

Spatio-Temporal Approaches for Handling Occlusions Based on Object Tracking

Vethamani SE* and Diale D

¹Department of Computer Science and Engineering, AP/CSE, Chandy College of Engineering, Thoothukudi District, Tamil Nadu, India

²Department of Computer Science and Engineering, AP/CSE, Dr. GU Pope College of Engineering, Thoothukudi District, Tamil Nadu, India

Abstract

This paper proposes two different approaches to tracks two occlusion targets. Occlusion means hiding an object by another object. In this paper, two approaches are devised to handle the occlusion detection. In a video, during the tracking process, the states of the object are gradually adjusted one by one to eliminate the occlusion effects. The edges are detected and identified whether the occlusion is partial or full. Spatial approach is used to handle partial occlusion and temporal approach is used to handle full occlusion detection. The spatial approach uses edge-based restoration scheme that 1) identifies areas that are likely to contain wrong motion vectors, 2) finds artifacts within these areas, and 3) restores these artifacts. The temporal approach is planned to fill the missing parts from the past history of a person if available.

Keywords: Occlusion; Spatial approach; Temporal approach

Introduction

In multiple human tracking, one of the most challenging issues is occlusion detection and handling. A number of techniques have been developed in recent years to deal with this problem directly or indirectly. The proposed system is used to detect and handle occlusion. The first challenge is how to robustly determine the portion of the target that is occluded. Determining occlusion status is very hard for general purpose trackers, where the only knowledge available on the target is its initial appearance, and the camera itself could have arbitrary motions. Both spatial and temporal techniques are proposed a solution to this problem the novelty of edge-based detection algorithm consists of the approximation of the incoming edges, with circle arcs, the use of the spatial order of edges, and the directional interpolation scheme that restores missing areas parallel to the covered edge. In the temporal approach, the occlusion is handled by matching the current frame with previous frame, image region that yields the maximal likelihood. Section 2 describes the related work of spatial and temporal occlusion handling technique. Section 3 and section 4 describes the proposed work. Section 3 depicts preprocessing work. Section 4 describes occlusion handling approaches in which spatial method uses edge based restoration and temporal method uses frame matching algorithm. Section 5 draws conclusions and indications on future work.

Related Work

Various approaches that analyze several spatial and temporal occlusion detection approaches address the problem of filling in missing data from different points of view. In the following, a short categorized overview is given that presents the most popular approaches. In Extracting spatio-temporally consistent segments from a video sequence [1] introduces an iterative optimization scheme by first initializing segmentation maps for each frame independently, and then link the correspondences among different frames and iteratively refine them with the collected statistics, so that a set of spatio-temporally consistent volume segments are finally achieved. In the random subspace method (RSM) [2] is chosen for building a classifier ensemble robust against partial occlusions. The component classifiers are chosen on the basis of their individual and combined performance. The main contribution of this work lies in our approach's capability to improve the detection rate when partial occlusions are present without compromising the detection performance on non-occluded data.

In Tracking Pedestrians Using Local Spatio-Temporal Motion Patterns [3] predict the next local spatio-temporal motion pattern a tracked pedestrian will exhibit based on the observed frames of the video. In Texture-Based Restoration [4] the nonparametric texture synthesis algorithm based on Markov random fields. Their approach restores pixel based on the similarity between their local neighborhood and the surrounding neighborhoods. From the candidate neighborhood, one is randomly selected and the value of its central pixel is pasted at current location, a process which is able to intelligently imitate the natural randomness of textures. In Structure-Based Restoration [5] the algorithm recovering missing blocks in video transmission. It uses only the information existing in the same frame, by making a "sketch" of the edges around the missing blocks. The approach which is based on joint interpolation of the image gray-levels and gradient/isophotes directions, smoothly extending in an automatic fashion the isophote lines into the holes of missing data. This interpolation is computed by solving the variation problem via its gradient descent flow, which leads to a set of coupled second order partial differential equations, one for the gray-levels and one for the gradient orientations. Applications of this technique include the restoration of old photographs and removal of superimposed text like dates, subtitles, or publicity is discussed on restoration based on partial differential equations and variation methods [6,7].

For the improved tracking algorithm [8] occlusion layers are introduced to represent occlusion relation and the non-occlusion parts of the persons are obtained according to the occlusion relation and used for tracking. For the non-occlusion persons, each one can be labeled as one patch and for the occlusion persons, more than one person could be labeled as one patch and used for tracking. In Ref. [9], a

***Corresponding author:** Vethamani SE, Department of Computer Science and Engineering, AP/CSE, Chandy College of Engineering, Thoothukudi District, Tamil Nadu, India, Tel: +919791589181; E-mail: ezramani@gmail.com

Received September 02, 2017; **Accepted** September 21, 2017; **Published** October 04, 2017

Citation: Vethamani SE, Diale K (2017) Spatio-Temporal Approaches for Handling Occlusions Based on Object Tracking. J Comput Sci Syst Biol 10: 093-097. doi:10.4172/jcsb.1000256

Copyright: © 2017 Vethamani SE, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

symmetric patch-Based correspondence model for occlusion handling is introduced. The system proposes two new methods called soft constraints and re-segmenting. In soft constraints, systems segment the reference image in two levels: the coarse and fine levels. In resegmenting phase, extract the occluded parts from the entire mixed segment, now this new matching process can make the correct disparity assigned to the new segment.

In Ref. [10], a local best match authentication (LBMA) algorithm is devised to handle complete occlusions, so that it achieves a much more trustworthy detection of the end of an arbitrarily long complete occlusion. The tracking solution could be further improved by “softening” the outlier map so that a smooth transition from non-occlusion to occlusion could be realized. In Ref. [11], it presents a novel method for cost aggregation and occlusion handling for stereo matching. In order to estimate optimal cost, given a per-pixel difference image as observed data, we define an energy function and solve the minimization problem by solving the iterative equation with the numerical method. Tracking rigid objects in image sequences, using template matching. In essence, object tracking is the process of updating object attributes over time. The complete set of attributes includes position, motion, shape, and appearance. The appearance is comprised of a set of photometric features representing the object region in a frame. A new template updating algorithm that satisfies the two qualities: simplicity and robustness. Simplicity implies that the algorithm is easy to implement and has the minimum number of parameters. Robustness implies the ability of the algorithm to track objects under difficult conditions is discussed in Ref. [12].

Proposed Work

The proposed system is used to detect and handle occlusion. Also it uses spatial method, the edge base restoration technique for the partially occluded portion and temporal method to fill the missing parts from the past history of a person if available for the full occlusion. The following Figure 1 shows the functional diagram for an efficient occlusion handling. Figure 1 illustrates the solutions for both partial and full occlusion handling. The input video is converted into frames and the edges are identified in each frame. Using histogram matching, check that all individual pixels are adjacent to 255. Find the artifacts after the edges are detected and calculate the area for the artifact region. Set the threshold values for the area to be calculated. If the threshold value is small, it is partial occlusion, and the edge based restoration technique is used to handle the partial occlusion. If the threshold value is large, it is full occlusion, and the frame matching technique is used to handle the full occlusion. For partial occlusion the edge features are extracted, the image structure is reconstructed and the edge based implemented to display the needed object. In the case of full occlusion, the reference frame is matched with previous based on texture or color to fill the missing parts and the result is displayed. Since the human is a bounded non-rigid object, the occlusion may occur at the boundary at any part and at any time. Occlusion may be partial or full.

Moving object segmentation

Motion segmentation in video sequences is known to be significant and difficult problem, which aims at detecting regions corresponding to moving objects such as vehicles and people in natural scenes. Background subtraction is a particularly popular method for motion segmentation, especially under those situations with a relatively static background. It attempts to detect moving regions in an image by differencing between the current image and a reference background

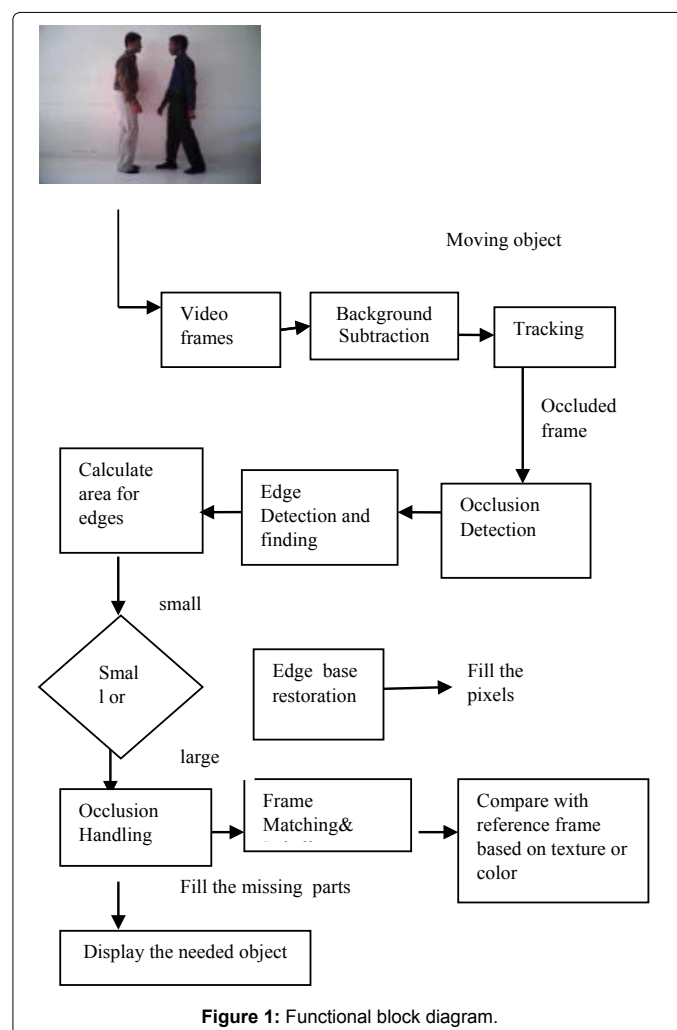


image in a pixel-by-pixel fashion. The simplest background model is the adaptive background subtraction.

Adaptive background subtraction model

This method uses a reference background to detect the foreground. Then subtract the intensity value of each pixel the current image from the corresponding value in the reference background image. This difference is then filtered with an adaptive threshold to provide insensitive solution for pixels. Let $I_n(X,Y)$ represents gray level intensity value at pixel positions x,y and at a time instance n of video image sequence I . Let $B_n(x,y)$ be the corresponding background intensity value for the pixel position x,y estimated over time from the video images B_0 through B_{n-1} . As a generic background subtraction scheme suggests, a pixel at position x,y in the current video image belong to the foreground if it satisfies:

$$\text{If } |I_n(x,y) - B_n(x,y)| > T_n(x,y) \text{ a}_1 \text{ else } a_0 \quad (1)$$

where $T_n(x,y)$ is an adaptive threshold value estimated using the image sequence I_0 through I_{n-1} . The above equation is used to generate the foreground pixel map which represents the foreground region as binary array where a_1 corresponds to foreground pixel and a_0 stands for background pixel. The Figure 2 shows an example of adaptive background subtraction model.

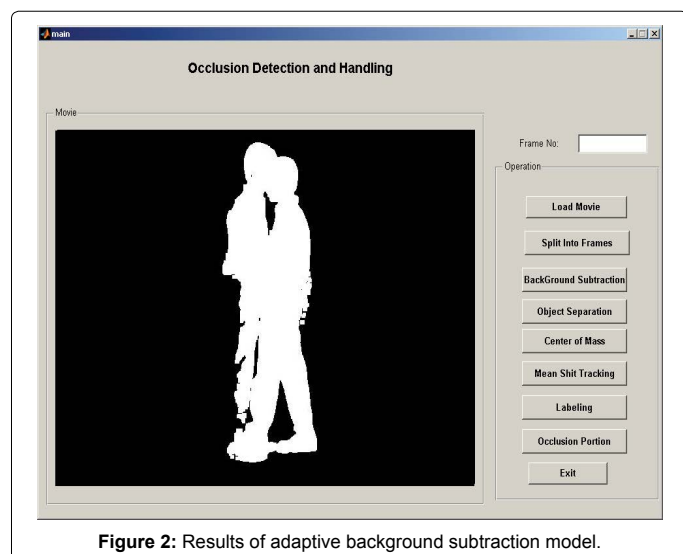


Figure 2: Results of adaptive background subtraction model.

Noise cancellation: While detecting the moving object using background model, it produces some noise. This noise affect the output of many calculation stages during the processing of price and causes inaccurate results. In order to get improved results, noise removal is crucial step. For this purpose, simple but effective algorithm is used in the proposed system. These are

- Erosion
- Dilation

In the proposed system, Erosion and Dilation are applied to the images which have large connected components, and thus noisy pixels are eliminated.

Tracking

For the improved tracking algorithm, occlusion layers are introduced to represent occlusion relation and the non-occlusion parts of the persons are obtained according to the occlusion relation and used for tracking. During the tracking process, the states of the persons are gradually adjusted by one to eliminate the occlusion effects. A Gaussian-mixture based adaptive background modeling algorithm is used to detect the foreground mask and label them as different connected patches.

For improved Mean Shift

- Rectangle position of image patch

$$B_j(t) = ((x_j, y_j), h_j, w_j) \quad (2)$$

- Occlusion layers

$$(\Phi(t) = \{q_j(t) = k\} n(t), j=1 \quad (3)$$

For the non-occlusion persons, each one can be labeled as one patch and the approach of histogram matching is applied to estimate their states. For the occlusion persons, more than one person could be labeled as one patch. An improved mean shift tracking algorithm which is special for occlusion target tracking is used. The iteration number of improved mean shift tracking algorithm is smaller than that of traditional tracking algorithm. Figure 3 shows the results of improved mean shift tracking algorithm.

Occlusion detection

Occlusion is detected when the objects are tracked. While tracking, the occluded object is separated by taking each pixel from the foreground image [White(1), Black(0)]. If the pixel value is zero, change the corresponding red, green and blue pixels in the current frame to zero. To find the centre of mass, find the extreme points of each object. Select any two diagonal points and calculate the centre of mass.

$$(c(2)+c(2)+c(4))/2, (c(1)+c(1)+c(3))/2 \quad (4)$$

where (c1,c2) - starting co-ordinates and (c3,c4) - co-ordinate value of column increment and row increment respectively. Figure 4 shows the results of object separation and Figure 5 shows the results of centre of mass calculation for the occluded object.

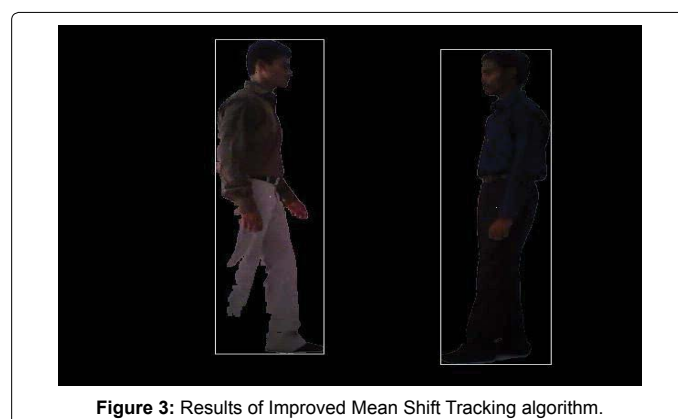


Figure 3: Results of Improved Mean Shift Tracking algorithm.

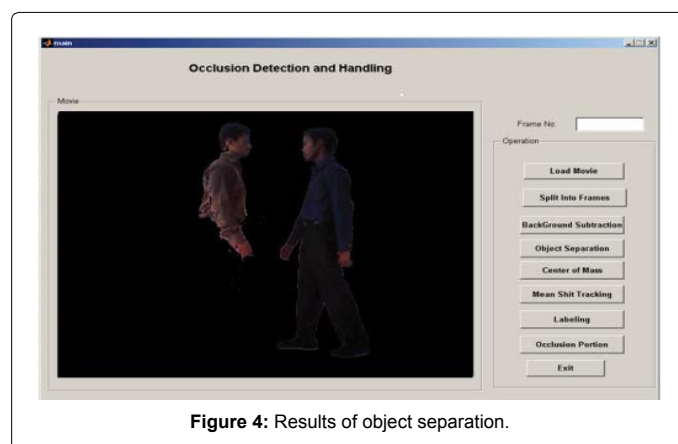


Figure 4: Results of object separation.

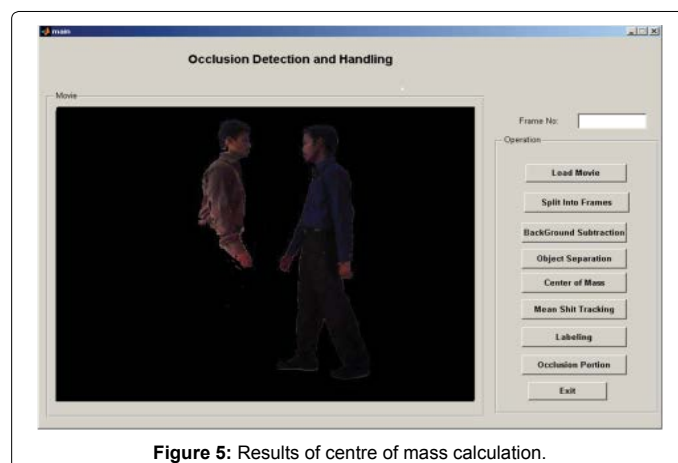


Figure 5: Results of centre of mass calculation.

Occlusion handling

Occlusion is efficiently handled for both the partial and full occlusion. Spatial approach is used to handle partial occlusion and the temporal for full occlusion. If a person is partially occluded the tracking algorithm will represent two bounding box for each of the two persons. In that case spatial method is applied to handle the occlusion. If the person is fully occluded or at the most occluded then temporal method is used to handle the occlusion.

Edge-based restoration for spatial approach

Edges generally separate areas with different content. Edges are more robust than isophote -based algorithm. Uses edge information that extracted from the surrounding area of the artifact. It is used both for the reconstruction of a skeleton image structure and guide for interpolation. The image structure is preserved as much as possible, by performing a non-linear interpolation based on edge information. Using histogram matching, check that all individual pixels are adjacent to 255. Find the artifacts after the edges are detected and calculate the area for the artifact region. Set the threshold values for the area to be calculated. If the threshold value is small, it is partial occlusion, and the edge based restoration technique is used to handle the partial occlusion. The spatial restoration algorithm consists of three main steps, 1) edge detection and edge feature extraction; 2) image structure reconstruction; and 3) edge-based inpainting.

Edge detection and edge feature extraction: In this step, edges are detected around the artifact, based on the contours of the segments that result from a watershed segmentation. The object edges are extracted in clockwise order, from a point of view lying inside the artifact. Only relevant edges are then kept for the next steps.

Image structure reconstruction: In this step, recover the structure of the image within the artifact area. Structure reconstruction step invent content in places where it was last, based on some assumptions about the usual image properties. The input to this step represents a list of edges coming into the artifact, in clockwise order. The output of this step is a list of edge couples arranged in groups of edges, and a list of spare edges. To build accurate pair wise connections between edges, use of local features, as well as global features and prediction of the final configuration and spare edge reconstruction.

$$C^{cdg}(Z) = \frac{4 \times c^{loc}(Z) + 4 \times c^{glb}(Z)}{5} \quad (5)$$

Edge-based inpainting: In this step, the artifact is restored by inpainting, taking into account the recovered image structure. Essentially, the inpainting procedure restores a pixel based on the surrounding recovered edges. The surrounding edges indicate which pixels on the artifact border are used for the interpolation. Then, based on the distance to these border pixels, the pixel inside the artifact is interpolated. The interpolation method tries to draw strips "parallel" to the nearby edges, resulting in smooth patches. The inpainting algorithm consider inpainting of a side strip with continuous contour bounded only by an edge couple, Spare Edge and inpainting of a middle strip with continuous contours. Figure 6 shows the results of occlusion handling by spatial method.

Frame matching

Each video frame is captured to maintain the history of a person. To keep track of a particular object appearance model defined as the collection of photometric feature vectors for pixels inside the target region. The image region of rigid moving objects can be obtained from

a fixed template region via a coordinate transformation; it is convenient to map the feature vector at a point in the target region at time to the template feature vector. The feature vectors are tracked independently by individual temporal filters. These vectors are matched to the current image in order to determine the measurement for the appearance filter at a time. The matching is performed by finding the image region that yields the maximal likelihood.

A template pixel is regarded as outlier if the measurement error exceeds the threshold. An occlusion is declared when the fraction of outliers exceeds a predefined percentage. During the occlusion, the template and parameters are not updated. By labeling, the occluded video frame is compared with the previous frames to find the occluded portion.

Feature correspondence

Texture and color: If the occluded part is matched with any of the same part in the previous frames, that particular part is extracted in the case of partial occlusion. In the case of full occlusion the occluded frame is matched with the reference frame and the human is extracted. Figure 7 shows the results of occlusion handling by temporal method.

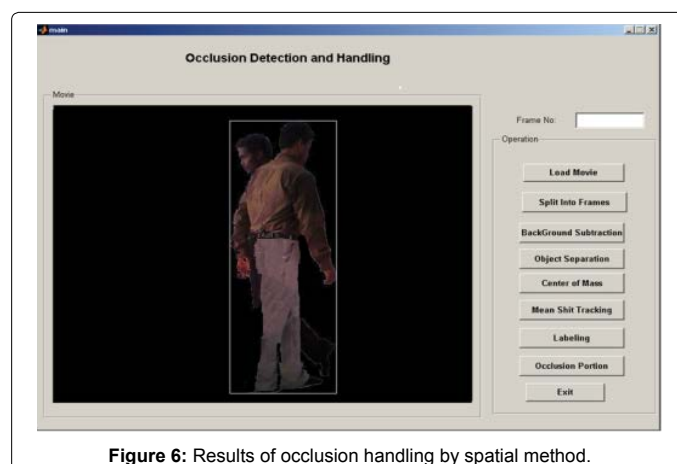


Figure 6: Results of occlusion handling by spatial method.

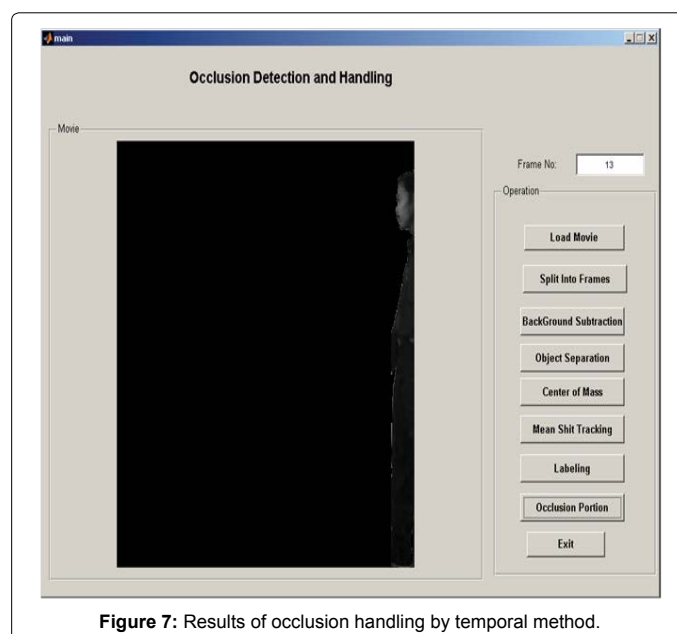


Figure 7: Results of occlusion handling by temporal method.

Conclusions and Future Work

The proposed work preserves both sharp and smooth edges. It makes use both local and global features of the edges in the image. Sophisticated approach can be developed - available temporal information, along with the spatial information. The proposed system handles the occluded part even when the occluded part and the background are in same color. The proposed system can be extended to track the objects correctly when the skin or dress color of the objects and the background are in same color even though the objects can be separated correctly. Also the complete occlusion of similar looking objects may cause ambiguities.

References

1. Jiang H, Zhang G, Wang H, Bao H (2015) Spatio-Temporal Video Segmentation of Static Scenes and Its Applications. *Multimedia*, IEEE Transactions 17: 3-15.
2. Martin J, Vazquez D, Lopez AM (2014) An Occlusion Handling via Random Subspace classifiers for Human Detection" *Cybernetics*. IEEE Transactions on Image Processing 44: 342-354.
3. Kratz L, Nishino K (2012) Tracking Pedestrians Using Local Spatio-Temporal Motion Patterns in Extremely Crowded Scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 34: 987-1002.
4. Kokaram A (2002) Parametric texture synthesis for filling holes in pictures. *Proceedings International Conference on Image Processing*, Rochester, NY, USA.
5. Rane SD, Sapiro G, Bertalmio M (2003) Structure and texture filling-in of missing image blocks in wireless transmission and compression applications. *IEEE Transactions on Image Processing* 12: 296-303.
6. Ballester C, Bertalmio M, Caselles V, Sapiro G, Verdera J (2001) Filling-in by joint interpolation of vector fields and gray levels. *IEEE Trans Image Processing* 10: 1200-1211.
7. Bornard R, Lecan E, Laborelli L, Chenot JH (2002) Missing data correction in still images and image sequences. *IEEE Trans Multimedia* 3: 326-338.
8. Li Z, Tang QL, Sang N (2008) Improved Mean Shift Algorithm for Occlusion Pedestrian Tracking. *Electronics Letters* 44: 622-623.
9. Pan J, Hu B, Zhang JQ (2008) Robust and Accurate Object Tracking Under Various Types of Occlusions. *IEEE Transactions on Circuits and Systems for Video Technology* 18: 223-236.
10. Guha P, Mukerjee A, Venkatesh KS (2005) Efficient Occlusion Handling for Multiple Agent Tracking by Reasoning with Surveillance Event Primitives. *ICCCN '05 Proceedings of the 14th International Conference on Computer Communications and Network*, Washington, DC, USA.
11. Pan J, Hu B (2007) Robust Occlusion Handling in Object Tracking. *IOP Conf. Ser.: Earth Environ*, USA.
12. Yang T, Li SZ, Pan Q, Jing L (2005) Real-time Multiple Objects Tracking with Occlusion Handling in Dynamic Scenes. *IEEE Conference* 1: 970-975.