

# Wood Detection and Tracking in Videos of Rivers

Imtiaz Ali<sup>1,2</sup>, Julien Mille<sup>1,3</sup>, and Laure Tougne<sup>1,2</sup>

<sup>1</sup> Université de Lyon, CNRS

<sup>2</sup> Université Lyon 2, LIRIS, UMR5205, F-69676, France

<sup>3</sup> Université Lyon 1, LIRIS, UMR5205, F-69622, France

**Abstract.** Rivers during floods bring a lot of fallen trees and debris. Video surveillance systems are installed on strategically important places on the rivers. To protect these places from destructions due to accumulation of wood, such systems must be able to automatically detect wood. Image segmentation is performed to separate wood and other moving elements from the rest of the water. Moving objects are detected with respect to brightness and temporal variation features. The floating wood is then tracked in the sequence of frames by temporal linking of the segments generated in the detection step. Our algorithm is tested on multiple videos of floods and the results are evaluated both qualitatively and quantitatively.

## 1 Introduction

Video monitoring systems installed on rivers record videos throughout the year. During floods, rivers carry many fallen trees, bushes, branches of fallen trees and other small pieces of wood. Automatic detection and counting of these fallen trees and other wooden pieces will help to protect infrastructures like bridges and dams from hazardous accumulation of trees. This automation will also decrease the manual efforts involved in supervised surveillance. In this paper, automatic wood detection in the river is performed by image segmentation and motion tracking. Dynamic nature of such application implies many constraints and limitations.

This paper is organized as follows. Section 2 presents a review of relevant works. Section 3 summarizes observations and assumptions made on available videos. The image segmentation method is presented in 4. Moving objects are related within successive frames thanks to a temporal linking method, which is described in section 5. This section also presents the method for counting wood pieces in the river. The experimental results, including comparison with ground truth data are presented in section 6. Section 7 concludes and presents perspectives to video analysis in outdoor scenarios.

## 2 Related Works

Wood detection and tracking in rivers is an example of moving object detection within moving background. There are two major theories for object recognition

in fixed camera videos [1]. In the first approach, the moving object is identified first and then the motion it performs in the image sequence is sought (*e.g.* background subtraction methods). In the second approach, the motion information is used directly to recognize moving objects (*e.g.* optical flow based techniques). For moving object detection, the adaptive background model was proposed for non-stationary backgrounds by [2]. It is constructed by adapting the changes during the training period. Gaussian Mixture Model (GMM) method is used by many researchers [3,4], where one or more Gaussian(s) are used to represent a pixel-wise background model. The Gaussian model parameters are recursively updated in order to follow the gradual background changes. The Weiner filter is used by [5] to learn and predict color changes in each background pixel. In [6] a spatio-temporal filtering method is proposed for compensation of the limitations of region-based image blocks, applied in an aquatic context. A filtering method based on spatial features with spectral features is also presented in [7].

Unlike previously discussed methods, optical flow-based methods proposed by [8] directly detect moving objects from their motion information. [9] used the estimation of the consistency of the optical flow over a short duration of time.

### 3 Limitation of Existing Methods in Our Case

Available videos of wood imply many constraints and difficulties which are summarized. Fig. 1 represents a few images extracted from the videos. The presence of bridge (top left corner of images), moving branches of tree in front of the camera (right middle portions of images) and the shadows of surrounding trees over the river are evident from these images. The detection of wood depends on the intensity difference between wood and water. But water waves in the presence of sunshine resemble wood pieces as shown in Fig. 2. The distinction between waves and floating wood must be made for correct wood detection and tracking. Waves and wood move at similar speed and hence cannot be distinguished with respect to motion dissimilarity criterion. The motion of wood is not purely translational. Finally, due to remote location of the monitoring scene and the limitations of transfer rate of data networks, the frame rate in the video is very low ( $\sim 4$  fps). Consequently, the object displacement is large between consecutive frames of videos as shown in Fig. 6.

Due to above properties of videos, existing methods may suffer from several limitations. The GMM method may lead to misclassification when the



**Fig. 1.** Few original images of, (a) a wood piece under the shadows of surrounding trees, (b) surrounding building cast shadows over water, (c) a fallen tree in the cloudy weather and (d) a wood piece having reflection of sun shine from the surface of water

background scene is complex [10]. For tracking objects in consecutive frames, the consistency of differential optical flow methods require small object displacements, which is not the case in current videos. Larger water waves have strong consistent movements in multiple frames. Moreover the object shape also play an important role in such methods, but in our case there is no specific shape or size of wood objects. Consequently, due to the complex nature of our application, we chose not to construct a background model. As a matter of fact the background is dynamic with water waves and wood in motion with the same speed.

## 4 Segmentation of Floating Wood

This paper is an improvement of our previous work [11]. We have developed the mathematical modelization of the problem. The intensity based segmentation method described in this new paper gives better results. Automatic detection of wood begins with the image segmentation step, which is a pixel-based probabilistic approach based on intensity and its temporal variation. We take as an input a sequence of  $T$  frames  $\{I(\cdot, t)\}_{1 \leq t \leq T}$ . Basically, we rely on two observations: wood is darker than water and undergoes permanent motion. The intensity probability map  $P_i$  contains the likeliness of pixels to be wood with respect to their brightness, whereas the temporal probability map  $P_t$  contains this information with respect to the brightness temporal variations at each pixel level.

### 4.1 Intensity Probability Map

The brightness of floating wood pieces is lower than water, even under the shadows of surrounding trees. Moreover, it does not change significantly in the presence of sunlight. Fig. 5 shows intensity histograms of wood pieces as an example. It seems relevant to approximate the intensity distribution of wood by Gaussian distribution with fixed mean and variance. The input grayscale value of each pixel  $\mathbf{x}$  at time  $t$  being denoted by  $I(\mathbf{x}, t)$ , the probability of the current pixel to belong to wood is

$$P_i(\mathbf{x}, t) = g_{\mu, \sigma^2}(\mathbf{x}, t) \quad \text{and} \quad g_{\mu, \sigma^2}(\mathbf{x}, t) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(I(\mathbf{x}, t) - \mu)^2}{2\sigma^2}\right)$$

where  $g$  is a Gaussian probability density function. To find  $\mu$  and  $\sigma^2$ , we led experiments on different wood pieces under various lighting conditions. Selection of mean and variance is discussed in section 4.4. Fig. 4(b) shows an example



**Fig. 2.** An image with water wave and wood piece are highlighted, (b) upper portion of image is zoomed to show shape and color intensity of wave, (c) part of image is zoomed to show shape and color intensity of wood piece

of  $P_i$  for a wood piece. In the presence of cast shadows of surrounding tree, the intensity probability map  $P_i$  has higher value in wood regions but also at undesirable shadowed regions.

### 4.2 Temporal Probability Map

Wood cannot be extracted relying solely on intensity considerations. Indeed, some objects like bridge pillars or cast shadows of surrounding trees have the same intensity as wood. To remove these static objects, we rely on pixel-wise temporal variations. The temporal probability is partially based on the normalized inter-frame difference  $\Delta_t I$ :

$$\Delta_t I(\mathbf{x}, t) = \frac{I(\mathbf{x}, t) - I(\mathbf{x}, t - 1)}{255} \tag{1}$$

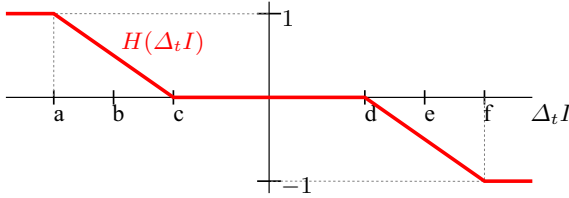
which takes its values within range  $[-1, 1]$ . Hard thresholding the absolute interframe difference  $|\Delta_t I|$  has been extensively tested for object detection. By nature, this technique only detects new object pixels and inevitably removes object areas that overlap in time. This is the case here with big wood pieces. Our temporal probability  $P_t$  is defined in order to avoid this drawback. We design it according to the observation that, when wood passes through a given pixel,  $\Delta_t I$  dips to a negative value and then to a positive value afterwards. Moreover,  $P_t$  should naturally remain constant if  $\Delta_t I = 0$ . This is achieved using a recursive definition in time:

$$P_t(\mathbf{x}, t) = P_t(\mathbf{x}, t - 1) + H(\Delta_t I(\mathbf{x}, t)) \tag{2}$$

where  $H \in [-1, 1]$  is an updating function, mapping the inter-frame difference to the amount of changes in the temporal probability. We express it in accordance with the considerations previously addressed. To handle noise and ignore insignificant intensity variations due to the non-uniformity of wood or water,  $H(\Delta_t I)$  should be null for relatively small values of  $|\Delta_t I|$ . It allows to handle slow illumination variations as well. Beyond certain threshold value,  $H$  should increase or decrease as  $\Delta_t I$  gets significantly negative or positive, respectively. Instead of using hard thresholding which would cause  $H$  to jump suddenly from 0 to 1 or  $-1$ , we use a soft approach less critical with respect to the choice of threshold parameters leading to the following piecewise linear definition:

$$H(\Delta_t I) = \begin{cases} 1 & \text{if } \Delta_t I \in [-1, -\tau - \frac{B}{2}] \\ \alpha \Delta_t I + \beta & \text{if } \Delta_t I \in [-\tau - \frac{B}{2}, -\tau + \frac{B}{2}] \\ 0 & \text{if } \Delta_t I \in [-\tau + \frac{B}{2}, \tau - \frac{B}{2}] \\ \alpha \Delta_t I - \beta & \text{if } \Delta_t I \in [\tau - \frac{B}{2}, \tau + \frac{B}{2}] \\ -1 & \text{if } \Delta_t I \in [\tau + \frac{B}{2}, 1] \end{cases} \tag{3}$$

where  $\alpha = \frac{-1}{B}$  and  $\beta = \frac{1}{2} - \frac{\tau}{B}$ . Fig. 3 plots  $H$  versus  $\Delta_t I$ . Variation of  $H$  in turns requires a threshold  $\tau$  and transition length  $B$ . The probability of a wood pixel must have higher value than surrounding. It should be noted that  $P_t(\mathbf{x}, t)$  in Eq. (2) is truncated between 0 and 1 afterwards to remain a probability. In



**Fig. 3.** Representation of updating function  $H(\Delta_t I)$

the first frame, we set  $P_t(\mathbf{x}, 1)$  to 0 everywhere, as it is very unlikely that wood pieces appear at initial time. Temporal probability  $P_t$  is non-null only if temporal brightness variation is negative enough, *i.e.* if a pixel gets significantly darker or has the same brightness as it had in the previous frame. It helps in removing stationary objects in the scene (*e.g.* pillars of bridge). Fig. 4(c) highlights the fact that  $P_t$  has higher values for a wood piece than water and static areas.

### 4.3 Combination of Intensity and Temporal Probability Maps

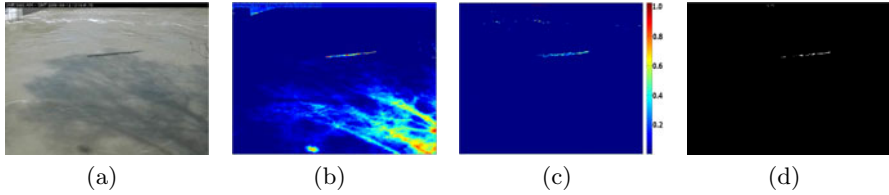
Since we expect wood to be simultaneously dark and under motion, wood pixels should have both high intensity and temporal probabilities, hence it is relevant to multiply the two probability maps. The product yields the joint probability  $P_{global}$ , representing the likelihood of a given pixel to be wood with respect to its intensity and corresponding variation:  $P_{global}(\mathbf{x}, t) = P_i(\mathbf{x}, t)P_t(\mathbf{x}, t)$ . The foreground image is obtained by simple thresholding of the joint probability map:

$$FG(\mathbf{x}, t) = \begin{cases} 1 & \text{if } P_{global}(\mathbf{x}, t) \geq G_{Th} \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

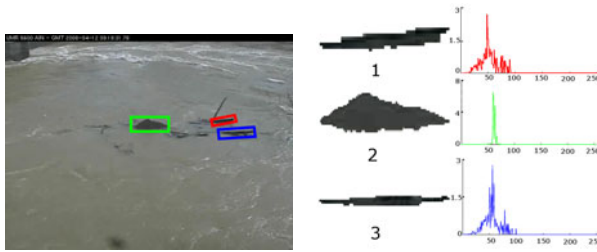
The joint probability  $P_{global}$  is high for wood pixels but also unfortunately for pixels located on dark waves. Hence, global threshold  $G_{Th}$  should be chosen in order to limit the number of false detections without removing significant parts of real wood pieces (choice of  $G_{Th}$  is discussed in section 4.4). An example of final foreground image is shown in Fig. 4(d), which clearly indicates that the algorithm can detect moving wood pieces under difficult weather conditions. Also, we have evaluated the segmentation results with ground truth images in section 6.

### 4.4 Selection of Parameters

In previous sections, we introduced some parameters which need to be investigated thoroughly. For this purpose, we extracted small and large wood pieces under different weather conditions. An example of floating wood along with three portions highlighted in different colors is shown in Fig. 5. We fix  $\mu = 55$  and  $\sigma^2 = 225$ . Computation of the temporal probability involves threshold  $\tau$  and its transition length  $B$ , whereas extraction of the final foreground image requires global threshold  $G_{Th}$ . Parameter tuning was performed through a brute-force



**Fig. 4.** (a) An example of wood piece under sunlight with corresponding, (b) intensity probability map  $P_i$ , (c) temporal probability map  $P_t$  and (d) resulting  $FG$  after thresholding



**Fig. 5.** Floating wood piece; (1) Zoomed portion of wood pixels and corresponding intensity histogram highlighted by red, (2) green and (3) blue rectangle

approach, by maximizing the overlap between the foreground image generated with current parameter values on one hand and ground truth segmentations on the other hand, on a training dataset. The overlap was measured using the Dice coefficient, which is a commonly used measure of segmentation quality (see for example [12]). It is expressed as  $S = \frac{2|X \cap Y|}{|X| + |Y|}$  where  $X$  is the result of image segmentation and  $Y$  is the corresponding ground truth image.  $S$  is equal to 1 when the segmented region and the ground truth region perfectly overlap, and 0 when they are disjoint. Firstly, for each parameter, the range of values giving satisfactory results was coarsely located by successive attempts. We determined that the triplet  $(\tau, B, G_{Th})$  leading to the best segmentation was located in range  $[0.1, 0.4] \times [0.1, 0.4] \times [0.01, 0.2]$ . Then, all parameter values within these ranges were tested, with respective steps 0.05, 0.05 and 0.02.

The optimal values for these parameters for which  $S$  values are maximum for both small and big wood pieces are  $\tau = 0.3, B = 0.3$  and  $G_{Th} = 0.05$ . For big wood pieces,  $S$  average value obtained is 0.8, with a maximum value equals to 0.9 and a minimum value equals to 0.71. For small wood pieces, average  $S$  is 0.71 with a maximum value equals to 0.8 and a minimum value equals to 0.6.

## 5 Tracking of Floating Wood

During flood water waves and turbulences are prominent. As shown in Fig. 2 waves and wood resemble and therefore, the distinction should be made between

them. We propose a temporal linking method based on the segmentation method for this purpose.

### 5.1 Extraction of Representative Points

The size, shape and orientation of the floating wood do not remain the same in consecutive frames. Moreover, submergence causes partial occlusions resulting in variable number of connected components, which is shown in Fig. 6. In order to group several connected components which may correspond to the same object, we first rely on centroids. The centroid  $c_{R_i}$  of a given component  $R_i$  is taken as the representative of  $R_i$ . To evaluate the closeness between two connected components  $R_1$  and  $R_2$ , we choose to consider the euclidean distance between  $c_{R_1}$  and  $c_{R_2}$ . This distance allows us to label the connected components of same object even if their size vary from one frame to another due to occlusion. We perform hierarchical grouping of connected components as long as the distance between their centroids is below a threshold  $s$ . At each step, the two closest connected components  $R_a$  and  $R_b$  are merged in a new region whose representative center is assigned to the average  $(c_{R_a} + c_{R_b})/2$ , until  $\|c_{R_a} - c_{R_b}\| < s$ . Let  $R = \{R_i\}_{i=1\dots n}$  be a set of gathered connected components. Its representative center  $c_R$  is the average of centroids  $c_{R_i}$ , and the following relation is verified:

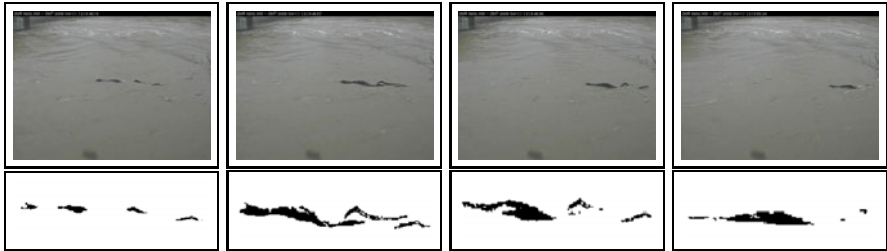
$$R = \{R_1, \dots, R_i, \dots, R_n\} \Rightarrow c_R = \frac{1}{n} \sum_{i=1}^n c_{R_i} \quad \text{and} \quad \|c_{R_i} - c_R\| < s \quad \forall i \in 1\dots n$$

This method is robust to partial occlusion of wood in water. Hence, every object in the frame is localized by a representative point, which is linked to its corresponding point in the next frame.

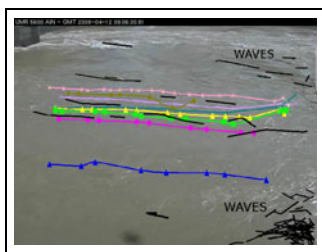
### 5.2 Temporal Linking of Floating Wood

Let  $c_R(t)$  and  $c_R(t + 1)$  be representative points of object  $R$  matched in two consecutive frames. This is actually verified if their distance is below  $\delta$ :

$$\|c_R(t) - c_R(t + 1)\| \leq \delta \tag{5}$$



**Fig. 6.** Four consecutive frames of a moving wood object, zoomed resulting segmented object regions show the appearance of floating wood in the images



**Fig. 7.** A summary of small video portion represents wood (in different colors) and water waves (black color) trajectories

where  $\delta = 100$  pixels, which is the maximal displacement of wood pieces we learned after experimentally testing on different videos. Incidentally, it allows us to give a lower bound of threshold  $s$ . If  $s$  is lower than  $\delta$ , objects may be mismatched in consecutive frames. A component of an object may be mistakenly matched with another component of the same object, which does not happen if  $s > \delta$ . Similarly two objects moving simultaneously can be counted separately.

The positions of the representative point of an object in consecutive frames can be linked by line segments, which yields a trace in the *summary image*. Such an image is a graphical representation of trajectories during a given duration, an example is shown in Fig. 7. We can notice from the summary image that wood objects make longer traces than waves. The summary image also exhibits that water waves disappear after some frames. This property is used in the following to distinguish wood pieces from waves.

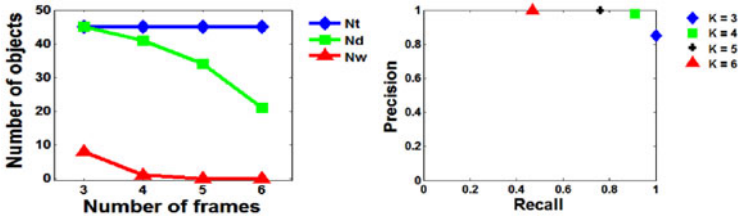
### 5.3 Counting Wood Pieces

For each object  $R$ , we determine the number of consecutive frames in which it appears. We obtain a sequence of  $n$  representative points  $\{c_R(t), c_R(t+1), \dots, c_R(t+n-1)\}$  in which each couple  $(c_R(t+i), c_R(t+i+1))$  verifies Eq. 5. Wood pieces and water waves are separated from one another according to their persistence in the consecutive frames. Hence, the chosen criterion to consider  $R$  as a wood piece is  $n \geq K$ . If a wood piece is not totally submerged, its representative point at different times should all be linked two by two. Unlike wood pieces, waves generally disappear after three or four frames. Hence, floating wood is counted on this basis. The relevancy of this limit is evaluated in section 6.

## 6 Experimental Results

The studied videos were generated by a monitoring system set up on the river Ain in France. Notice that even if all our videos come from a unique camera, weather conditions vary a lot. No assumption is made about the wood position in the water or according to static parts such as bridge. Consequently our method could be used with another camera in a likewise scene. We tested our algorithm of wood counting on five videos of 1500 frames each, for which we had ground truth





**Fig. 8.** (on the left) Quantitative comparison of selection the number of consecutive frames for wood attribution for a video (on the right) precision and recall for video evaluated for  $K = 3, 4, 5$  and 6 consecutive frames

**Table 1.** Quantitative evaluation of wood counting and segmentation in videos

Video	Wood counting				Wood segmentation
	$N_t$	$N_d(\%)$	$N_{pd}(\%)$	$N_w(\%)$	Overlap ratio(%)
1	41	91	9	1	79
2	37	90	10	0	76
3	47	90	10	0	81
4	80	92	8	3	83
5	85	93	7	4	80

data for validation. Frames have size  $640 \times 480$  and are extracted from MPEG4-compressed streams tested on an Intel Core2 Duo 2.66GHz with 4GB RAM running C code. Average computation time per frame is 210 ms and it can be used smoothly in the on-line scenario. In section 5.3, we introduced the minimal number of consecutive frames  $K$  during which wood pieces should appear and to be counted as wood. This number is evaluated in Fig. 8. Qualitative evaluation is given by precision  $Pr$  and recall  $Re$ , defined as:

$$Pr = \frac{N_d}{N_d + N_w} \quad ; \quad Re = \frac{N_d}{N_t}$$

where  $N_d$  is the number of detected wood pieces by algorithm,  $N_w$  is the number of waves detected as wood and  $N_t$  is the total number of wood objects *i.e.* ( $N_t = N_d + N_{pd}$ ) where  $N_{pd}$  number of non detected wood pieces. We can see that the best trade-off between false positive and negative detections is obtained with  $K = 4$ . Successful wood counting is validated manually by visual inspection frame per frame. Results are summarized in Table 1. Moreover, we randomly selected a wood piece per video and compute the overlap ratio of segmented wood object with ground truth.

After visual inspection, it turns out that undetected wood pieces correspond to very small parts, which are not critical with respect to the application. These small pieces are often totally submerged in some frames.

## 7 Concluding Remarks

In this paper we presented an automatic method for detecting and counting the floating wood in rivers. Intensity and temporal probability maps are computed

for every incoming frame. These two probability maps are combined and resulting image is segmented by selecting a threshold. The resulting segmented image contains wood pieces along with some water waves. Water waves are separated from wood pieces by temporal linking method. Due to outdoor environment there are many constraints in our case. The experimental results are evaluated on every step. This algorithm could be extended to any object detection within multiple motions in background. Future work will be dedicated to the incorporation of prior object motion knowledge in both segmentation and tracking processes to reduce again omissions and false detections.

## References

1. Shah, M.: Motion-based recognition: A survey. *Image and Vision Computing* 13, 129–155 (1995)
2. Li, L., Huang, W.M., Gu, I.Y.H., Tian, Q.: Statistical modeling of complex background for foreground object detection. *IEEE Trans. on Image Processing* 13(11), 1459–1472 (2004)
3. Stauffer, C., Grimson, W.: Learning patterns of activity using real-time tracking. *IEEE Trans. Pattern Anal Machine Intell.* 22(8), 747–757 (2000)
4. Wren, C., Azarbayejani, A., Darrell, T., Pentland, A.: Pffinder: realtime tracking of the human body. *IEEE Trans. on Pattern Anal. Machine Intell.* 19(7), 780–785 (1997)
5. Toyama, K., Krumm, J., Brumitt, B., Meyers, B.: Wallflower: principles and practices of background maintenance. In: *IEEE Int. Conf. Computer Vision (ICCV)*, pp. 255–261 (1999)
6. Eng, H.L., Wang, J., Wah, A.H.K.S., Yau, W.Y.: Robust human detection within a highly dynamic aquatic environment in real time. *IEEE Trans. on Image Processing* 15(6), 1583–1600 (2006)
7. Mittal, A., Paragios, N.: Motion-based background subtraction using adaptive kernel density estimation. In: *IEEE Conf. Comp. Vision and Pattern Recog. (CVPR)*, pp. 302–309 (2004)
8. Horn, B.K.P., Schunck, B.G.: Determining optical flow. *Artificial Intelligence* 17, 185–203 (1981)
9. Wixson, L.: Detecting salient motion by accumulating directionally-consistent flow. *IEEE Trans. Pattern Anal Machine Intell.* 22(8), 774–780 (2000)
10. Gao, X., Boulton, T., Coetzee, F., Ramesh, V.: Error analysis of background adoption. In: *IEEE Conf. Comp. Vision and Pattern Recog. (CVPR)*, pp. 503–510 (June 2000)
11. Ali, I., Tougne, L.: Unsupervised video analysis for counting of wood in river during floods. In: *Bebis, G., Boyle, R., Parvin, B., Koracin, D., Kuno, Y., Wang, J., Pajarola, R., Lindstrom, P., Hinkenjann, A., Encarnaç o, M.L., Silva, C.T., Coming, D. (eds.) ISVC 2009. LNCS, vol. 5876, pp. 578–587. Springer, Heidelberg (2009)*
12. Cárdenes, R., Bach, M., Chi, Y., Marras, I., de Luis, R., Anderson, M., Cashman, P., Bultelle, M.: Multimodal evaluation for medical image segmentation. In: *Kropatsch, W.G., Kampel, M., Hanbury, A. (eds.) CAIP 2007. LNCS, vol. 4673, pp. 229–236. Springer, Heidelberg (2007)*