the best accuracy or single activity recognition using clean motion captured data with reduced skeleton (=13 joints). However, they also claimed that velocity features are not robust to noise by using synthesized noisy skeleton. As we have seen in Section 3, our dataset contains a significant amount of incorrect tracking and noise, which might explain the lower accuracy of velocity features. Note that the raw position feature uses the position of all 30 joints at a frame. The low accuracy using the raw position feature means that the actions in the dataset are difficult to classify.

2. Experiments on Whole Action Sequence Classification

In this section, we compare the classification performance between the MILBoost classifier and SVM classifier with segmented videos. Note that this experiment is not real-time detection. We compute joint distance feature for each frame and the size of window is the maximum frame size of a sequence. For MILBoost classifier, the size of the bag is the same as the window size for SVM classifiers. In SVMs each frame is treated equally during learning, while in MILBoost frames are re-weighted so that keyframes for the interaction receive larger weights. For example, a 'punching' classifier gives higher weights the feature of the frame, where the arm of 'active' actor is stretched and almost hit the head of 'inactive' actor. Since the lower weights will be assigned to the frames having irrelevant actions, the classification performance increases. To evaluate this, we first use the sequences with our ground truth labels in the training set, called Set 1. To amplify the effect of irrelevant actions in the training set, we create a different training set, called Set 2, with more irrelevant actions. Specifically, we segment the original recorded sequence by starting from five frame earlier than the original start frame and ending five frame later than the original final frame. Set 2 contains more irrelevant actions since participants randomly moved between action categories when we collected data. On Set 1, MILBoost has slightly better performance than SVMs. However, we found accuracy dramatically dropped with SVMs from Set 1 to Set 2, while MILBoost retain high classification accuracy with training data including more irrelevant actions. We conclude the MILBoost classifier outperforms SVMs if there exist irrelevant actions in the training set.

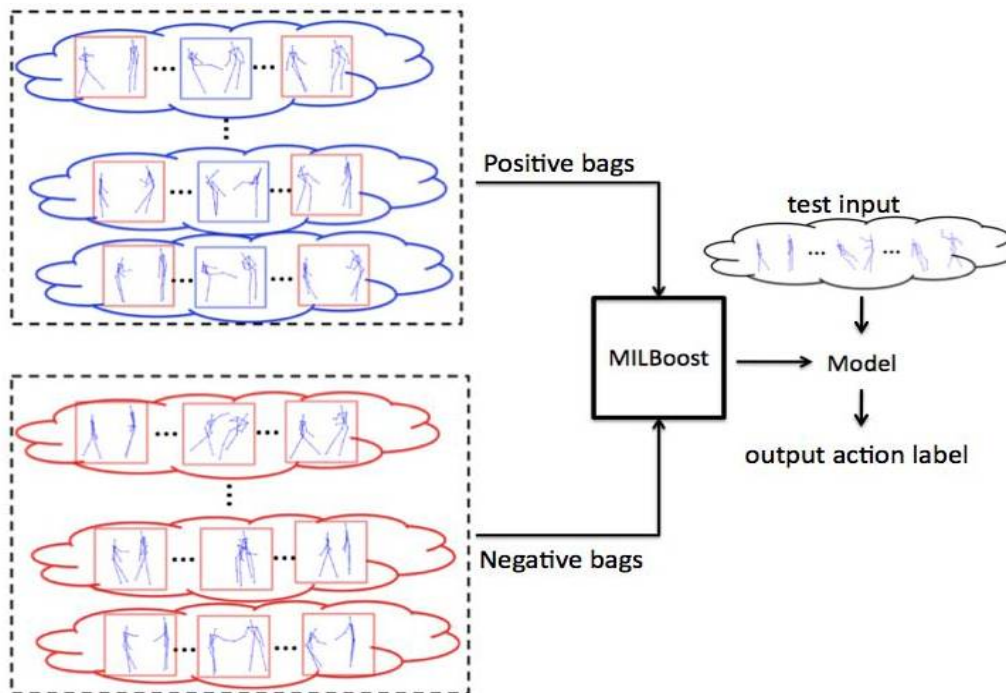


Figure 2: The overview of the 'kicking' MILBoost classifier.

VI. CONCLUSION

We have created a new dataset for two-person interaction using the Microsoft Kinectsensor including eight interactions, color-depth video and motion capture data ateach frame. Using this dataset, we have evaluated body-pose features motivated from 3D skeleton features for indexing and retrieval of motion capture data. Geometric relational features based on distance between all pairs of joints (i.e. joint features) outperformed other features for real-time interaction detection on noisy 3D skeleton data. Moreover, we have shown that the MIL Boost classifier outperforms SVMs if there exist irrelevant actions in the training data.

REFERENCES

- [1] A. Nelson, "The UK approach to managing risk at passive level crossings,"in Proc. 7th Int. Symp. RailRoad-Highway Grade Crossing Res. Safety, 2002, pp. 1–11.
- [2] E. Griffioen, "Improving level crossings using findings from human behaviour studies," in Proc. 8th Int. Level Crossing Symp., Sheffield, U.K., 2004.
- [3] [Online]. Available: http://www.uic.org/cdrom/2008/11_wcr2008/pdf/G_2.4.5.4.pdf
- [4] L. Khoudour, M. Ghazel, F. Boukour, M. Heddebaut, and M. El-Koursi, "Towards safer level crossings: Existing recommendations, new applicable technologies and a proposed simulation model," Eur. Transp. Res. Rev., vol. 1, no. 1, pp. 35–45, Mar. 2009.
- [5] J. J. Garcia et al., "Efficient multisensory barrier for obstacle detection on railways," IEEE Trans. Intell. Transp. Syst., vol. 11, no. 3, pp. 702–713, Sep. 2010.
- [6] E. Schnieder, R. Slovak, E. M. El Koursi, L. Tordai, and M. Woods, "A European contribution to level crossing safety," in Proc. FOVUS, Stuttgart, Germany, Sep. 2008, pp. 222–228.
- [7] Government's IT Strategy Headquarters: New IT Reform Strategy, The Realization of Society, inWhichEveryone Can Benefit From IT, Anytime and Anywhere, Jan. 19 2006, 19.
- [8] L. Khoudour et al., "PANsafer project: Towards a safer level crossing,"presented at the 11th Level Crossing Symposium, Tokyo, Japan, 2010.



ISSN(Online) : 2319-8753
ISSN (Print) : 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 3, March 2017

- [9] N. Fakhfakh et al., "Background subtraction and 3D localization of moving and stationary obstacles at level crossings," in Proc. Int. Conf. Image Process. Theory, Tools Appl., Paris, France, 2010, pp. 72–78.
- [10] A. Prioletti et al., "Part-based pedestrian detection and feature-based tracking for driver assistance: Real-time, robust algorithms, and evaluation," IEEE Trans. Intell. Transp. Syst., vol. 14, no. 3, pp. 72–74, Sep. 2013.
- [11] B. U. Toreyin, A. E. Cetin, A. Aksay, and M. B. Akhan, "Moving object detection in wavelet compressed video," Signal Process., Image Commun., vol. 20, no. 3, pp. 255–264, Mar. 2005.
- [12] B. Han, D. Comaniciu, and L. S. Davis, "Sequential Kernel density approximation through mode propagation: Applications to background modeling," in Proc. Asian Conf. Comput. Vis., Jan. 2004, pp. 818–823.
- [13] N. M. Oliver, B. Rosario, and A. P. Pentland, "A Bayesian computer vision system for modeling human interactions," IEEE Trans. Pattern Anal. Mach. Intell., vol. 22, no. 8, pp. 831–843, Aug. 2000.
- [14] A. R. J. Francois, "Real-time multi-resolution blob tracking," Univ. Southern California, Los Angeles, CA, USA, IRIS Tech. Rep. IRIS-04- 422, 2004.
- [15] A. Blake and M. Isard, Active Contours. New York, NY, USA: Springer- Verlag, 1998.