

Stereo System for Remote Monitoring of River Flows

Konstantinos Bacharidis¹  · Konstantia Moirogiorgou¹ ·
Georgia Koukiou² · George Giakos³ · Michalis Zervakis¹

Received: 30 January 2017 / Revised: 15 June 2017 / Accepted: 23 August 2017 /

Published online: 22 September 2017

© Springer Science+Business Media, LLC 2017

Abstract In this article we present a video-based method for river flow monitoring. The proposed method aims at deriving efficient approximations of the river velocity using natural formations on the river surface. In order to overcome peculiarities of the flow, we propose to uniformly exploit all such structures that appear locally with short temporal duration. Towards this direction we explore the expanded capabilities of a stereoscopic camera layout with the dual observation fields and the potential of reverting projective deformations. By mapping to world coordinates, all spatial locations in the video reflect velocity as a uniform field, except for local flow variations. The velocity estimation is performed by computing the optical flow using a series of video frames, combining the information of the views of both cameras. The novelty of the proposed river flow estimation scheme lies on the fact that the accuracy of motion estimation is increased due to the use of the complementary views, which also enables the transition from a 2-Dimensional image-based velocity estimate to 3-Dimensional estimates. The estimated optical velocity is back-projected to the real world coordinates using the parameters extracted using the stereoscopic layout. The results on simulated and real conditions demonstrate that the proposed method is efficient in the estimation of the surface velocity

✉ Konstantinos Bacharidis
kbacharidis@isc.tuc.gr

Konstantia Moirogiorgou
dina@display.tuc.gr

Georgia Koukiou
gkoukiou@upatras.gr

George Giakos
george.giakos@manhattan.edu

Michalis Zervakis
michalis@display.gr

¹ Electrical and Computer Engineering School, Technical University of Crete, Chania, Greece

² Electronics Lab, Physics Department, University of Patras, Patras, Greece

³ Electrical and Computer Engineering School, Manhattan College, Bronx, NY, USA

and robust against locally disappearing formations, since it can compensate for a loss with other formations active in the field of view.

Keywords River flow monitoring · Remote sensing · Optical flow · Stereo Vision

1 Introduction

NOWDAYS the deforestation and the sudden filling of riverbeds and torrents at ephemeral rivers as an effect of the climate change phenomenon has led to the increase of flash flood events and to variations on the depth and the water quality for the river flows [17]. The impact of these events to both the ecosystem and human life poses a scientific interest for the hydrological community [19]. Monitoring river flow, especially during flash flood events provides a better understanding of the impact of the human intervention in the natural ecosystem that leads to an incremental occurrence of these phenomena. River flow monitoring with conventional equipment, especially under extreme weather conditions, is a difficult task and can lead even to life risking conditions for the experts that handle the equipment. Moreover, such events usually have short duration and abrupt occurrence. This means that fast responsiveness or continuous human presence is needed to monitor the events to their full extent.

Remote sensing has been proposed as a solution to these problems e.g. [8, 15, 22]. Such systems are low-cost solutions and can provide continuous site monitoring, without human intervention. Moreover, their capabilities can be extended to support real-time site monitoring, combining the data of multiple distributed sensors. This feature combined with service oriented architectures can further expand the capabilities of the system, allowing for efficient planning and decision making in extreme environmental crisis management scenarios [31]. For river flow surveillance the most recent approaches utilize radar, satellite or image sensor-based systems for river surface velocity measurement e.g. [1, 3, 8, 15, 20, 22]. For the case of image-based remote sensing systems, a single camera or a pair of cameras is used to monitor the site. The video data collected, are used to derive 2D image motion estimates of the flow field. Finally, the estimation of the water surface velocity occurs by relating the motion vectors with known ground points in the site. In other words, the main parts of an image based monitoring system are (a) the installation of the appropriate hardware, (b) the mapping of the image coordinate system to the real-world one (c) the selection of an appropriate motion estimation method for the computation of the 2D velocity field and (d) the correct association of the 2D velocity to the 3D velocity estimate.

The derivation of motion estimates in a physical environment through an optical system is quite challenging due to the non-rigid nature of the water surface. The non-rigid motion of the water is multi - directional and is affected by external forces [6], such as the wind, gravity, etc. Image based motion estimation techniques usually make use of motion models in order to derive the motion field. On the other hand, non-rigid motion varies in a dynamic way, which means that it cannot be described using a specific motion model. For this reason, traditional optical flow methods, that are based on affine neighborhood matching or that make use of a' priori motion model assumptions, cannot provide accurate calculations when dealing with fluid flow images.

In most cases of river monitoring using camera sensors, the estimation of the water motion usually uses the existence of specific formulations that can be tracked in time, such as waves, foam or discrete objects, like leaves etc. Based on this concept a number of methods have been proposed, which rely either on the rigidity of discrete objects flowing the river surface in order

to apply classic motion transformation models e.g. [12, 18, 23, 32], or by modeling the image intensity variations based on Bayesian probabilistic schemes e.g. [1, 4, 14]. Finally, fluid motion model schemes that make use of fluid properties to constrain the modeled intensity variation can be applied [24, 26, 27].

However, the formations being tracked can be observed anywhere in the monitored surface. The distance from the camera due to the perspective effect determines the relation between the estimated and real motion at each point. One of the aims of this study is to derive a uniformly defined motion everywhere in the picture by reverting the perspective effect. This is achieved through the use of a stereo camera system which enables the estimation of transform parameters and the derivation of depth information without the use of control points.

The 3D motion lacks of perspective effects and provides information uniformly for all points in the scene. More specifically, the motion of a river surface can be thought in a 2D surface plane. The proposed approach aims to transform the perspective projection into projective, parallel to the surface plane. Given the physical to image coordinate system relation we are able to go, in 2D, from perspective to projective motion with all motion vectors to share the same importance and relative extent. In this way, we can utilize the full potential of the fluid flow field instead of tracking regions of particles, increasing the amount of information for the flow field and removing particle-tracking related problems that affect the tracking accuracy such as viewpoint and scene illumination changes, object deformation due to viewpoint and motion change, etc.

In order to perform the proposed methodology to the entire fluid surface, we need to use an optical flow estimation method that can retain both the local motion details at all different parts (regions) of the flow and the global ones through the definition and use of a relation scheme between these local regions. The calculation of the local motion is important since the movement of a dynamic body like that of a fluid mass is based on the interaction of both external and internal forces and is characterized by a unique model of the overall motion field e.g. [5, 33]. Local interactions between the molecules of the liquid are linked to arrive at a unified field of motion which characterizes the liquid as a whole. The global motion field can involve a variety of motion trends that change from the center of the river to its banks, depending on the morphology of the river basin. Thus, we would be interested in recovering the main trend and possibly some prevailing local trends. On the other hand, the accurate estimation of the global motion is related with the process of the 2D velocity field estimation. In our study, we have defined a new stereo probabilistic optical flow estimation framework that utilizes the additional information provided by the stereo layout in a Bayesian inference scheme to accurately estimate the motion field of the river's surface. This probabilistic formulation initially, allows the estimation of local displacement probabilities for every formulation present in the river surface, grouped into pixel neighborhoods and finally, relates each local motion distribution estimate with the global one through a global estimation framework. This framework leads to a global motion field estimate, thus approximating non-rigidity in motion e.g. [1, 4, 14]. This local probability derivation can be performed more accurately by combining the pixel relation between the two image planes of a stereo system.

Finally, an additional use of the proposed formulation is the tracking of compact objects such as leaves or artificial particles on the river surface. The initial position of the object in the stereo camera system along with the estimation of 3D motion vectors, based on the 2D motion field estimate, enables the estimation of the 3D position of the object after time t . The projection of this position on the two 2D camera planes gives us the estimated position of

the object in cameras and limits considerably the search area for locating the object at time t . The fast tracking of objects is used for the validation for the accuracy of the 3D motion estimation based on the stereo system.

2 Related work

Despite the existence of several optical-based motion estimation methods for fluid flows, only a few have been deployed to real-world applications, such as river flow monitoring systems. The existing optical river flow estimation systems make use of particle based methods for the calculation of the surface river velocity vector field. The idea is that physical formations, such as leaves or foam [7, 11, 21], or artificial tracers can act as indicators of the river flow motion. These tracers are tracked through the frame series for extracting the associated optical displacements. There exist two major variations of this methodology, concerning whether we track one specific particle (PTV method) or a group of particles (PIV method). From the two strategies, PIV is widely used in river flow monitoring systems, due to the difficulties in isolating a single particle in the natural environment, where the illumination variations have a more severe effect when dealing with one particle/tracer.

The estimation of the optical displacement of the particle, in PIV methods, is usually performed using correlation techniques, examining the similarity of a region in the reference frame to candidate regions in the next frame. The similarity is assessed using correlation metrics [3, 7, 15] based on constant or shifting interrogation windows. However, there exist PIV approaches that combine particle tracking with classic optical flow methods [9, 10, 30]. The key concept of these methods is the existence of particles in the flow, which is also their disadvantage, since we cannot provide a constant seeding of artificial particles in the flow, especially during flood events. Even for the case of physical tracers such as waves or foam, there are cases of flow where neither of these formations can be present so the estimation process can't be performed. There are approaches in order to overcome the particle presence necessity, focused on the river monitoring case. Bacharidis et al. [1] made an early attempt to use a probabilistic approach, first introduced by Chang et al. [4].

Given the motion field of the fluid flow, the next step is to actually convert the 2D image velocity into the 3D real world velocity estimate. This part is the tricky one, since we need to convert the 2D world of the image plane into the 3D world coordinate system i.e. we need to introduce the depth perception that relates the image to its physical co-ordinates. Existing image monitoring systems use an eight-parameter projective transformation, presented first by Fujita et al. [11] with a number of control points (minimum four) with known coordinates to perform the transformation [7, 21]. However, this method assumes horizontal water surface and requires a careful and detailed selection of the control points at the site in order to perform an accurate estimation. In our approach, we propose the use of stereo layout, instead of a single camera that is usually used in the state of the art systems. This change eliminates the need for control point selection and additionally enables a more accurate modeling of the distortion effects during image acquisition, thus, reducing the error in the transformation process.

3 River flow estimation framework

The majority of the current image-based river flow estimation systems consist of three methodological modules, (a) the methodology relating the coordinate systems and scene

projections, (b) the methodology for image-based motion estimation and (c) the relation between the image-based estimated velocity with the real-world velocity.

Regarding the first module, the most common approach is to associate 3D world points with known coordinates with their corresponding 2D image points in order to estimate the projection parameters between the two coordinate systems. These projection parameters enable the relation between the image-based and the real-world velocity estimates. The second module, involves the estimation of the river's velocity field in the 2D image domain using motion and optical flow estimation approaches with the simplest and most popular, being the block matching approaches. Similarity metrics are employed for this purpose such as, Mean Squared Error (MSE) and Sum of Squared Distances (SSD). However, such metrics are based on motion rigidity, and thus require objects to be present on the river flow to estimate the fluid's velocity. Our earliest contribution on this topic, presented in [1], is the application of a Probabilistic Optical Flow Estimation Scheme which applies Bayesian inference to estimate the optical flow field of the non-rigid motion of the fluid. In fact, a major contribution of the proposed method is an improvement in both the computational cost as well as in the relation between the 2D image based motion field and the corresponding 3D world fluid motion.

The probabilistic method uses a conditional Bayesian model to estimate the unknown 2D velocity field $u = (v, w)$. The Bayesian formulation is based on the assumption that a pixel's flow vector is actually a random variable described by a probability distribution function. So, the unknown velocity can be formed as a posterior distribution that is estimated based on the Maximum a posteriori rule which is defined from prior likelihood motion model assumptions:

$$\hat{u} = \arg \max_u p(\phi|u) \cdot p(u) \quad (1)$$

where $p(\phi|u)$ is the conditional probability describing the observed data given the actual realization of the underlying global flow field, $p(u)$ is a probability describing a prior knowledge for the motion, usually a Gibbs distribution and \hat{u} the estimate of the optical flow field.

The crucial step in this formulation is the selection of the appropriate function ϕ for the observed data representation. In our approach ϕ , as Chang et al. [4] suggested, follows a stochastic formulation, assigning a probability of selection for each destination position of the candidate region in the next frame:

$$\phi \approx \frac{A_i}{\sum_{j=1}^{D_s} A_j} \quad (2)$$

where D_s is the size of the candidate neighborhood in which our pixel can be positioned in the next frame.

The coefficient A_i denotes this transition probability and is estimated by means of the relation of the image intensities of both frames, reference and next, based on a Spatio-Temporal Autoregressive model:

$$I(x, y, t) = \sum_{i=1}^D A_i I(x + \Delta x_i, y + \Delta y_i, t + \Delta t) \quad (3)$$

where $I(x, y, t)$ is the intensity of the pixel (x, y) at time t in the reference frame and the intensities of the pixels $(x + \Delta x_i, y + \Delta y_i)$ at the destination neighborhood at time $t + \Delta t$.

The estimation of the transition probabilities A is performed through a least squares scheme that utilizes the pixel intensity information of the neighborhood of the pixel (x, y) in the

reference frame and the pixel intensity information of the candidate neighborhood in the current frame. We essentially end up with a minimization of the cost function:

$$J(A) = \sum_{s=1}^N (I(x_s, y_s, t) - k_s^T A)^2 \quad (4)$$

with k_s being a vector of size D containing the intensities of pixels belonging to the candidate neighborhood D_s for the pixel (x_s, y_s) in the next frame, N is the number of elements in the spatial neighbourhood N_s of pixel (x_s, y_s) .

If we express Eq. (4) in a matrix form we end up with a least squares estimation problem of the form $Ax = b$, with A being a matrix containing the transition coefficients, x being a matrix containing the intensities of pixels belonging to the spatial neighborhood N_s centered at the pixel (x_s, y_s) and finally, b being a matrix containing the intensities of the pixels belonging to the candidate neighborhood D_s in the frame $t + 1$ where the pixel (x_s, y_s) is expected to be displaced at. The least squares problem is estimated, in our case, with respect to A instead of x since the matrix A contains the coefficients to be estimated. The solution of such least squares scheme is given by:

$$A = (K^T K)^{-1} K^T M \quad (5)$$

with K being the matrix containing all the vectors k_s consisting of pixel intensities for all the possible transitions for each pixel contained in the spatial neighborhood N_s defined by a central pixel (x_s, y_s) , i.e. each destination neighborhood D_s defined by the pixel (x_s, y_s) belonging to N_s and M being the matrix containing the intensities of the pixels belonging to the spatial neighborhood N_s . In order to cope with ill-conditioned cases, i.e. when K is not always full ranked, leading to Eq. (5) being violated, we apply the Singular Value Decomposition (SVD) method to solve the least squares problem and estimate the A matrix.

Finally, the third module is the one relating the 2D motion field with the real world to produce the 3D motion field. This process is a simple inverse mapping between two coordinate systems using the parameters estimated in the first module. The 3D velocity is computed by taking the difference between the 3D point coordinates of the initial and final position of the tracked object that have been defined based on the 2D image point positions estimated using the 2D motion field. The second contribution of the proposed methodology is on the mapping process between the 2D and 3D motion estimates.

4 Proposed method

Our approach presents a novel formulation of a video-based system for river flow estimation utilizing any available pattern formation on the surface of the water basin. Our method utilizes a stereo camera layout to derive the necessary relation between the physical and image coordinate systems. The estimation of the optical velocity of the fluid is performed using a stereo probabilistic framework for the computation of the optical flow field. To increase the accuracy of the estimation as well as to remove erroneous or unwanted motion vectors, image segmentation and machine learning classification methods are incorporated. Our approach is composed of the following steps:

- **Stereo Layout** - Use the stereo layout to derive the relation between the 3D physical and 2D image coordinate systems and formulate a region of examination with known real

world dimensions. This process determines the projection parameters for the transformation of 3D world to 2D camera coordinates.

- **Optical Flow Estimation** - Combine a probabilistic optical flow estimation methodology with the additional information provided by the stereo (layout) to estimate the optical flow field of the fluid employing the entire image domain of the stereo image pair.
- **Coordinate System Normalization for Velocity Estimation** - Associate the estimated motion field with the corresponding 3D physical velocities based on the 3D physical coordinate change of the 2D image points as defined by the 2D motion vector. In this way, the perspective distortion introduced by the transformation from the 3D world to 2D image domain is removed. This step allows the uniform consideration of spatial location for optical flow estimation, irrespective of the distance from the camera, through the inversion of perspective projection effects.
- **Velocity Estimate Validation** - Consider the average 3D velocity estimate as constant over the monitored area and validate the estimation by examining whether an existing particle in the flow, assuming to have the same velocity, travels along the examination area at the expected number of frames. By validating our estimate, we ensure that the estimated velocity accurately expresses the fluid's motion.

4.1 Stereo layout in world coordinates

As mentioned before, existing monitoring systems make use of a single camera set up. The relation between the physical and image coordinate systems is established by solving an eight-parameter transformation system, which, given the appropriate amount of data (at least 8 control points), leads to a system of linear equations. However, such formulations require the prior appropriate selection of known control points which in accordance with the horizontal viewing position assumption leads to harsher geometrical reconstructions of the scene (affine reconstruction). To overcome these disadvantages and reach to more detailed scene reconstructions, we propose the use of a stereo layout allowing the computation of both the stereo rigs' characteristics and the distortion values. This method does not require the use of control points to relate the physical and image coordinate systems. Moreover, it provides additional scene information, such as depth perception, that can be utilized in the processing part to increase the accuracy of estimation e.g. stereo motion estimation for the flow. The continuous debate when adopting a stereo rig formation is whether a convergent or a non-convergent layout is more suitable for the specific application. Each of these has its advantages and disadvantages. A parallel layout requires simpler transformations in order to move from the physical to image plane coordinate systems and provides more valid points for 3D scene reconstruction i.e. denser depth information field compared to a convergent layout. On the other hand, parallel layouts provide less information about the depth perception. Convergent layouts, have better viewing angles and depth perception but, introduce keystone distortion which reduces the number of valid points in both image planes, and thus, leading to sparser scene reconstructions.

The answer to this problem in the river monitoring case is that the selection is river dependent. Streams or rivers with small width allow for both parallel and convergent layouts, which can be placed on the center of a bridge (Fig. 1). However, for the case of a river with large width only a convergent layout will be feasible due to the fact that the parallel layout has a restricted distance between the camera pair (≈ 5.5 cm) and thus, a restricted field of view.

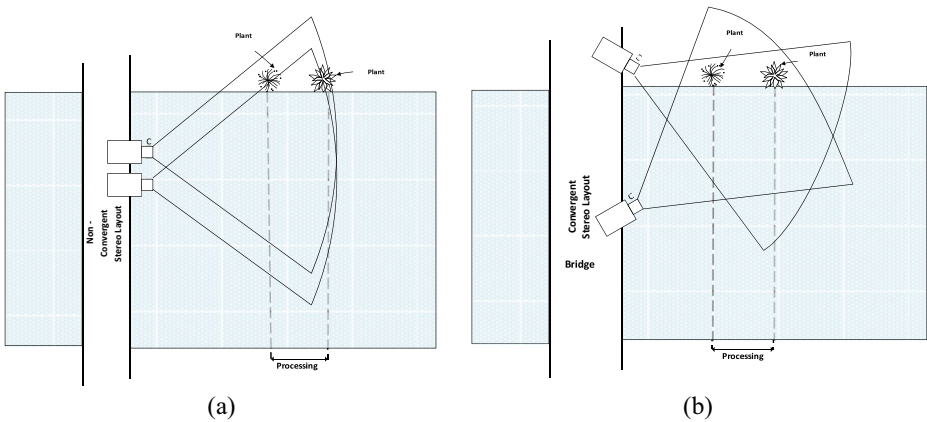


Fig. 1 Stereo layouts for river monitoring, (a) Non Convergent/Parallel layout and (b) Convergent Layout

- 1) *Physical to image plane coordinates*: The role of the stereo rig is to connect the physical and image plane coordinate systems so that an estimate of the depth of the scene can be acquired. This will allow defining a region of known dimensions that will provide the information required to perform the transformation of the image-based velocity estimate to its corresponding real world velocity approximation. The relation between the physical and image plane coordinate systems is given by:

$$\begin{aligned}
 \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} &= K \cdot [R|T] \cdot \begin{bmatrix} x_{world} \\ y_{world} \\ z_{world} \\ 1 \end{bmatrix} \\
 \Rightarrow \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} &= \begin{bmatrix} f & & x_o \\ & f & y_o \\ & & s \end{bmatrix} \cdot \begin{bmatrix} R_{11} & R_{12} & R_{13} & | & T_x \\ R_{21} & R_{22} & R_{23} & | & T_y \\ R_{31} & R_{32} & R_{33} & | & T_z \end{bmatrix} \cdot \begin{bmatrix} x_{world} \\ y_{world} \\ z_{world} \\ 1 \end{bmatrix} \\
 \Rightarrow x_{im} &= P \cdot X_{world}
 \end{aligned}
 \tag{6}$$

with K : matrix containing the intrinsic characteristics of the camera, required for the relation between the camera and the image plane coordinate systems, $[R|T]$: matrix denoting the extrinsic parameters of the system, i.e. the rotation (3x3 matrix) and translation (3 x 1 vector) required to match the world and camera coordinate systems and P : the projection matrix, defining the geometric mapping of points from one plane to another.

The correspondences between the image points and the world points are established based on Eq. (6), and concern both the intrinsic camera characteristics (if not known) as well as the extrinsic characteristics of the camera to world relation. This relation is expressed based on Zhang’s methodology [34] combined with Heikkila’s and Silven’s intrinsic model [16]. The estimation approach follows the implementation employed in Bouquet’s camera calibration toolbox from Caltech1 involving the use of a targeted pattern. In our case the targeted pattern is a chessboard, placed in the river banks, moved at different heights and positions, acting as a reference coordinate system.

- 2) *Relating the image planes of the camera pair*: Having related the physical and image coordinate systems for each camera the next step is to relate the camera pair together. This process involves the disparity map estimation. For the case of a non-convergent layout the

image planes are parallel on the y-axis, meaning that the difference between a point in the first camera is just a translation in the x-axis, i.e. a point at the position (x, y) in the first image will be located in the position $(x + d_b, y)$ in the second image. For the case of a convergent layout in order to relate the image planes we need first to rectify them (Fig. 2) so that the points can be associated with a translation in the x-axis only, and finally compute the disparity map.

The relation between the 3D coordinate vector $X_{world} = (x_w, y_w, z_w)$ and the corresponding point vector in the camera reference coordinate point $X_{camera} = (x_c, y_c, z_c)$ is defined by the extrinsic parameters (rotation R_{camera} , translation T_{camera}). We assume here that the intrinsic camera parameters are the same and that this mapping is easily inverted. In the case of the camera pairs, the camera projection matrices can be expressed as follows:

$$\begin{aligned} X_{left} &= P_{left} \cdot X_{world} = R_{left} \cdot X_{world} + T_{left} \\ X_{right} &= P_{right} \cdot X_{world} = R_{right} \cdot X_{world} + T_{right} \end{aligned} \tag{7}$$

The aim is to relate the two image planes in order to associate corresponding points but also to define the relations between the image planes and the physical world. For this purpose, we employed Bouquet’s algorithm for stereo rectification [29]. This method relates the left camera’s image plane to the right camera’s image plane by applying appropriate rotational and translational transformations forming the projection matrix of the left image plane in accordance with the one of the right camera’s:

$$P_{left} = R \cdot P_{right} + T \tag{8}$$

with R and T being the rotation matrix and translation vector respectively relating the two planes formed as follows:

$$R = R_{right} \cdot R_{left}^T \text{ and } T = T_{left} - R^T \cdot T_{right} \tag{9}$$

where R and T essentially rectify the coordinates of the right camera X_{right} to those of the left camera as $X_{Rect_right} = R X_{right} + T$. Furthermore, although these transformations lead to coplanar camera systems, they do not achieve the desirable parallelism of the (coinciding) epipolar lines to the x-axis of camera systems. This alignment between the epipolar lines and the horizontal axis is crucial, since it restricts the search for matching points on the two images within a single line. To achieve epipolar line alignment with the x-axis, a rotation matrix R_{rect} must be computed, consisting of three epipolar unit vectors with mutual

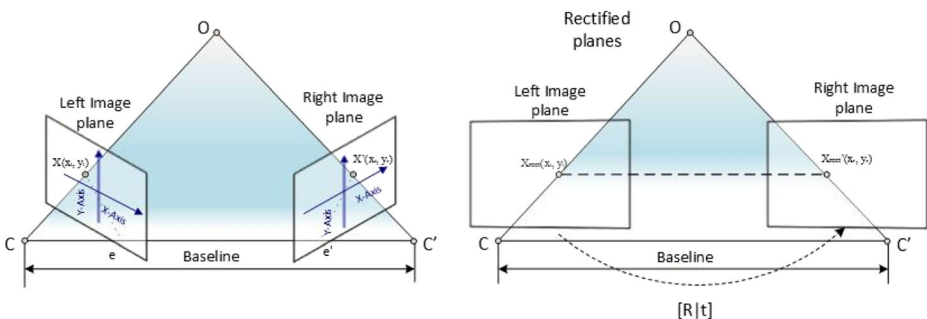


Fig. 2 The rectification process of transforming the image pair, with C, C' being the principal points, and $[R|T]$ the transformations relating the two planes

orthogonality. This matrix is then applied to the left projection matrix in order to move the left camera’s epipole to infinity and align the epipolar lines horizontally, leading to row alignment of the image planes. Overall, the epipolar line alignment of the two cameras with the x-axis is achieved by setting the rotation matrices as:

$$R_{left} = R \cdot R_{rect} \quad \text{and} \quad R_{right} = R$$

By rectifying the image planes, the epipolar lines of the two camera images become parallel to the x-axis, leading to simple x-axis disparities of corresponding points in the two images. Essentially, based on these linear relations, we can derive the physical point coordinates on which each image point is back projected, through a process known as triangulation [13], which will be briefly presented in the following paragraphs. Nevertheless, in many real-world applications including traffic monitoring, surveillance, human motion and river flow, the motion is still in two dimensions in the world coordinate system, with the third dimension of height being of minor importance. Focusing on such applications, we can limit the back-projection to only two dimensions, which is exactly the case of mapping the perspective onto the projective mapping (Fig. 3), which performs a distance and motion scaling from the camera plane to the 2D world plane parallel to the observed surface. This orthorectification process allows the extraction of accurate flow data that correctly approximate the real motion of the river.

- 3) *Disparity and depth map estimation:* Having related appropriately the image planes, the points of each plane are associated in the form of a translation in x-axis. The disparity map d contains the distance between corresponding points in the left and right image of a stereo pair. For the computation of the disparity map various approaches have been proposed, with the theoretical basis being similar with the methods used for pixel motion estimation. In our case we employ a variational approach developed by J.Ralli [25], due to the fact that it produces denser and more accurate disparity maps compared to a block matching-based disparity estimation method.

The depth estimation is performed through the back-projection of each 2D image plane point into its corresponding point position in the 3D world coordinate system. This process, known as triangulation, utilizes image-point correspondences in the image pair of a stereo layout to derive estimates of the Homography matrices, essentially the projection matrices P , which map the 3D world point to its corresponding 2D image point (Eq. (1)). The derivation of the point correspondences in the two image planes is performed using the disparity map d , which indicates the point relation between the two planes. As mentioned before, the rectification process attains the matching of the two camera planes through a simple translation in the x-axis, implying that a point $x_{left} = (x, y)$ in the left camera image plane is matched to the point $x_{right} = (x', y) = (x + d_i, y)$ in the right camera image plane, with d_i being the disparity for the given point. Utilizing these point correspondences among matched pixels and the corresponding disparities can recover the 3D world coordinates through the mapping relations of Eq. (7). More specifically, following the DLT method in Hartley and Zisserman [13], we end up in a system of linear equations of the form $B \cdot X_{world} = 0$, in which $X_{world} (x_w, y_w, z_w)$ denotes the 3D world point to be estimated and B being a matrix of the following form:

$$B = \begin{bmatrix} x \cdot P_{left}^3 T - P_{left}^1 T \\ y \cdot P_{left}^3 T - P_{left}^2 T \\ x' \cdot P_{right}^3 T - P_{right}^1 T \\ y' \cdot P_{right}^3 T - P_{right}^2 T \end{bmatrix} \tag{10}$$

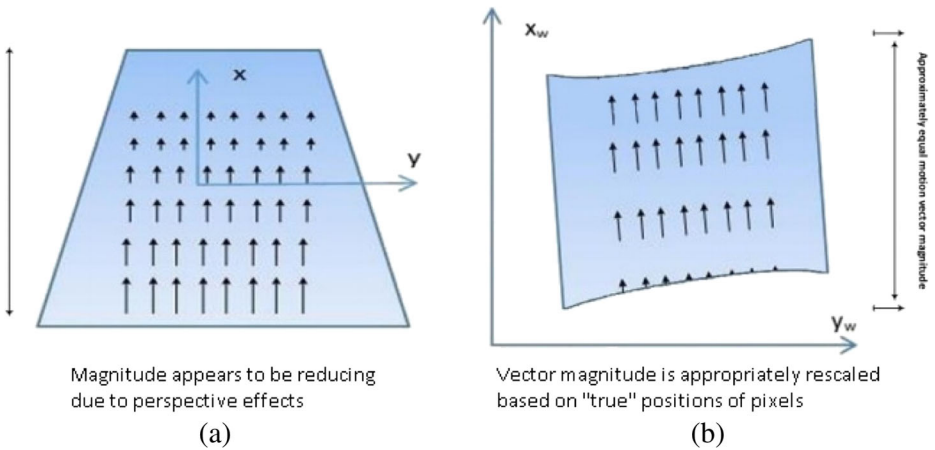


Fig. 3 (a) The distorted scene representation as recorded from the camera, (b) the undistorted scene representation after the orthorectification process

with p_{left}^i, p_{right}^i being the i -th row vector of the projection matrices P_{left}, P_{right} for the points $x_{left} = (x, y), x_{right} = (x', y')$ respectively.

This is a direct benefit of our approach using a calibrated stereo layout, where the projection matrices for homographic mapping between the 2D image coordinates and the 3D world coordinates of each point can be estimated. This process takes advantage of the fact that paired points in the associated image planes are related to the corresponding 3D world point through the projective mappings presented in Eq. (6). The solution of the linear system and consequently the derivation of the 3D world point are achieved through the singular value decomposition (SVD) of the matrix B . This matrix is factorized in the form of UDV^T , with U and V being orthogonal matrices and D being a diagonal matrix with non-negative entries. The 3D world point is determined by the unit singular vector that corresponds to the smallest singular value of the matrix B , which is equivalent to the (unit) eigenvector of $B^T B$ with least eigenvalue, under the assumptions that $B^T B$ is invertible and B is a $m \times n$ matrix, with rank n and $m > n$ [13].

4.2 Optical flow estimation

The next step of the estimation process is the computation of the 2D image velocity field. For this task as presented in the previous section we use a probabilistic method for the computation of the optical flow field of a fluid based on the methodology introduced by Chang et al. [4]. However, the disadvantage of a probabilistic methodology is that its accuracy depends upon the prior assumptions about the data distribution characteristics which in turn are highly affected by the amount of the available data. Fewer data result in an over fitting estimate with the outliers further reducing the accuracy of the estimate.

To solve this problem, we can use a number of subsequent frames to increase the amount of data used in the estimation process, or simply increase the size of the interrogation window for velocity estimation. Such approaches, however, increase the spatial or temporal extent of the assumption on velocity uniformity. In our scheme, we use a regularization factor λ to penalize the outliers' importance, keeping the estimate robust without the use of extra data. In addition,

we incorporate the extra information available in the stereo layout in the estimation process, by means of the complementary pixels in the two camera views. By using supplementary information, we can not only increase the estimation accuracy, but also reduce the window size of velocity estimation, increasing the detail of the estimated velocity field. At a second stage of development, we can also exploit the stereo layout in order to define the appropriate coordinate transformations as above and directly recover an estimate of the 3D motion field. These developments are considered in the next two subsections.

4.2.1 Exploitation of the stereo layout for improving estimation accuracy

The estimation of the local displacement probabilities A_i for each pixel position in the candidate neighborhood is an inference problem leading to a least squares solving scheme which is high dependent on the amount of data used. One way to increase the estimation accuracy and at the same time reduce the amount of data used is to utilize the extra information provided by the bi-channel formulation of the stereo scheme. The pixel intensities of the two camera pairs, that have been set up in a parallel layout or have been rectified when a convergent layout is used, are related based on the disparity d as follows:

$$I_{left}(x, y, t) = I_{RectRight}(x + d, y, t) \tag{11}$$

In our approach the disparity is computed based on the work of J. Ralli [25] who developed a variational framework for pixel-wise relation between two different views. For the subsequent frames, where a displacement for the pixel (x, y) occurs the relation between the paired images for the stereo layout is defined based on both the displacement $(\Delta x, \Delta y)$ as well as the disparity estimate (Fig. 4):

$$I_{left}(x + \Delta x, y + \Delta y, t + 1) = I_{RectRight}(x + d' + \Delta x, y + \Delta y, t + 1) \tag{12}$$

with $d' = d + D_t$ indicating the disparity of the shifted pixel, which is essentially a change in the disparity value of the previous frame at a constant D_t .

For each pixel intensity at each camera based on the STAR model (Eq. (3)) we have the relation:

$$I_K(x, y, t) = \sum_{i=1}^D A_i \cdot I_K(x + \Delta x_i, y + \Delta y_i, t + 1) \tag{13}$$

where $K = \{left, right\}$ denoting the intensity of the image for each camera.

Combining Eqs. (12) and (13) allows us to express Eq. (11) in the following form:

$$\sum_{i=1}^D A_{left,i} \cdot I_{left}(x + \Delta x_i, y + \Delta y_i, t + 1) = \sum_{j=1}^D A_{RectRight,j} \cdot I_{RectRight}(x + \Delta x_j + d', y + \Delta y_j, t + 1) \tag{14}$$

If we compare Eq. (14) in parts and with consideration that the intensities of each image of the camera pair at time $t + 1$ are connected based on Eq. (12) we can conclude

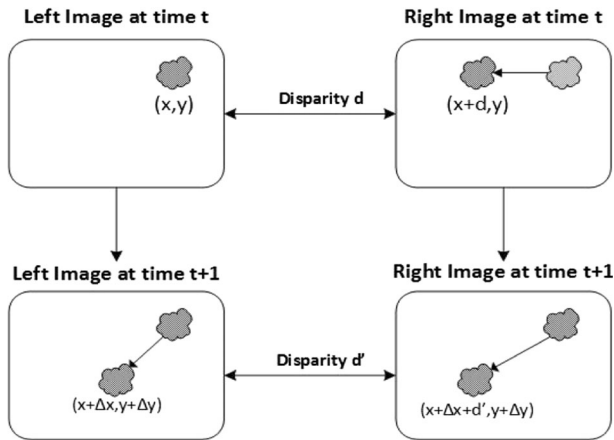


Fig. 4 Relation between the stereo image pairs and their optical flow estimates based on the disparity map

that that given the disparity estimate the coefficient depicting the translation to a certain position must have the same value in both the images in the image pair, i.e. $A_{left,i} = A_{right,j}$ with $i = j$.

The coefficient estimation for the case of a single camera, in a matrix formation solving the least square estimation scheme of the form $Ax = b$ with respect to A instead of x . For the case of a stereoscopic layout the matrices A , x , b are defined as:

$$x = \begin{pmatrix} I_{left}(x_1 + \Delta x_1, y_1 + \Delta y_1, t + 1) & \dots & I_{left}(x_N + \Delta x_N, y_N + \Delta y_N, t + 1) & I_{Right}(x_1 + d, y_1 + \Delta y_1, t + 1) & \dots & I_{Right}(x_N + d, y_N + \Delta y_N, t + 1) \\ 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \dots & \vdots & \vdots & \dots & \vdots \\ I_{left}(x_N + \Delta x_N, y_N + \Delta y_N, t + 1) & \dots & I_{left}(x_N + \Delta x_N, y_N + \Delta y_N, t + 1) & I_{Right}(x_N + d, y_N + \Delta y_N, t + 1) & \dots & I_{Right}(x_N + d, y_N + \Delta y_N, t + 1) \end{pmatrix}$$

$$A = \begin{bmatrix} A_{left,1} \\ \vdots \\ A_{left,D} \\ A_{right,1} \\ \vdots \\ A_{right,D} \end{bmatrix}$$

and

$$b = \begin{bmatrix} I_{left}(x_1, y_1, t) \\ \vdots \\ I_{left}(x_N, y_N, t) \\ I_{Right}(x_1 + d_1, y_1, t) \\ \vdots \\ I_{Right}(x_N + d_N, y_N, t) \end{bmatrix}$$

with A being a $2N \times 2D$ matrix, x being a $2D \times 1$ vector and b being a $2N \times 1$ vector, and N being the size of the spatial neighborhood N_s and D the size of the destination neighborhood D_s .

Given the assumption that the transition coefficients (A_{left} and A_{right}) for each translation in each image pair must have the same value we can express the coefficient vector A as:

$$A = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \cdot \begin{bmatrix} A_{left,1} \\ \vdots \\ A_{left,D} \end{bmatrix} = [I \ : \ I]^T A_{common}$$

Essentially expressing the coefficient matrix A in form, $A = C * A_{common}$, where $C = [I \ : \ I]^T$ is a $2D \times D$ matrix and A_{common} being a $D \times 1$ vector. This formulation allows us to estimate the transition coefficients, removing the duplicates since we have shown that $A_{left,i} = A_{right,j}$, with $i = j$.

By doing this we have doubled the information used to approximate the transition probabilities, modeling the estimate considering the illumination variation encountered between the stereo image pair. This formulation enables us to use smaller windows sizes to reach the same accuracy compared to the case of a single camera layout. This formulation allows us to estimate a common set of transition coefficients for the two cameras since $A_{left,i} = A_{right,j}$, with $i = j$, utilizing matched intensity information from both cameras as to increase estimation accuracy. Essentially, in this form we double the intensity values used to approximate the transition probabilities for each point and model the estimate by also considering the illumination variation encountered between the stereo image pair. This formulation enables the use of smaller windows sizes to reach the same accuracy compared to the case of a single camera layout.

A possible drawback of introducing stereoscopy in the estimation of displacement distribution pertains to the introduction of an error factor associated with the stereo camera model and the computation of an accurate disparity map estimate that provides the correct relation between the pixel points of the stereo image pair. The error sensitivity can be reduced by constraining the estimated model with the addition of a regularization parameter λ , which controls both the modeling and computation errors for the original ill-posed inversion problem, leading to Eqs. (4) and (5) being reformulated as:

$$J(A) = \sum_{s=1}^N (I(x_s, y_s, t) - k_s^T A)^2 + \sum_{j=1}^D \lambda^2 \cdot A_j \tag{15}$$

$$A = (K^T K + N \cdot \lambda^2 I)^{-1} K^T M \tag{16}$$

with K being the matrix containing the vectors k_s with the intensities for all the possible transitions for each pixel contained in the spatial neighborhood N_s defined by a central pixel (x_s, y_s) , and M being the matrix containing with the intensities of the pixels belonging to spatial neighborhood N_s of the central pixel.

Finally, the estimation of the global motion field is achieved through the utilization of the Bayesian inference scheme, described in detail in our earliest work [1], using a Maximum a posteriori (MAP) formulation that enables the relation the regional information contained in

the likelihood models with the addition of a smoothness factor leading to a dense and highly coherent flow field.

4.2.2 Optical flow estimation in world coordinate system

Up to this point we have computed the optical flow field of the scene using the full intensity information from the stereo camera system, but all the computations pertain to the common 2D reference system used by the camera planes. Since we have also derived the disparity map for the two cameras in parallel concatenation, as well as the general concatenation scheme, we also have the means of transforming the 2D back to the 3D world space. Thus, in this section we derive the mapping of motion vector field from the 2D reference system of cameras to the 3D world coordinate system. Relating the 2D image based motion vector to its corresponding 3D velocity vector can be viewed as a mapping problem similar to the one of relating a 2D image point to its corresponding 3D world point.

Let us consider the 2D motion vector $u_{im} = (v, \nu)$ and its corresponding 3D motion vector $u_w = (v_w, \nu_w, z_w)$ in the 3D world coordinate system. For each point X_w in the 3D world coordinates system based on Eq. (6) and the fact that the motion vector is expressed as a displacement between two point positions:

$$X_{w,t+1} - X_{w,t} = P \cdot (x_{im,t+1} - x_{im,t}) \quad (17)$$

For the corresponding motion fields, we also need to find the transformation H that maps the two motion fields in the form of:

$$u_{im} = H \cdot u_w \quad (18)$$

By expanding this equation, we can show that the transformation H that maps the two motion fields is actually the projection matrix P that relates the 2D image and the 3D world coordinate systems:

$$\begin{aligned} u_{im} &= H \cdot u_w \\ \Rightarrow x_{im,t+1} - x_{im,t} &= H \cdot (X_{w,t+1} - X_{w,t}) \end{aligned} \quad (19)$$

Assuming that the temporal matching of 3D points is accurately mapped on 2D, i.e. the recovered and actual pairs $(X_{wi,t+1}, X_{wi,t}), (X_{w,t+1}, X_{w,t})$ define the same relation, then the required transformation for the motion vector is the projection matrix P . Thus, there can be defined an intermediate relation between the 3D velocity field u_w and the 2D motion field u_{im} based on the knowledge of the projection matrix, which defines the relation between the two coordinate systems. The 3D motion vector of a point X_w in the world coordinate system can be now be defined as the displacement difference between the initial reference position and the estimated final position of the point derived through the back-projection of the 2D image plane position of the point to its corresponding position in the 3D world coordinate system.

The main trend in the river velocity can be considered as the average of several instantaneous velocities in the world system, i.e. for M relevant points of interest:

$$u_{river} = \frac{\sum_{i=1}^M u_{w,i}}{M} \quad (20)$$

In order for this velocity estimate to be considered as the river's surface velocity we need to validate the assumption that this velocity will almost be constant over the area of examination. Of course, this assumption will only show validity in cases of rivers where the flow is unobstructed and is not accelerated due to the change of the ground's inclination. Notice that the use of a world coordinate system and the projective mapping in the expression of the motion permits the direct averaging of vectors estimated at different points of the river, since they are devoid of perspective effects depending on the viewing depth.

4.2.3 Area of interest for motion estimates

The estimation of the optical flow field of the river flow includes some unwanted motion vectors. For the case of the river flow motion field as unwanted we consider the motion vectors that describe the dispersion phenomena of the water hitting the river banks or the rock formations inside the flow. These motion vectors should not be involved to the main motion trend of the river flow and thus, they are not useful for the estimation of the velocity of the flow.

In order to isolate the area that is covered with water, we implement a segmentation approach. This process will separate the river body from the vegetation around it. To do so, we transformed the river image into the HSV color space and performed a K-means classification task on the hue and saturation channels of the image. In this way, the image is separated into two classes (Fig. 5), one containing the vegetation, rock formations, leaves flowing in the water etc. and the other containing the water pixels. In order to distinguish which of these clusters contains the water pixels we used an evaluation metric. The group class that has the largest average value, computed by the associated pixels, contains the water pixels.

By isolating the river water body from the rest of the information that is included in the cameras view, we reduce the unwanted flow vectors that represent the vegetation motion and the diffusion effects in the river banks and at the rock formations. However, this is not enough since this process does not isolate the surface flow vectors of the river flow that are produced by the wind or the illumination variance. Concerning the wind induced motions the conventional measuring systems form a tunneling effect in which the water flows to its natural motion direction, without being disturbed by wind changes in the surface of the flow. This is something that we have to deal with as to reduce the influence of such motion types in our method.

Since we cannot constrain the river flow conditions, we need to define a means of distinguishing the main motion of the flow from the motion vectors caused by external forces, such as wind. As a solution, we applied a classification method on the motion vectors of the estimated flow field. In our earliest work [1], we showed that a supervised classification method provides the best results compared to an unsupervised classification method. Specifically, we concluded that a supervised Naive Bayes classifier is the most accurate classification method to distinguish the main motion trend in the flow. The supervised classifier is trained to classify the motion vectors into 8 motion directions (North, South, East, West and the four subsequent motion directions). Since there is always a hint for the main motion direction of the river we can easily mark which of these motion directions are considered valid for the main trend of motion, e.g. the river flows to the North so we accept North, North-East and North-West motion directions. The optical velocity of the fluid can be computed by simply taking the average magnitude of the motion vectors belonging to the main motion trend of the fluid. In the next subsection, we present the method for the computation of the 3D motion vectors of the points associated with the main trend of motion.

4.3 Using the 3D velocity estimate for validation

To validate the estimated velocity, we can make use of the existence of physical particles in the flow, such as leaves or flowing wood parts. In fact, if we estimate the motion field in world coordinates, then we can assume uniform 3D velocity over the river surface. Thus, if we locate one drifting object at one position, we can estimate its position after certain time. Then we can map this position on the 2D camera space through the perspective and search for the object in this position. Such validation may be performed under the notion that the computed leave velocity might be different from the main trend of the river motion or the fluid velocity on the surface of the river. When using natural particles with physical dimensions, i.e. objects such as leaves, we cannot guarantee that the influence of external forces, such as wind or turbulence, will not dominate the particle's motion forcing the particle to follow a different motion from the typical fluid field. Nevertheless, under the assumption of minimal external forces, the validation process can be performed as follows.

- (1) *Select the particle to be tracked* - The selection can be either user assisted, i.e. the user selects the particle he considers the most appropriate manually in the image, or an automated particle identification can be performed. For the latter, we can make use of the segmentation method used in previous sections. We can observe in Fig. 5 that the particles flowing in the river case (e.g. leaves), have been assigned to the non- water class. We can perform a supervised segmentation process on this class moving to the YCbCr color space, aiming at a specified value range of the channels Cb and Cr that corresponds to the particle's color values, brown color in the leaf's case. The identified particle is then stored as a reference pattern.
- (2) *Search and Validate* - The validation is performed by searching the specified particles after a number N of frames, under the assumption that the estimated 3D velocity estimate remains constant over the river length. The reference patterns are compared with each isolated particle in the subsequent frames by assessing their similarity rates on their SURF features [2]. The constant velocity is used to match the particle's 3D position at the initial time stamp and the estimated position based on motion estimation. Afterwards, by back-projecting the estimated position on the camera plane(s) we examine the particle's presence within the specified region of the frame(s).

In essence we need to find the time required for the particle to travel from its current computed 3D position (defined by its centroid) to the position of the first indicator (entry to

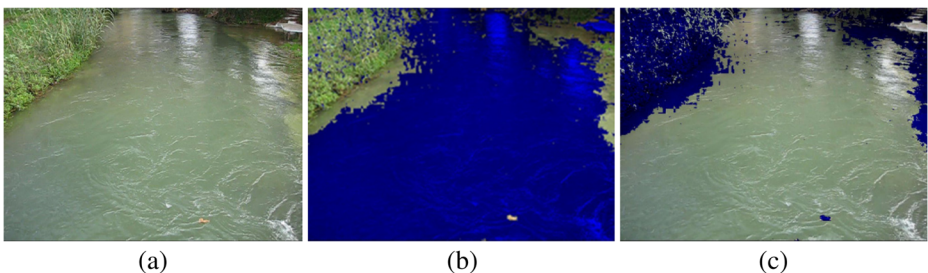


Fig. 5 River isolation via segmentation, (a) Initial image of Koiliaris River, (b) Class containing vegetation, rock formations, leaves and (c) Class containing the pixels belonging to the river

region of interest) as well as the second indicator (exit from region of interest) given the estimated average 3D velocity:

$$\begin{aligned} X_{w,indicator} - X_{w,c.p.} &= u_{river3D} \cdot t \\ \Rightarrow t &= \frac{X_{w,indicator} - X_{w,c.p.}}{u_{river,p}} \end{aligned} \quad (21)$$

where $X_{w,indicator}$, $X_{w,c.p.}$ are the 3D point coordinate vectors representing the current particle position and the position of the region's indicator.

The number of frames we need to go forward to search the particle can be approximated by dividing the previously computed time interval t with the camera's frame rate value.

The verification of the particle at this position validates both the estimation accuracy as well as the velocity-constancy assumption. The most important assumption behind the velocity estimation method and its validation is the velocity constancy in its 3D form, which is devoid of perspective effects associated with the 2D mapping. In real-world applications, the 3D velocity constancy assumption can be justified better than in the 2D case, where velocity needs to be adjusted with time and scene depth as to accommodate the perspective effects.

The reference particle pattern will be searched within this region using the same SURF feature extraction and identification process mentioned earlier. The presence of a particle or a characteristic particle portion within this region can be considered for the accuracy assessment of the estimated instantaneous velocity of the river (Fig. 6).

To further increase the estimation confidence, we can also back-project to the expected particle position in the two cameras. In such case, the particle's position will be different on each image plane of each camera defined by the direct mapping from 3D to 2D via the corresponding matrix projection. Then, by taking advantage of the disparity map we can relate the pixel positions and the motion vector estimates between the two image planes. We can use this relation to search the particle in the two camera planes thus, increasing the tracking validity of the particle.

5 Results

We have designed an initial version of the monitoring system presented above which has been tested in Koiliaris river, Chania, Crete. In our study site, Koiliaris river has a small cross-section (≈ 8.5 m) so both parallel and convergent layouts can be implemented. We selected a nearby bridge and placed there a convergent layout to examine how the convergent layout will perform at a small-scale experiment (Fig. 7). The camera pair is rotated about 12 degrees in the y -axis and ± 4 degrees in the x -axis respectively in order to have a panoramic view of the scene. The baseline between the camera centers is 88 cm and the cameras are rotated about 2 degrees and -3 degrees in the x -axis for the left and right camera respectively. Finally, the camera pair placed at tripods have 5.25 m distance from the river's surface. The final system will have the cameras placed at permanent positions on the bridge and with a wireless transmission system in place (in our initial experiment an Ethernet connection to a modem was used).

Starting from the stereo layout and the depth estimate, we first calibrated the cameras by placing a chessboard pattern at the river banks, as mentioned earlier, gaining the information required to relate the physical and the image coordinate systems. As indicators for the region formulation we used a bunch of reeds present at the river banks (Fig. 7). The distance of the

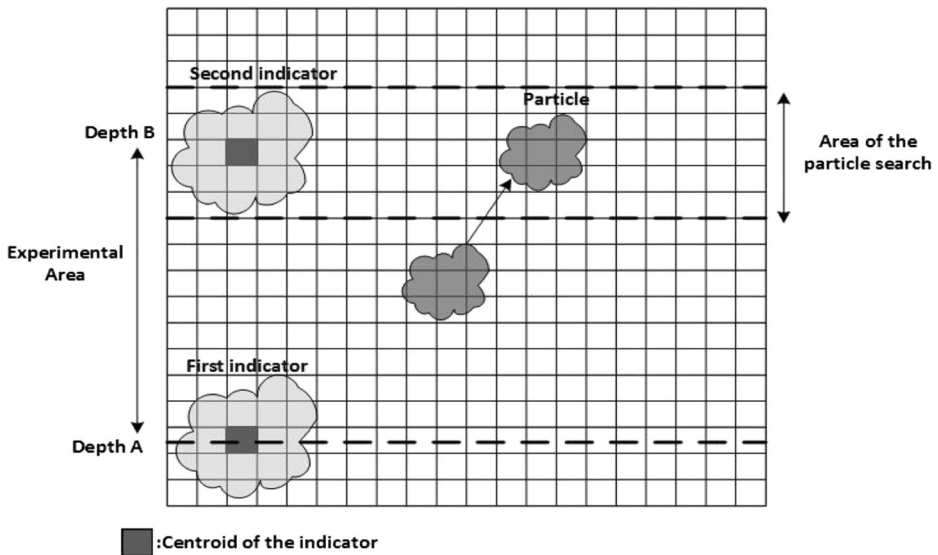


Fig. 6 Formulation of the search area for the particle in each image plane by defining a window with a horizontal size equal to the image's width and vertical size the size of the particle plus M pixels up and down. In the figure the particle is about 3 pixels in size and the search window has a height 5 pixels, meaning that $M = 1$. The window is formulated around the centroid of the second indicator

first reed bunch (indicator 1) from the stereo layout was 16.83 m and the second bunch 18.93 m. Our estimates are 18.12 m and 20.01 m respectively. We can see that although the deviation between the indicator estimates is 1.29 m, yet the distance between the indicators, i.e. the region of interest has a deviation of 0.21 m, meaning that the region's length is computed very accurately. Regarding the acquired dataset size that was used during experimentation, we acquired 2-min videos (with 1024×768 resolution, and a frame rate of 30fps) of the flow during each of our 4 measurement sessions. The computational time required for extracting the optical and real-velocity estimate, given a pre-calibrated layout, was approximately 128 s, given 2 subsequent frames of the stereoscopic layout.



Fig. 7 The viewed scene from the stereo camera pair (convergent camera placement) without rectification in Koiliaris River, (a) left camera and (b) right camera

All numerical computations were performed in an ordinary laptop, consisting of an Intel Core (TM) i5-3230 CPU and 4Gb RAM, using a GUI tool developed and executed in MATLAB programming environment. The aforementioned computational timeline involves the execution of the following processes: (a) Rectifying frames, (b) Disparity computation, (c) Color-based river segmentation, (d) Optical flow estimation, and, (e) 2-D to 3-D coordinate mapping and 3-D displacement vector computation. The validation process was considered as a post-processing step introducing an additional 30 s fraction to the total processing time involving the particle detection and tracking process in the corresponding frames. However, it is important to mention that the proposed methodology is in its design and prototype build phase thus, no optimization in terms of computational efficiency has been performed, that would allow real-time computing. To our knowledge so far, real time or close to real time estimations is a limitation that characterizes similar river flow monitoring systems face. Future work will aim to address this issue.

5.1 Camera-based and conventional equipment-based real world velocity estimations

Starting from the outmost aim of this paper, i.e. the estimation of the 3D surface velocity of the river, our experiments, under almost ideal measuring conditions (no rain or cloudy days) show an average deviation of ± 0.01745 m/s between the real river surface velocity measured using conventional equipment (accelerometers and Doppler-based devices) and the velocity estimate derived using the proposed image based method. Table 1 depicts the velocity estimates for both the conventional equipment- based and the image based measuring approaches for a number of measuring sessions.

Table 1 shows that our approach shows an average deviation of ± 0.01745 m/s between the real river surface velocity measured using conventional equipment (accelerometers and Doppler-based devices) and the velocity estimate derived using the proposed image based method. The velocity estimate using conventional equipment (i.e. accelerometers) is considered to be the average surface velocity \pm the variance, as measured throughout the monitoring session. This deviation shows that the proposed approach even at this current prototype stage produces close to actual surface velocity estimates. Even more this deviation can be further reduced through the performance and accuracy optimization of each of the system's components.

Commenting further on our result, the majority of the state of the art systems measures the system's accuracy with respect to estimated water discharge instead of using the surface velocity. Moreover, we have not yet collected measurements of Koiliaris river surface

Table 1 Deviation between the estimated river surface velocities measured with conventional equipment and the proposed image-based method

Test cases	Conventional equipment	Proposed method	Deviation
1	0.3993 ± 0.0331 m/s ¹	0.4124 m/s	+0.0131 m/s
2	0.3993 ± 0.0331 m/s ¹	0.3864 m/s	-0.0129 m/s
3	0.6251 m/s	0.6009 m/s	-0.0242 m/s
4	0.7143 m/s	0.6947 m/s	-0.0196 m/s
Average	-	-	0.01745 m/s

¹ The first two measurements concern the same monitoring session during which two video sequences were acquired

velocities, since it is an ephemeral stream and we didn't have the chance to obtain and measure a flood event or, at least, surface velocities high enough to be used for validation of our methodology. Nevertheless, there do exist cases where other authors provide data for comparison and verification (Tsubaki et al. [30] and Bradley et al. [3]). With respect to the metrics used by these authors comparing the estimated and the actual surface velocity, we reached to the following comparative results:

	Metric	Their estimates	Our estimates
Tsubaki et al. 2011	Coefficient of determination (R2)	0.96	0.984
Bradley et al. 2002	Root Mean Squared Error (RMSE)	0.0032	0.0197

As we can observe, our estimates using those two metrics and comparing estimated and actual surface velocity, leads to better accuracy in both cases. However, a direct comparison between each system cannot be performed since the proposed systems have been tested on different rivers with different surface velocities and on different timelines and epochs, thus, the estimation was not performed using the same training and test datasets, i.e. monitoring sessions from the same river and ground-truth velocity field estimates.

5.2 Proposed river estimation method stage assessment

To sum up, the presented approach for the estimation of the 3D (real-world) average river surface velocity consists of the following steps: (a) Optical flow estimation and motion trend extraction, (b) 2D to 3D projection mapping and motion field relation, and (c) 3D motion field generation and velocity validation.

The first step, as presented in the previous sections, involves the extraction of the optical flow field and the derivation of the underlying motions. Figure 8 illustrates an instance of the extracted motion field of Koiliaris river, with different motion trends in the flow being colored

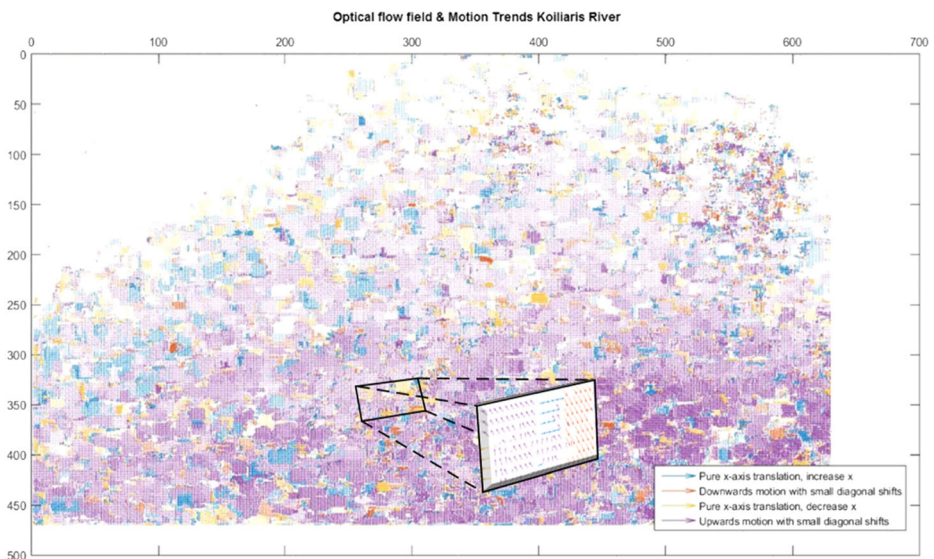


Fig. 8 An example of the estimated motion fields and the derived main trends using stereo camera layout and stereo-based probabilistic optical flow method

with different colors. In our test case, Koiliaris River, the flow trend (according to our monitoring position) is an upwards motion and, as it is shown in Fig. 8, this motion trend is verified by the main motion trend found in the derived motion field.

As far for the projection correction, which is the main factor related with the accuracy of the proposed work, we show that the perspective effect is corrected in the stereoscopic layout using the image plane and scene relation. This effect has to do with the fact that pixels in the far sight of the camera denote smaller motion vector magnitudes compared to the pixels in the front of the camera despite the fact that the motion is actually the same. To illustrate this effect, we isolated leaves flowing in the river surface and performed tracking to their motion in both viewing cases. The stereo based approach allows the correction of the perspective effect by associating the 3D world point information of the scene with the image plane information and the derived motion field. Figure 9 presents the test case and the rescaled optical flow field based on the real-world motion as estimated using the relations between the motion changes in each coordinate system.

The projection correction essentially scales the motion vectors according to the 3D world coordinate system. In Fig. 9, the 3D motion found in the initial and final leaf positions remains almost unchanged despite the fact that the image based motion differs by approximately 2 pixels.

5.3 Evaluation of the presented optical flow methodology

Beside the 3D river velocity estimation framework, in this article we have also presented a new stereo-based probabilistic optical flow estimation method. In the following paragraphs, we will present how the incorporation of the stereo – driven information in the optical flow estimation process enhances the estimation accuracy but also allows the reduction of the amount of data used in the estimation process. During this evaluation step, we compare the stereo-based approach with a single camera optical flow estimation methodology presented in our earliest work [1]. The two approaches are compared based on the estimation accuracy, the amount of data used and the post-processing capabilities that each method presents.

As far as the optical flow estimation accuracy is concerned, we observed small deviations the optical flow field estimates of both approaches. Specifically, based on the experimental results taken in Koiliaris River, approximately 22.7% of the main trend vectors change between the two optical flow derivation schemes. Both approaches succeed on retaining the main motion trend information as shown in Fig. 10. However, their main difference lies on their trade-off between the estimation accuracy and the amount of data used to achieve it.

Based on the amount of data used, the stereo case requires a smaller interrogation window size ~38% to reach an estimation accuracy with less than 10% deviation from the single camera approach. This is due to the incorporation of the data from the second camera which acts as an enhancement for the estimation process. To further back these observations we have tested both of these approaches with synthetic datasets using ground truth motion fields. The synthetic datasets were created based on the monitored scene. Motion fields were applied only to the leaves present at the flow. This serves a double cause, since despite the motion estimation accuracy testing it also indicates the ability of our approach to adapt to particle tracking, through the use of color-based segmentation approaches.

This dataset was generated for both single and stereo camera layouts. The motion formulations employed in the synthetic dataset consist of linear translations in each axis separately and diagonal translations ranging from 0 to 4 pixels. For the stereo case the disparity map d

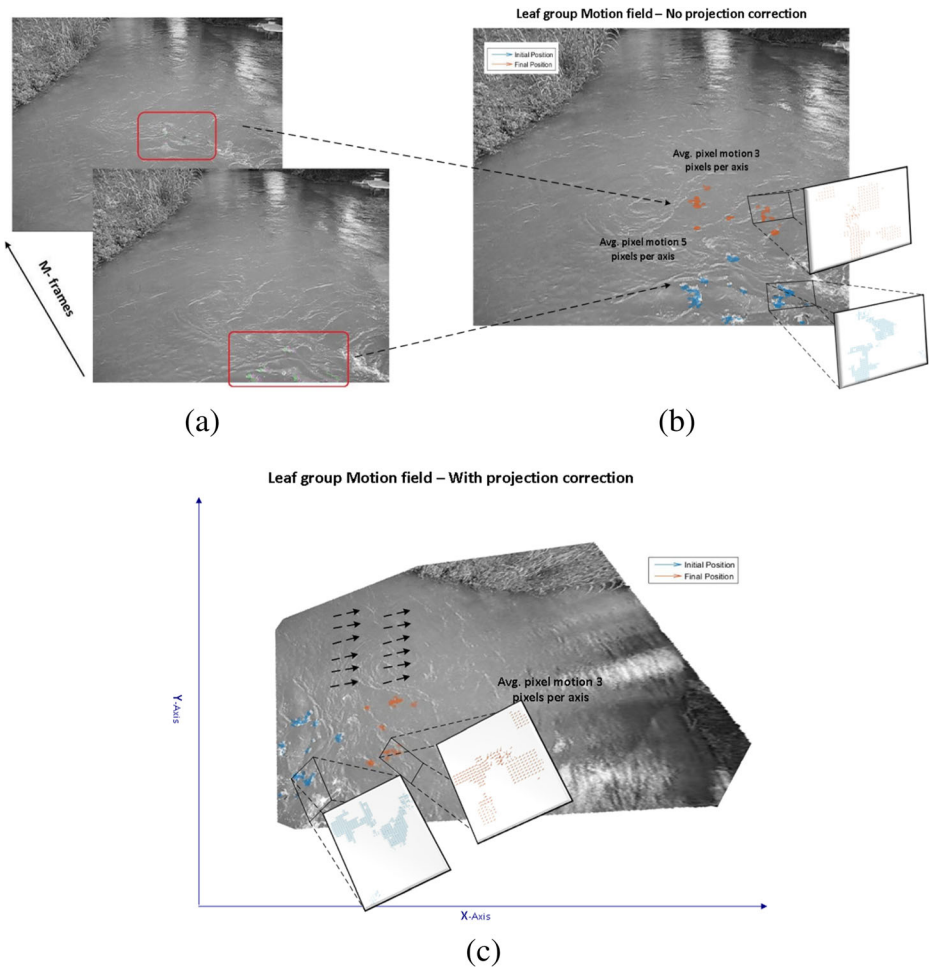


Fig. 9 Leaf bunch tracking in Koiliaris river test case, (a) leaf bunch being tracked, (b) corresponding motion field without projection correction placed in image coordinate system and (c) scene and motion field orthorectification based on the reconstruct 3D scene coordinates

maximum translation in the x-axis (due to parallel layout assumption) is of approximately 5 pixels. Figure 10 presents one of the generated synthetic datasets and the ground truth motion field. The selection of the translational motion patterns instead of rotational or spiral motions was based on the facts that the main trend of motion in the river flow which is followed ideally by all the tracked particles tends to follow ideally (assuming no rock formations exist within the river which may produce rotational or spiral motions due to diffuse effects) a linear translational motion pattern.

The single and stereo based approaches are compared both on the optical flow estimation accuracy as well as on the amount of data (window size) required. The optical flow field estimates of both the single and stereo optical flow estimation approach are compared to this ground truth optical field. The estimation accuracy assessment is performed based on the Angular (AE) and Endpoint (EE) error metrics [28]. Table 2 shows the average observed errors using the same window size (13×13) for each of the motion cases.

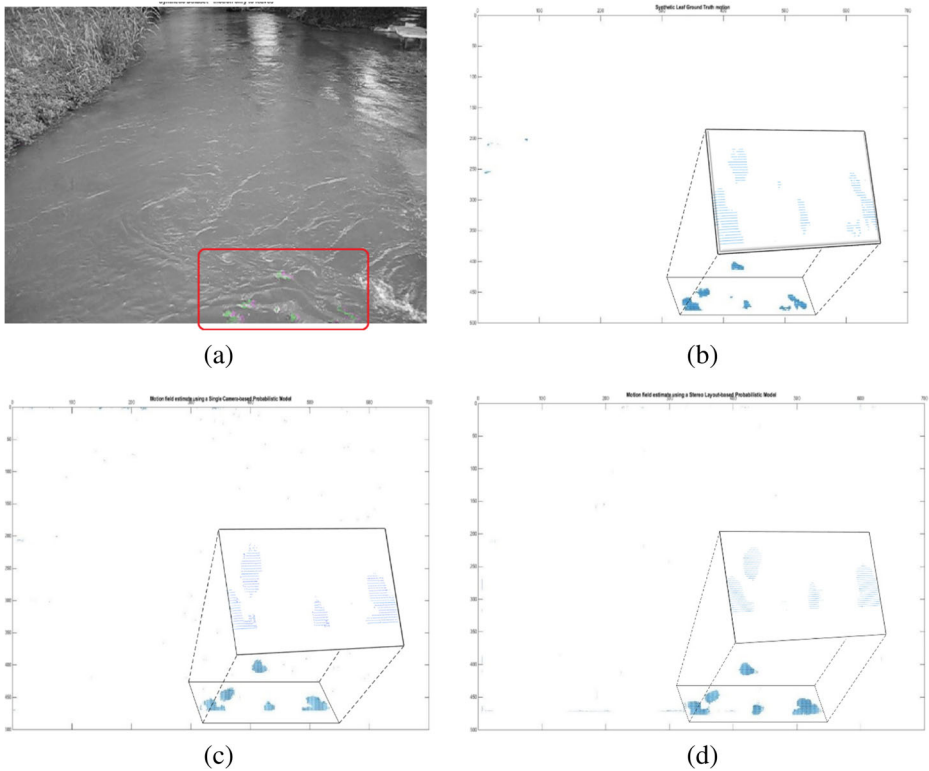


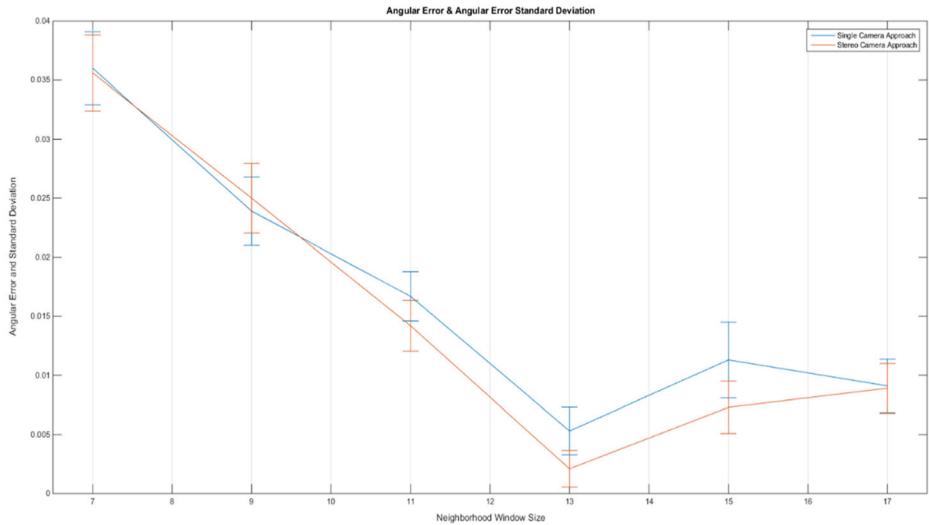
Fig. 10 (a) The synthetic reconstructed river scene with motion added on the leaves, (*green*) indicates the initial position and (*magenta*) the shifted position (a translational motion to the right direction), (b) the corresponding ground truth motion field, (c) the estimated optical flow field using the single camera probabilistic optical flow method, and (d) the estimated optical flow field using the stereo camera probabilistic optical flow method

We observe that both approaches lead to high accurate estimates with the stereo-based approach retaining more information about the magnitude information compared to the single-based method with a small loss in the directional information. The resulted errors show that a mean error Endpoint error of 0.2 pixels and an Angular error of lesser than 1 degrees is produced by the application of a Bayesian inference optical flow estimation technique. These results indicate a very good estimation accuracy in the estimation for both the directional and the amplitude information of the motion field. Comparing the stereo and the single camera approach we can observe that there exists only small deviation in the estimation accuracy between the two approaches.

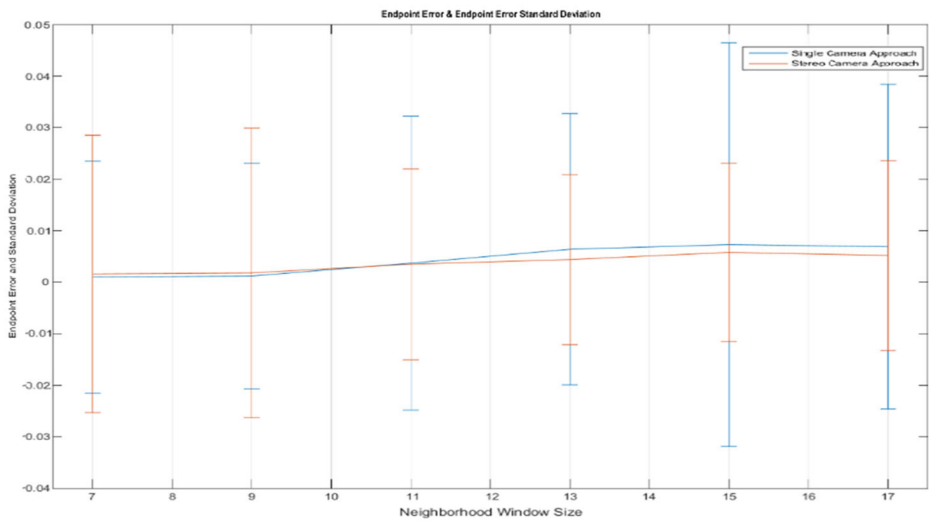
Table 2 Endpoint and Angular errors between the Stereo and the Single camera Probabilistic optical flow estimation approaches using for the stereo case a neighborhood size 20% smaller than the one used for the single camera optical flow probabilistic model

Optical flow method	Average AE	Average EE	Standard deviation AE	Standard deviation EE
<i>Single Camera Probabilistic</i>	0.0162	0.0066	0.0037	0.0283
<i>Stereo Camera Probabilistic</i>	0.0155	0.0039	0.0034	0.0209

However, as mentioned before, the most important benefit of the stereo method is the data reduction during the estimation process. In our test cases, we have observed that the use of the stereo based probabilistic optical flow estimation method can lead to approximately 20% window size reduction, without presenting loss in accuracy, compared to the use of the single camera probabilistic optical flow estimation method. Figure 11 presents the relation between the window block size variation (*window size range (7-17)*) and the estimation accuracy for



(a)



(b)

Fig. 11 Average Angular and Endpoint Errors and their standard deviations using various window sizes in the examination neighborhood formulation for the case of the stereo-based probabilistic optical flow estimation method. (a) AE and StD of AE for both methods, (b) EE and StD of EE for both methods, (b) AE, EE and their StDs for the proposed stereo-based approach compared to the optimal neighborhood window size found for the single camera probabilistic optical flow estimation method

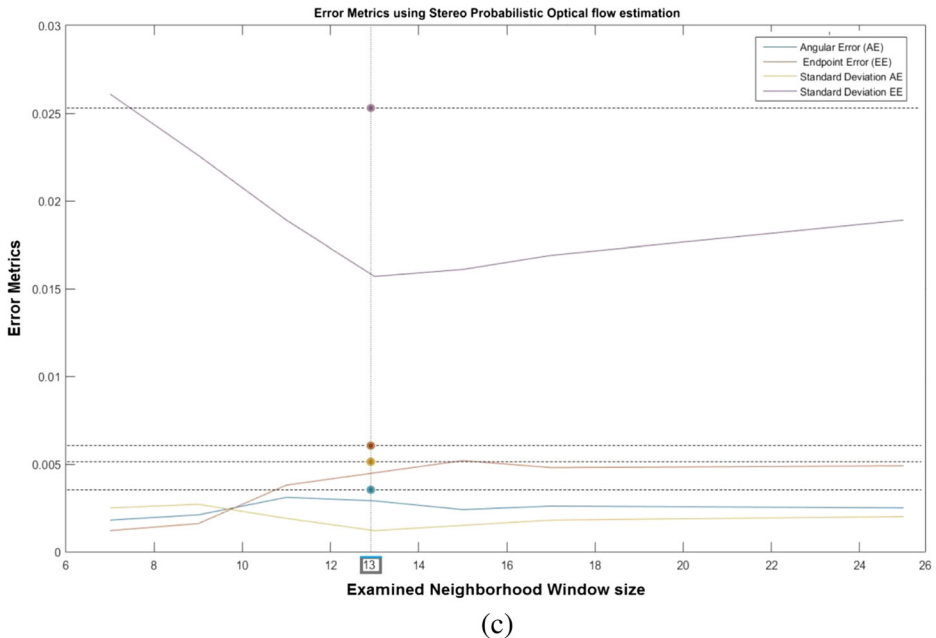


Fig. 11 (continued)

both the single-based as well as the stereo-based approaches. In subfigures 11.a and 11.b we present the error variance for different window sizes between the two approaches, whereas subfigure 11.c presents the fluctuations of the error metrics opposed to the optimal window size for the single camera approach (*solid points, window size 13×13*).

Commenting on the results, we can observe that the stereo-based approach outperforms the single camera-based method. The stereo based method achieves the same or better accuracy using smaller window sizes, thus reducing both the data amount as well as the computation time. Moreover, we can also observe that the standard deviation of the error metrics is also reduced despite the size increase of the interrogation region indicating that the produced flow field is more compound. As far as the estimation accuracy for the presented stereo-based probabilistic method, the Angular and Endpoint error metrics range below 1 degrees and 0.01 pixels, respectively, implying good estimation efficiency.

5.4 Velocity validation

The final step of the presented study, consists of the estimation validation which ensures the accuracy of the estimation. If natural particles, such as leaves, are present within the flow then they are used as validation points for the estimated velocity. The leaves are assumed to follow the river motion, thus having the similar velocities. Assuming an almost constant velocity we simply examine whether the leaf bunch enters an interrogation region, defined by two points or objects in the scene (indicators), after a specific number of frames since the frame that they were first identified. Following the methodology presented in Section 4 we have applied this validation scheme to our experimental scene in Koiliaris River (Fig. 12). The following

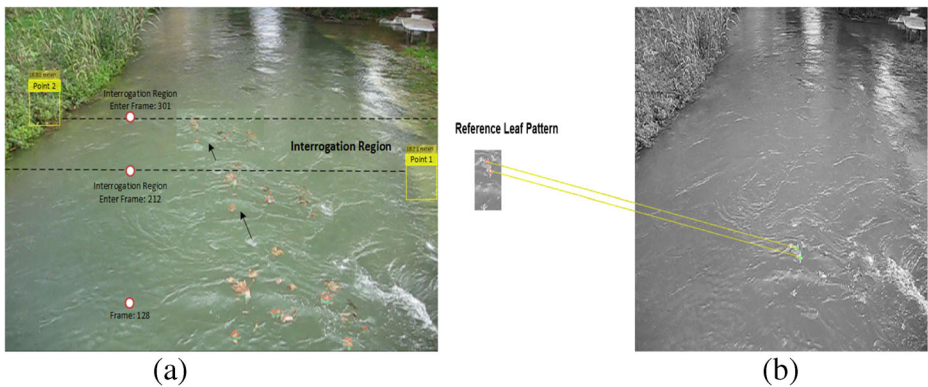


Fig. 12 Validation stage in Koiliaris River. Red and blue colors indicate the motion between consequent frames. (a) The leaf bunch formation across the frame series are fused into one unique frame to better represent the methodology stages across the frame series, (b) Identified and tracked leaf formation based on a reference leaf pattern taken from previous frames

validation concerns test case No 2 in Table 1. The red and blue color variations in the figure indicate the leaf motion between consequent frames and were added to highlight this motion. As indicator points/objects we used distinct rock and vegetation formations that were present in the scene (*any object can be selected*) forming an interrogation region of approximately 0.59 m in length. The first leaf encounter within the interrogation region was in frame 128 of our test video. The frame rate of the monitoring camera set was 30 fps. The estimated velocity according to our approach was 0.3864 m/s. The size of the interrogation region has been computed based on the selected identifiers (e.g. vegetation or rock formations as shown in Fig. 12a in points 1 and 2 respectively) using the methodology presented in subsection 4.1. Following our approach, we estimated that the leaf bunch will require, $t = 1.53$ s or 86 frames to first exit the interrogation region. As shown in Fig. 12, where the leaf bunch enters the interrogation region in frame 212, we estimate that it will start to exit (first leaf to exit) the interrogation region at frame 298, however we observe that this happens in frame 301, leading to a deviation of $f_{\text{real}} - f_{\text{expected}} = 301 - 298 = 3$ frames showing that our approach is close to the real. This result justifies our deviation in the estimated velocity field which in this test case was -0.0129 m/s.

Although this approach can be considered as an accurate validation step, however it depends on the existence of particles in the flow. This means that a constant velocity validation will not be possible since natural particles are not always present in the flow. A solution to this problem can be the incorporation of an automated artificial particle placement utility that will ensure the existence of validation particles during the frame acquisition period.

6 Conclusion and future work

In this article, a new image-based river monitoring framework is presented that allows the estimation of the real world average surface velocity of the river flow using image data. The proposed system shows high estimation accuracy, with a deviation of ± 0.01745 m/s from the estimate acquired using conventional equipment, it is autonomous and flexible allowing on-spot measurements at almost any river flow case. Furthermore, compared to previous work, the

system is free of the ground points requirement due to the novel use of a stereo camera layout that allows immediate mapping between the image plane and world coordinate systems. Moreover, for the estimation of the 2D motion field, we present a novel stereoscopic probabilistic optical flow method that utilizes the stereo data as a means of strengthening the displacement probability estimates of a pixel to a candidate position found by the use of Bayesian inference optical flow estimation scheme.

Future work will involve further assessment of the presented system in more river monitoring cases in order to further examine its accuracy and evaluate the role of its parameters in the deviation on the velocity estimate. Moreover, testing of monitoring under various weather conditions, will allow us to assess its performance based on the weather condition factor, since, so far, the system has only been tested under ideal weather conditions. Furthermore, more elaborate pre-processing and post-processing will be performed on the observed data in order to reduce the effect of noise or extreme illumination variations, due to outdoor monitoring conditions, on the estimated motion field. Techniques such as histogram equalization and motion vector filtering have already applied at the current version of the system, however, the methods used are basic and more problem-focused ones are expected to further increase the system's accuracy. The deviation on the validation stage will also be used as a means for readjusting the parameters of the system forming an autocorrecting loop, increasing the system's accuracy. Finally, the outmost goal will be to reach close-to real time computational execution, providing instant velocity estimates of the river's flow.

Acknowledgements This work is elaborated through an on-going THALES project (CYBERSENSORS – High Frequency Monitoring System for Integrated Water Resources Management of Rivers). The project has been co-financed by the European Union (European Social Fund - ESF) and Greek national funds through the Operational Program “Education and Lifelong Learning” of the National Strategic Reference Framework (NSRF) – Research Funding Program: Thales. Investing in knowledge society through the European Social fund.

References

1. Bacharidis K, Moirogiorgou K, Sibetheros I, Savakis A, Zervakis M, River Flow Estimation Using Video Data (2014) IEEE International Conference on Imaging Systems and Techniques (IST2014). Santorini Island, Greece, pp 173–178
2. Bay H, Ess A, Tuytelaars T, Van Gool L (2008) SURF: Speeded Up Robust Features. *Comput Vis Image Underst* 110(3):346–359
3. Bradley A, Kruger A, Meselhe EA, Muste M (2002) Flow measurement in streams using video imagery. *Water Resour Res* 38(12):1315
4. Chang J, Edwards D, Yizhou Y (2002) Statistical Estimation of Fluid Flow Fields. ECCV workshop on Statistical Methods in Video Processing, Copenhagen, pp 91–96
5. Chen X, Zillé P, Shao L, Corpetti T (2015) Optical flow for incompressible turbulence motion estimation. *Exp Fluids* 56:8. <https://doi.org/10.1007/s00348-014-1874-6>
6. Chorin AJ, Marsden JE (1990) *A mathematical introduction to fluid mechanics*. Springer-Verlag, New York **Texts in Applied Mathematics 4**
7. Creutin JD, Muste M, Bradley AA, Kim SC, Kruger A (2003) River gauging using PIV techniques: a proof of concept experiment on the Iowa river. *J Hydrol* 277:182–194
8. Fritz HM, Phillips DA, Okayasu A, Shimoazono T, Liu H, Mohammed F, Skanavis V, Synolakis CE, Takahashi TT (2011) Japan tsunami current velocity measurements from survivor videos at Kesennuma Bay using Lidar. *Geophys Res Lett* 2012:39

9. Fujita I, Tsubaki R (2004) Measurements of vorticity field from river flow images, vol V0032. Int. Conf. on Advanced Optical Diagnostics in Fluids Solids and Combustion, Tokyo
10. Fujita I, Muste M, Kruger A (1998) Large-scale particle image velocimetry for flow analysis in hydraulic applications. *J. Hydraul Res* 36(3):397–414
11. Fujita I, Watanabe H, Tsubaki R (2007) Development of a non-intrusive and efficient flow monitoring technique: the space-time image velocimetry (STIV). *Int J River Basin Manag* 5(2):105–114
12. Gui L, Wereley ST (2002) A correlation-based continuous window-shift technique to reduce the peak-locking effect in digital PIV image evaluation. *Exp Fluids* 32:50617
13. Hartley R, Zisserman A (2003) Multiple view geometry in computer vision. Cambridge University Press, New York
14. Has P, Herzet C, Mmin E, Heitz D, Mininni PD (2013) Bayesian estimation of turbulent motion. *IEEE Trans Pattern Anal Mach Intell* 35(6):1343–1356
15. Hauet A, Kruger A, Krajewski W, Bradley A, Muste M, Creutin J, Wilson M (2008) Experimental system for real-time discharge estimation using an image-based method. *J Hydrol Eng* 13(2):105–110
16. Heikkilä J, Silvén O (1997) A four-step camera calibration procedure with implicit image correction. In: Proceedings of the 1997 I.E. Computer Society Conference on Computer Vision and Pattern Recognition, 1997. IEEE, pp 1106–1112. <https://doi.org/10.1109/CVPR.1997.609468>.
17. Hoyt WG, Langbein WB (1939) Some general observations of physiographic and climatic influences on floods. *EOS Trans Am Geophys Union* 20:166–174
18. Huang HT, Fiedler HE, Wang JJ (1993) Limitation and improvement of PIV: part I. Limitation of conventional techniques due to deformation of particle image patterns. *Exp Fluids*
19. Jonkman SN (2005) Global perspectives on loss of human life caused by floods. *Nat Hazards* 34:151–175
20. Kääh A, Leprince S (2014) Motion detection using near-simultaneous satellite acquisitions. *Remote Sens Environ* 154:164–179
21. Kim Y (2006) Uncertainty analysis for non-intrusive measurement of river discharge using image velocimetry. Ph.D. dissertation, Univ. of Iowa, Iowa City
22. Le Boursicaud R, Pénard L, Hauet A, Thollet F, Le Coz J (2016) Gauging extreme floods on YouTube: application of LSPIV to home movies for the post-event determination of stream discharges. *Hydrol Process* 30:90–105
23. Lecordier B, Demare D, Vervisch LM, Riveillon J, Trinité M (2001) Estimation of the accuracy of PIV treatments for turbulent flow studies by direct numerical simulation of multi-phase flow. *Meas Sci Technol* 12:1382–1391
24. Nakajima Y, Inomata H, Nogawa H, Sato Y, Tamura S, Okazaki K, Torii S (2003) Physics-based flow estimation of fluids. *Pattern Recogn* 36:1203–1212
25. Ralli J (2011) PhD thesis, Fusion and regularisation of image information in variational correspondence methods, Universidad de Granada. Departamento de Arquitectura y Tecnología de Computadores. <http://hera.ugr.es/tesisugr/20702371.pdf>
26. Saikano H (2008) Fluid motion estimation method based on physical properties of waves. *Comp. Vision and Pat. Recognition (CVPR)*
27. Saikano H (2015) Motion estimation for dynamic texture videos based on locally and globally varying models. *IEEE Trans Image Process* 24(11):3609–3623. <https://doi.org/10.1109/TIP.2015.2447738>
28. Baker Simon, Scharstein Daniel, J. P. Lewis, Stefan Roth, Michael J. Black, Richard Szeliski (2007) A database and evaluation methodology for optical flow. *Int Conf Comput Vis* 1–8. <https://doi.org/10.1109/ICCV.2007.4408903>
29. Trucco E, Verri A (1998) Introductory techniques for 3-D computer vision. Prentice Hall, Upper Saddle River. http://cseweb.ucsd.edu/classes/fa12/cse252A-a/hw4/trucco_verri_pp178-194.pdf
30. Tsubaki R (2011) I Fujita, S Tsutsumi, Measurement of the flood discharge of a small-sized river using an existing digital video recording system. *J Hydro Environ Res* 5:313–321
31. Vescoukis V, Doulamis N, Karagiorgou S (2012) A service oriented architecture for decision support systems in environmental crisis management. *Futur Gener Comput Syst* 28(3):593–604
32. Westerweel J, Dabiri D, Gharib M (1997) The effect of a discrete window offset on the accuracy of cross-correlation analysis of digital PIV recordings. *Exp Fluids* 23:20–28 **Springer-Verlag**
33. Wildes RP, Amabile MJ, Lanzillotto A, Leu TS (2000) Recovering estimates of fluid flow from image sequence data. *Comput Vis Image Underst* 80:246–266
34. Zhang Z (1999) Flexible camera calibration by viewing a plane from unknown orientations. In: The Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999 ((Vol. 1, pp 666–673). IEEE. <https://doi.org/10.1109/ICCV.1999.791289>



Konstantinos Bacharidis was born in Greece in 1989. He received the B.Sc., M.Sc. from the School of Electrical and Computer Engineering (ECE) in Technical University of Crete (TUC) in 2014, and 2016, respectively. His main areas of research interest are in computer vision, machine learning and integrated imaging systems concerning the aspects of motion estimation, object detection and tracking, pattern recognition and stereoscopic vision.



Konstantia Moirogiorgou completed her BSc and MSc in Electronic & Computer Engineering at the Technical University of Crete (TUC), Greece. Currently, she is a PhD candidate and a research and teaching assistant in the Digital Signal & Image Processing Laboratory (Display Lab) at the School of Electrical & Computer Engineering (ECE) in TUC. Her research interests mainly focus on Digital Signal & Image Processing, real-time video processing systems, DSP-based imaging systems, video transmission over network. She has been involved in many EU funded R&D Projects, while she has co-authored many papers mainly in the above research areas.



Georgia Koukiou was born in Patras. She received the B.Sc. degree in Physics in 2004 and her M.Sc. in Electronics and Computers in 2006, both from the University of Patras. Her master thesis dealt with Current Feedback Operational Amplifiers. She is member of the research team of Digital Image and Signal Processing in the University of Patras. Her research interests are in Identification of Psychosomatic Conditions Using Infrared Facial characteristics. Emphasis is given in recognizing drunk persons using Thermal imagery.



George Giakos is a Chairman and Professor, Electrical and Computer Engineering, Director of Graduate Program, Manhattan College. Dr. Giakos research is articulated in the design of imaging systems, ladars and surveillance sensor platforms, multispectral polarimetry, nuclear instrumentation, design of cargo and airport security inspection systems for the Department of Defense and Homeland Security, as well as in the design of innovative molecular nanophotonics devices. His research group was among the first in the US to pioneer the characterization of the detection and imaging characteristics of Cadmium Zinc Telluride semiconductor substrates for flat-panel digital radiography applications. He has fostered several breakthrough inventions which have been rewarded with fifteen (15) US and international Patent Awards and more than 150 peer-review articles and journal publications. He is the recipient of a Distinguished Faculty Fellow Award, from the Office of Naval Research. He received numerous prestigious research faculty fellowship awards from the Department of the Air Force, NASA, National Academy of Sciences, and Naval Research Laboratory. Dr. Giakos is an IEEE Fellow, the Editor in Chief of the International Journal of Signal and Imaging Systems Engineering (IJSISE), the Chairman of the TC-19 IEEE Technical Committee on “Imaging Measurements”, for the Instrumentation and Measurements Society, a Guest Editor of the Special Issue on Imaging Systems and Techniques in the IEEE Transactions on Instrumentation and Measurement, 1996, a Guest Editor of the Special Issue on Imaging Systems and Techniques in the IEEE Transactions on Instrumentation and Measurement, 2006, a Guest Editor for the Special Issue on Imaging Systems and Techniques in the Measurement Science and Technology Journal, 2008, a Guest Editor of the Special Issue on Imaging System and Techniques, IET Image Processing Journal, IOP, 2009. He is a member of the IEEE Instrumentation and Measurement Society, IEEE Nuclear and Plasma Sciences Society, IEEE Photonics Society, IEEE Nanotechnology Forum, and SPIE. He maintains active

collaborations with the Department of Navy, Department of the Air Force, Lockheed Martin, NASA, NIH, and Cleveland Clinic. He is the founder, organizer and general chairman of the successful IEEE IM International Workshops and Conferences on Imaging Systems and Techniques.



Michalis Zervakis is serving as Professor at the School of Electrical and Computer Engineering and directs the Digital Image and Signal Processing Laboratory (DISPLAY) at the Technical University of Crete. His research interests include modern aspects of signal and image processing, data mining and pattern recognition, imaging systems and integrated automation systems. Application areas include surveillance and environmental applications, bioinformatics, biosignal analysis and medical Imaging. He has published over 100 scientific papers in these areas and he is involved on many national and European projects.