HDR Imaging for Enchancing People Detection and Tracking in Indoor Environments

Panagiotis Agrafiotis, Elisavet K. Stathopoulou, Andreas Georgopoulos and Anastasios Doulamis

Laboratory of Photogrammetry, School of Rural and Surveying Engineering, National Technical University of Athens, 15780 Zografou Athens, Greece

Keywords: HDR Imaging, People Detection, People Tracking.

Abstract: Videos and image sequences of indoor environments with challenging illumination conditions often capture either brightly lit or dark scenes where every single exposure may contain overexposed and/or underexposed regions. High Dynamic Range (HDR) images contain information that standard dynamic range ones, often mentioned also as low dynamic range images (SDR/LDR) cannot capture. This paper investigates the contribution of HDR imaging in people detection and tracking systems. In order to evaluate this contribution of the HDR imaging in the accuracy and robustness of pedestrian detection and tracking in challenging indoor visual conditions, two state of the art trackers of different complexity were implemented. To this direction data were collected taking into account the requirements and real-life indoor scenarios and HDR frames were produced. The algorithms were applied to the SDR data and their corresponding HDR data and were compared and evaluated for their robustness and accuracy in terms of precision and recall. Results show that that the use of HDR images enhances the performance of the detection and tracking scheme, making it robust and more reliable.

1 INTRODUCTION

Videos and image sequences of indoor environments with challenging illumination conditions often capture brightly lit or dark scenes where every single exposure may contain overexposed and/or underexposed regions. This common phenomenon is usually the result of human constructions (buildings, windows etc.) and trees that influence negatively the digital representation of the scene. Essential information loss, noise and saturation degrade image processing and computer vision algorithms that rely on accurate feature detection.

High Dynamic Range (HDR) images contain information that standard dynamic range ones, often mentioned in the bibliography also as low dynamic range images (SDR/LDR) cannot capture. A standard sensor is designed to depict a limited range of intensities at the same frame, in contrast with the human eye adaptation system. Thus, in extreme illumination conditions, overexposed and/or underexposed pixels may occur, causing serious information loss. HDR images can be captured either by specially designed sensors with extended dynamic range or by merging multiple SDR images of the same scene with different exposure time settings (Reinhard et al., 2005).

This study investigates the contribution of HDR imaging in people detection and tracking systems through the implementation of two detection and tracking systems of different complexity. The paper is organised as follows: Section 2 is a review of related previous work concerning HDR imaging and feature detection and tracking algorithms. Section 3 analyses our approach, followed by experimental results with captured datasets in Section 4 and conclusions in Section 5.

2 RELATED WORK

The current state-of-the-art in such computer vision algorithms study constrained visual environments like the ones within a research laboratory or of real-like sequences of one or few objects. To handle these bottlenecks, in the recent years, several research efforts have been also published using complicated visual environments, like the ones of outdoors surveillance of multiple persons (pedestrians), crowded conditions, aerial monitoring, teleconferences rooms, sports events, and daily activities within houses.

Creating an HDR image by merging multiple views of the same scene captured with different exposure values is commonly used in the bibliography.

Agrafiotis P., Stathopoulou E., Georgopoulos A. and Doulamis A..
HDR Imaging for Enchancing People Detection and Tracking in Indoor Environments.
DOI: 10.5220/0005456706230630
In *Proceedings of the 10th International Conference on Computer Vision Theory and Applications* (MMS-ER3D-2015), pages 623-630
ISBN: 978-989-758-090-1

(Debevec and Malik, 1997; Robertson et al., 1999; Rovid et al., 2007). These algorithms use the input images to create the HDR image by computing the camera response function (CRF) given the exposure value.

HDR images are basically radiance maps whose pixels reach a maximum of 32-bit floating point image representation. Common display media can visualise SDR/LDR images in a 16- or 8-bit format but are unable to render properly images in 32-bit format. Various compression operators to convert HDR images to displayable (LDR) ones without serious detail loss have been developed (Reinhard et al., 2002; Durand and Dorsey, 2002; Drago et al., 2003; Mantiuk et al., 2006; Fattal et al., 2007). This process is known as tone-mapping technique. Tone mapping operators can be classified into global and local with respect to their pixel region of impact. Global techniques apply a mapping function to the entire set of pixels, while local ones take into account the characteristics in the neighbourhood of each pixel.

Feature detection and tracking in images and videos is a common problem in computer vision and it has been extensively discussed in the literature during the previews years. First step of these approaches is the extraction of appropriate visual descriptors. The most commonly used detectors and descriptors in vision based applications are SIFT, SURF, GFTT, FAST, ORB and HOG. Scale Invariant Feature Transform (SIFT) (Lowe, 2004) extracts features invariant to image scale, rotation and translation and partially invariant to illumination changes. Speeded Up Robust Features (SURF) (Bay et al., 2008) and Good Features To Track (GFTT) (Shi and Tomasi, 1994), feature detectors are based on the assessment of the Hessian matrix. SURF also combines both detection and description but it outperforms SIFT in terms of speed. GFTT is an improved version of Harris corner detector (Harris and Stephens, 1988). Features from Accelerated Segment Test (FAST) (Rosten and Drummond, 2006) is based on the intensity difference of the candidate centre pixel and its surrounding neighbours lying on a circular ring about the centre. It is benefiting from low computation. The Oriented FAST and Rotated BRIEF (ORB) (Rublee et al., 2011). ORB is a combination of FAST and the recently developed BRIEF descriptor (Calonder et al., 2010), being rotation invariant and resistant to noise. Finally, HOG (Histograms of Oriented Gradients) were initially described by Dalal and Triggs (Dalal and Triggs, 2005) in the context of person detection in images and videos. These descriptors are fed as input to non-linear classifiers to detect the objects and track their position within the space (Liu et al., 2011; Kaaniche and Bremond, 2009; Alahi et al., 2010; Miao et al., 2011).

In addition, great effort has been dedicated to handle object tracking as a classification problem (Avidan, 2004; Lepetit et al., 2005). Some of the first approaches towards an adaptable classification for object tracking have been presented in (Collins et al., 2005; Doulamis et al., 2003). These studies, however, do not face efficiently the general trade-off between model stability and adaptability. Matthews et al. (Matthews et al., 2004), propose an updated template algorithm that avoids the ”drifting” inherent. Other methods exploit the semi-supervised paradigm (Stalder et al., 2009; Grabner et al., 2008), a co-training strategy (Tang et al., 2007), a combination of generative and discriminative trackers (Yu et al., 2008), or finally coupled layered visual models (Cehovin et al., 2011).

The fact that local information sometimes is not enough to make a correct decision led to the development of global optimization trackers. Examples are the use of a minimum cost graph matching that runs the Hungarian algorithm (Cehovin et al., 2011; Huang et al., 2008). Other works handle the problem as a minimum flow cost problem (Zhang et al., 2008; Henriques et al., 2011). However, in case that the background/foreground significantly changes, there is no way to estimate matches for longer time periods forcing the algorithm to fail. The aforementioned approaches exploit Standard Dynamic Range (SDR) images or videos that limit their performance.

Although HDR imaging is a broadly used technique, a few surveys regarding feature tracking on such images have been made. Cui et al. proved that the performance and robustness of SIFT operator on HDR images is superior than using SDR ones (Cui et al., 2011). The work of Ladas et al. (Ladas et al.,) takes advantage of HDR imaging and inverse illumination based on Precomputed Radiance Transfer (PRT) (Sloan et al., 2002). Chermak and Aouf investigated HDR imaging on a larger number of state of the art operators, SIFT, SURF, Harris, Tomasi and FAST, resulting in indeed higher performance than SDR images (Chermak and Aouf, 2012). A subsequent study (Chermak et al., 2014) elaborates further on this field by testing larger image sequences for feature tracking.

The main contribution of this paper is the use of the HDR imagery instead of the SDR already used, in order to eventually improve the performance and the reliability of pedestrian detection and tracking algorithms.

3 APPROACH

3.1 Tested Trackers

In order to evaluate the contribution of the High Dynamic Range imaging in the accuracy and robustness of pedestrian detection and tracking in challenging indoor visual conditions, two state of the art trackers of different complexity were implemented.

3.2 The Complex Tracker

We adopt and modify the tracking scheme presented in (Kokkinos et al., 2013). The algorithm operates in real time, or at least just in time and the computer vision tools have been developed using Intels Integrated Performance Primitives tools, exploiting processor hardware capabilities.

This exploited and modified version of the aforementioned human tracker, integrates on the one hand adaptive background models able to capture slight modifications of the background with, on the other hand, algorithms based on motion characteristics in order to define accurately and precisely the areas of the image to be considered as foreground or background.

In particular, for background modelling, local geometrically enriched mixture models were introduced like the Gaussian Mixture Models (GMMs) in order to deal with the problem of tracking over long periods. These models are exploiting the geometric properties of locally connected regions. Since, such an approach is very sensitive to noise and the alteration of the pixel values due to illumination changes, reducing the background modelling performance, the dimensionality of the hyper-space is reduced using the concept of saliency maps. This approach adopts a graph-based saliency map methodology so that the most important pixels are selected (pixels in which humans feature more than others), and these to be fed as GMM input. This inclusion of local connectivity in modelling the background content increases robustness and tolerance to noise owing either to camera defects and illumination fluctuation. In order to make the algorithm applicable in real-time applications, the background on the on-line captured video frames is re-modelled only in high confident areas of background. These areas are estimated using an iterative motion detection scheme constrained by shape and time properties. In particular, the foreground object is currently detected as a moving object presenting human shape constraints while retaining its continuity in time (temporal coherency). The motion field is detected through an iterative implementation of the Lucas-Kanade Optical Flow method (Lucas et al., 1981) while a constrained shape and time mechanism is developed in order to overcome introduced noise. For accelerating the optical flow algorithm, is adopted an iterative implementation of the algorithm starting with the creation of a stack of different image resolutions (Doulamis, 2010). This is achieved by low-pass filtering of the image content.

However, estimating the motion activity for all pixel of a frame is a time-consuming process while increases the computational complexity of the algorithm, threatening a real-time implementation. Moreover, the application of the aforementioned technique for all pixels would provide a large number of erroneous motion vectors mainly due to the high dynamics of the background content and the complexity of the visual environment. It is also clear that erroneous estimation of motion vectors results in an inaccurate detection, increasing false positives/negatives.

To address this difficulty, the implemented scheme estimates the motion vectors on particularly selected points on the image plane. This is done by detecting good image pixels. In this approach, is implemented the Good Features to Track method of Shi and Tomasi for pixel representation as feature vectors (Shi and Tomasi, 1994). The aforementioned scheme has been also used for activity and behavior recognition (Kosmopoulos et al., 2012; Voulodimos et al., 2014; Voulodimos et al., 2012).

3.3 The HOG-based Tracker

In parallel with the already described implementation, we test a simpler tracker based on Histograms of Oriented Gradients (HOG) (Dalal and Triggs, 2005) and a linear Support Vector Machine (SVM) classifier. A detection and tracking system based on this methodology is sensitive in luminosity conditions and this makes it ideal for testing a possible HDR enchantment for pedestrian detection and tracking.

Subsequently, a detailed review of Dalal and Triggs method is given. The HOG human detector is based on evaluating well-normalized local histograms of image gradient orientations in a dense grid. The main concept of this methodology is that local object appearance and shape can often be characterized rather well by the distribution of local intensity gradients or edge directions, even without precise knowledge of the corresponding gradient or edge positions (Dalal and Triggs, 2005). In order to construct this dense grid of oriented gradient histograms, several stages are implemented. Initially, gamma and colour normalization steps are carried out to reduce the influence of shadows and illumination. The intensity gradient at a pixel is then computed for each colour channel. The next stage is gradient computation at each spatial region or (cell). A range of adjacent cells are grouped into larger spatial blocks so as to normalize edge contrast and illumination. The final step is to concatenate all of the normalized block vectors over each detection window to form a final window descriptor. HOG descriptors are computed over a number of scales on overlapping windows in the image. The high dimension of the HOG descriptors allows a simple linear SVM classifier to successfully classify detection windows into non-pedestrian and pedestrian. In order to train the a linear Support Vector Machine (SVM) classifier, Dalal and Triggs select 1239 of the images as positive training examples, together with their left-right reflections (2478 images in total) (Dalal and Triggs, 2005). This database includes pedestrian in many different standing/walking poses in complex scenes. In addition, these images include indoor and outdoor scenes taken at different times of the day. Nevertheless, the use of HOG descriptors is a memory consuming process because of the high dimension of the descriptors. In addition, the HOG human detector requires the pedestrians to be in upright positions. For poses other than standing and walking, the performance of the HOG human detector deteriorates (Jiang et al., 2010).

4 EXPERIMENTAL RESULTS

4.1 Test Dataset

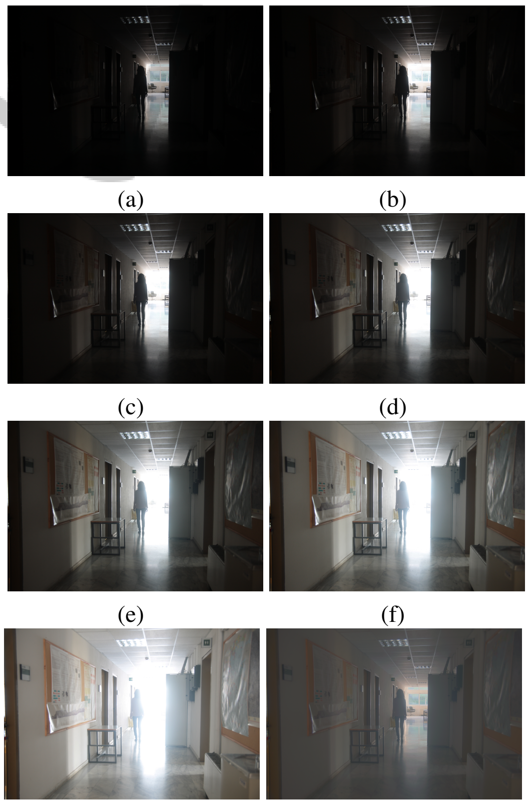
Since there are not any standard datasets including HDR sequences, in order to evaluate the performance of HDR imaging on pedestrians tracking and detection algorithms, data were collected taking into account the requirements and real-life indoor scenarios. More specifically an indoor sequence of a moving person acquired having challenging illumination conditions. In more detail, a person is walking away from the camera, moving firstly almost in parallel with the optical axis and then perpendicular to it. In addition, overexposed and underexposed areas appear in the frames. The sequences were recorded with recording rate 5 fps and 640×430 resolution. Furthermore, other data including pedestrians videos were used in order to assess the detection and tracking performance of the applied algorithms (Agrafiotis et al., 2014b; Agrafiotis et al., 2014a).

(a) (b) (c)

Figure 1: Three consecutive frames of ± 0 exposure level.

4.2 HDR Image Creation

In order to create each HDR image, 7 images were captured with different exposure level $(-3,-2,-1,0,+1,+2,+3)$. Having collected the necessary data, HDR images (radiance maps) were created for each frame implying the algorithm presented in (Debevec and Malik, 1997), which uses the reciprocity law in order to recover the response curve. Since these HDR images are not displayable, Durand and Dorsey (Durand and Dorsey, 2002) operator is used afterwards for the tonemapping process which calculates the corresponding LDR of the HDR image with minimum information loss.



(a) (b) (c) (d) (e) (f) (g) (h)

Figure 2: (a)-(f) The 7 simultaneous images captured with different exposure level $(-3,-2,-1,0,+1,+2,+3)$ respectively and (g) the corresponding HDR image.

This operator decomposes the HDR image into two layers, a base and a detail one, using edge-preserving bilateral filter. To avoid any confusion, from now on we will refer to the original captured data as SDR images, while LDR images will mean the one produced after the tonemapping procedure.

4.3 Tracking Results Comparison and Evaluation

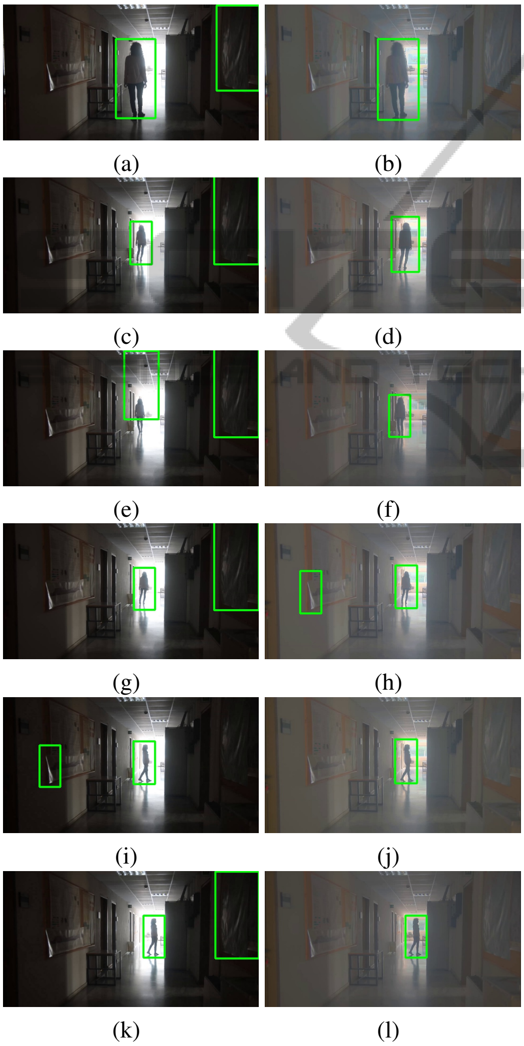


Figure 3: Tracking results from original SDR images of ± 0 level of exposure (left) and from LDR images (right) for HOG based algorithm.

Aiming to evaluate the contribution of the HDR Imaging in detection and tracking systems, initially the algorithms described in Section 3.2 are applied in the original SDR test data presented in Section 4.1 and their corresponding LDR images of HDR data as to be compared and evaluated for their robustness and accuracy in terms of precision and recall.

Figures 3 and 4 present examples of detection and tracking results of the implemented algorithms using original SDR images of ± 0 level of exposure on the left column and using LDR images on the right column. It is clear that the use of HDR images (in fact their corresponding LDR ones) enhances the performance of the detection and tracking scheme making it robust and more reliable. In more detail, HDR imaging in dark indoor visual environments with overexposed and underexposed images areas, improves the accuracy of the tracker by reducing greatly false positives and increasing true positives. As it is observed, in original SDR images are detected in average 90% more false people detections than in LDR images.

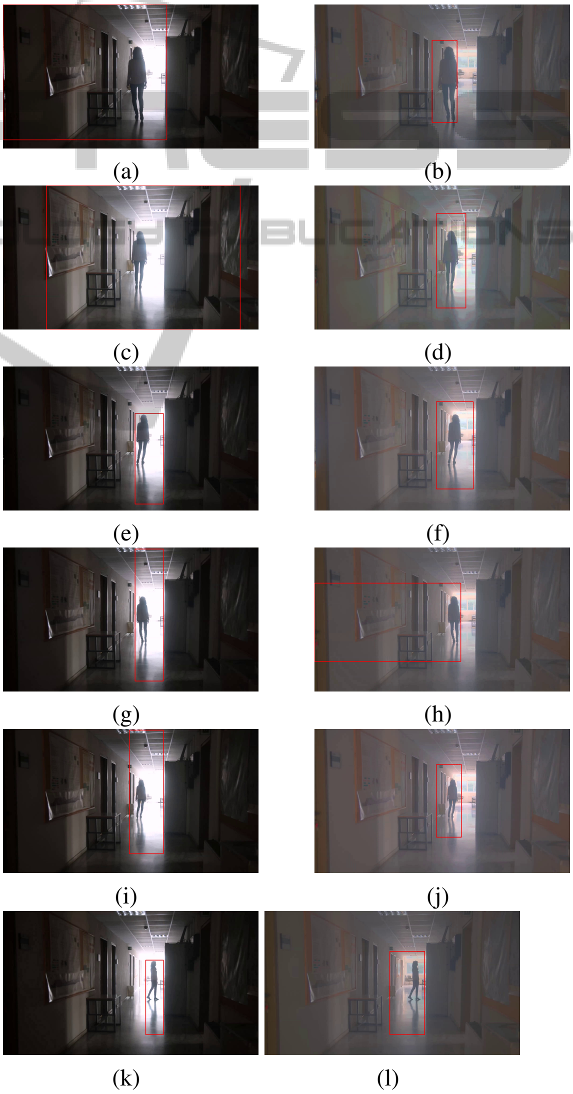


Figure 4: Tracking results from original images of ± 0 level of exposure (left) and from LDR images (right) for the complex algorithm (Kokkinos et al., 2013).

The graph in Figure 5 presents the average precision and recall from all tested data and both implemented detection and tracking schemes. In terms of precision, the SDR dataset achieves 30.3% and the HDR dataset 75%. It seems that the use of HDR images leads in more than doubling the percentage of precision term. Such increase in precision means that the tested algorithms returned substantially more relevant results than irrelevant through the use of HDR data.

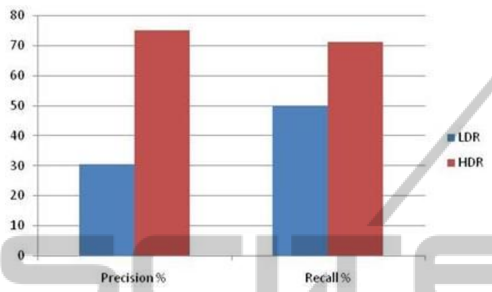


Figure 5: Average precision and recall from all tested data and both implemented detection and tracking schemes.

In terms of recall, the SDR dataset achieves 50% and the HDR dataset 71%. These results denote that the algorithms returned most of the relevant results. As before, significant improvement of the performance of the algorithms after using HDR imagery is observed here.

5 CONCLUSIONS

We have proven and presented that HDR imaging could enhance the performance and the reliability of pedestrian detection and tracking algorithms. Our experimental results show that the HDR imagery increases over 100% the precision term and about 50% the recall term. This technique can reduce the use of IR cameras and consequently reduce the cost of such systems.

Our future work includes extending our experiments to implement more state of the art trackers of different complexity combining them with more HDR creation algorithms. In addition, our future work includes capturing datasets from outdoor environments with reduced illumination, foggy or snowy areas where the latter are characterized for containing overexposed and underexposed regions. Finally, sequences from urban environments are to be tested in order to deal with the problem of building shadows and their underexposed areas.

ACKNOWLEDGEMENTS

The research leading to these results has been supported by European Union funds and National funds (GSRT) from Greece and EU under the project JASON: Joint synergistic and integrated use of eArth obServation, navigatiOn and commuNication technologies for enhanced border security funded under the cooperation framework. The contribution of Ms. Elisavet K. Stathopoulou has been supported by the European Unions Seventh Framework Programme for research, technological development and demonstration under grant agreement no 608013, titled ”ITN-DCH: Initial Training Network for Digital Cultural Heritage: Projecting our Past to the Future.”

REFERENCES

Agrafiotis, P., Doulamis, A., Doulamis, N., and Georgopoulos, A. (2014a). Multi-sensor target detection and tracking system for sea ground borders surveillance. In *Proceedings of the 7th International Conference on PErvasive Technologies Related to Assistive Environments*, page 41. ACM.

Agrafiotis, P., Georgopoulos, A., Doulamis, A. D., and Doulamis, N. D. (2014b). Precise 3d measurements for tracked objects from synchronized stereo-video sequences. In *Advances in Visual Computing*, pages 757–769. Springer.

Alahi, A., Vandergheynst, P., Bierlaire, M., and Kunt, M. (2010). Cascade of descriptors to detect and track objects across any network of cameras. *Computer Vision and Image Understanding*, 114(6):624–640.

Avidan, S. (2004). Support vector tracking. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(8):1064–1072.

Bay, H., Ess, A., Tuytelaars, T., and Van Gool, L. (2008). Speeded-up robust features (surf). *Computer vision and image understanding*, 110(3):346–359.

Calonder, M., Lepetit, V., Strecha, C., and Fua, P. (2010). Brief: Binary robust independent elementary features. In *Computer Vision–ECCV 2010*, pages 778–792. Springer.

Cehovin, L., Kristan, M., and Leonardis, A. (2011). An adaptive coupled-layer visual model for robust visual tracking. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 1363–1370. IEEE.

Chermak, L. and Aouf, N. (2012). Enhanced feature detection and matching under extreme illumination conditions with a hdr imaging sensor. In *Cybernetic Intelligent Systems (CIS), 2012 IEEE 11th International Conference on*, pages 64–69. IEEE.

Chermak, L., Aouf, N., and Richardson, M. (2014). HDR imaging for feature tracking in challenging visibility scenes. *Kybernetes*, 43(8):1129–1149.

Collins, R. T., Liu, Y., and Leordeanu, M. (2005). On-line selection of discriminative tracking features. *Pat-*

tern Analysis and Machine Intelligence, IEEE Transactions on, 27(10):1631–1643.

Cui, Y., Pagani, A., and Stricker, D. (2011). Robust point matching in hdri through estimation of illumination distribution. In Mester, R. and Felsberg, M., editors, *DAGM-Symposium*, volume 6835 of *Lecture Notes in Computer Science*, pages 226–235. Springer.

Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893. IEEE.

Debevec, P. E. and Malik, J. (1997). Recovering high dynamic range radiance maps from photographs. In *Proceedings of the 24th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '97, pages 369–378, New York, NY, USA. ACM Press/Addison-Wesley Publishing Co.

Doulamis, A., Doulamis, N., Ntalianis, K., and Kollias, S. (2003). An efficient fully unsupervised video object segmentation scheme using an adaptive neural-network classifier architecture. *Neural Networks, IEEE Transactions on*, 14(3):616–630.

Doulamis, N. (2010). Iterative motion estimation constrained by time and shape for detecting persons' falls. In *Proceedings of the 3rd International Conference on PErvasive Technologies Related to Assistive Environments*, page 62. ACM.

Drago, F., Myszkowski, K., Annen, T., and Chiba, N. (2003). Adaptive logarithmic mapping for displaying high contrast scenes. *Computer Graphics Forum*, 22:419–426.

Durand, F. and Dorsey, J. (2002). Fast bilateral filtering for the display of high-dynamic-range images. *ACM Trans. Graph.*, 21(3):257–266.

Fattal, R., Agrawala, M., and Rusinkiewicz, S. (2007). Multiscale shape and detail enhancement from multi-light image collections. *ACM Transactions on Graphics (Proc. SIGGRAPH)*, 26(3).

Grabner, H., Leistner, C., and Bischof, H. (2008). Semi-supervised on-line boosting for robust tracking. In *Computer Vision–ECCV 2008*, pages 234–247. Springer.

Harris, C. and Stephens, M. (1988). A combined corner and edge detector. In *Alvey vision conference*, volume 15, page 50. Manchester, UK.

Henriques, J. F., Caseiro, R., and Batista, J. (2011). Globally optimal solution to multi-object tracking with merged measurements. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2470–2477. IEEE.

Huang, C., Wu, B., and Nevatia, R. (2008). Robust object tracking by hierarchical association of detection responses. In *Computer Vision–ECCV 2008*, pages 788–801. Springer.

Jiang, Z., Huynh, D. Q., Moran, W., Challa, S., and Spadaccini, N. (2010). Multiple pedestrian tracking using colour and motion models. In *Digital Image Computing: Techniques and Applications (DICTA), 2010 International Conference on*, pages 328–334. IEEE.

Kaaniche, M. B. and Bremond, F. (2009). Tracking hog descriptors for gesture recognition. In *Advanced Video and Signal Based Surveillance, 2009. AVSS'09. Sixth IEEE International Conference on*, pages 140–145. IEEE.

Kokkinos, M., Doulamis, N. D., and Doulamis, A. D. (2013). Local geometrically enriched mixtures for stable and robust human tracking in detecting falls. *Int J Adv Robotic Sy*, 10(72).

Kosmopoulos, D. I., Doulamis, N. D., and Voulodimos, A. S. (2012). Bayesian filter based behavior recognition in workflows allowing for user feedback. *Computer Vision and Image Understanding*, 116(3):422–434.

Ladas, N., Chrysanthou, Y., and Loscos, C. Improving tracking accuracy using illumination neutralization and high dynamic range imaging.

Lepetit, V., Lagger, P., and Fua, P. (2005). Randomized trees for real-time keypoint recognition. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 2, pages 775–781. IEEE.

Liu, C., Yuen, J., and Torralba, A. (2011). Sift flow: Dense correspondence across scenes and its applications. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(5):978–994.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110.

Lucas, B. D., Kanade, T., et al. (1981). An iterative image registration technique with an application to stereo vision. In *IJCAI*, volume 81, pages 674–679.

Mantiuk, R., Myszkowski, K., and Seidel, H.-P. (2006). A perceptual framework for contrast processing of high dynamic range images. *ACM Trans. Appl. Percept.*, 3(3):286–308.

Matthews, I., Ishikawa, T., and Baker, S. (2004). The template update problem. *IEEE transactions on pattern analysis and machine intelligence*, 26(6):810–815.

Miao, Q., Wang, G., Shi, C., Lin, X., and Ruan, Z. (2011). A new framework for on-line object tracking based on surf. *Pattern Recognition Letters*, 32(13):1564–1571.

Reinhard, E., Stark, M., Shirley, P., and Ferwerda, J. (2002). Photographic tone reproduction for digital images. In *ACM Transactions on Graphics (TOG)*, volume 21, pages 267–276. ACM.

Reinhard, E., Ward, G., Pattanaik, S., and Debevec, P. (2005). *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting (The Morgan Kaufmann Series in Computer Graphics)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.

Robertson, M. A., Borman, S., and Stevenson, R. L. (1999). Dynamic range improvement through multiple exposures. In *In Proc. of the Int. Conf. on Image Processing (ICIP99*, pages 159–163. IEEE.

Rosten, E. and Drummond, T. (2006). Machine learning for high-speed corner detection. In *Computer Vision–ECCV 2006*, pages 430–443. Springer.

Rovid, A.and Varkonyi-Koczy, A., Hashimoto, T., Balogh, S., and Shimodaira, Y. (2007). Gradient based synthesized multiple exposure time hdr image. In *Instrumentation and Measurement Technology Conference Proceedings, IMTC 2007*, pages 1–6.

Rublee, E., Rabaud, V., Konolige, K., and Bradski, G. (2011). Orb: an efficient alternative to sift or surf. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2564–2571. IEEE.

Shi, J. and Tomasi, C. (1994). Good features to track. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on*, pages 593–600. IEEE.

Sloan, P.-P., Kautz, J., and Snyder, J. (2002). Precomputed radiance transfer for real-time rendering in dynamic, low-frequency lighting environments. In *ACM Transactions on Graphics (TOG)*, volume 21, pages 527–536. ACM.

Stalder, S., Grabner, H., and Van Gool, L. (2009). Beyond semi-supervised tracking: Tracking should be as simple as detection, but not simpler than recognition. In *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*, pages 1409–1416. IEEE.

Tang, F., Brennan, S., Zhao, Q., and Tao, H. (2007). Co-tracking using semi-supervised support vector machines. In *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, pages 1–8. IEEE.

Voulodimos, A. S., Doulamis, N. D., Kosmopoulos, D. I., and Varvarigou, T. A. (2012). Improving multi-camera activity recognition by employing neural network based readjustment. *Applied Artificial Intelligence*, 26(1-2):97–118.

Voulodimos, A. S., Kosmopoulos, D. I., Doulamis, N. D., and Varvarigou, T. A. (2014). A top-down event-driven approach for concurrent activity recognition. *Multimedia Tools and Applications*, 69(2):293–311.

Yu, Q., Dinh, T. B., and Medioni, G. (2008). Online tracking and reacquisition using co-trained generative and discriminative trackers. In *Computer Vision–ECCV 2008*, pages 678–691. Springer.

Zhang, L., Li, Y., and Nevatia, R. (2008). Global data association for multi-object tracking using network flows. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE.