1Research on Intelligent Perception Technology of Dynamic Flood Conditions Based on Video Images

Ke Zhang, Yun Li, Suchuang Di, Yongkun Li, Junxiong Huang, Xuli Zan, Pengyu Liu, Jiali Chen, Bowen Yuan, Xuemin Li, Yuguang Lu, and Ying Lu, China

Key words: Video Image; Surface Flow Velocity; Optical Flow; Flood Monitoring

1. SUMMARY

In catastrophic rainstorm and flood scenarios, monitoring equipment often fails, making it difficult to obtain timely flood data. This study utilizes widely available and information-rich video footage, combined with the RAFT deep learning optical flow model, to develop an intelligent perception system for estimating river surface flow velocity from images, thereby enhancing flood emergency decision-making capabilities.

The RAFT-based model analyzes pixel intensity changes between consecutive video frames to derive a dense optical flow field, enabling tracer-free surface velocity measurement. Tests conducted in both indoor simulated channels and at an outdoor site on Beijing's Qinghe River showed stable frame-by-frame recognition with an average velocity error of less than 0.1 m/s.

This approach addresses the challenge of acquiring hydrological data when conventional monitoring systems fail, providing non-contact, pixel-level surface flow estimation. It significantly improves pervasive flood monitoring and supports more effective disaster assessment and early warning.

Research on Intelligent Perception Technology of Dynamic Flood Conditions Based on Video Images

Ke Zhang, Yun Li, Suchuang Di, Yongkun Li, Junxiong Huang, Xuli Zan, Pengyu Liu, Jiali Chen, Bowen Yuan, Xuemin Li, Yuguang Lu, and Ying Lu, China

2. Intelligent Recognition Model for Surface Flow Velocity Without Tracers

2.1 Dataset Construction

Intelligent Perception of Surface Flow Velocity Without Tracers is constructed based on the deep learning optical flow method (Optical Flow Visualization, OFV). The OFV dataset adopts a data construction method that automatically generates optical flow ground truth from real-world videos[1-2]. This method extracts and matches objects from adjacent video frames to compute dense optical flow values, and then applies deformation to the objects of interest using these values to produce synthetically simulated images. The original video frames and the synthetically simulated images form an image pair as input to the model, while the dense optical flow values serve as the ground truth for training. The construction process is illustrated in the figure below. Based on real videos of converging water flow captured at the experimental site and surveillance footage from the Shaziying Gate on the Qinghe River, the aforementioned algorithm was used to generate computer-simulated image pairs along with their corresponding optical flow annotations, resulting in a final dataset of 900 simulated images and their optical flow annotation files.

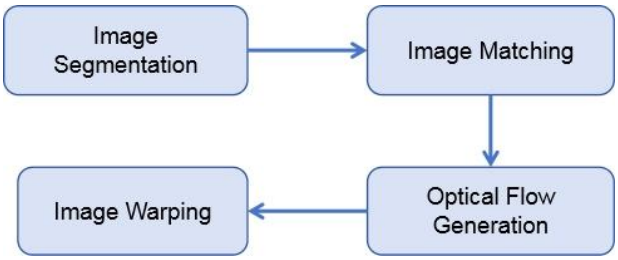

Fig.1 Dataset Construction Workflow for the OFV Method

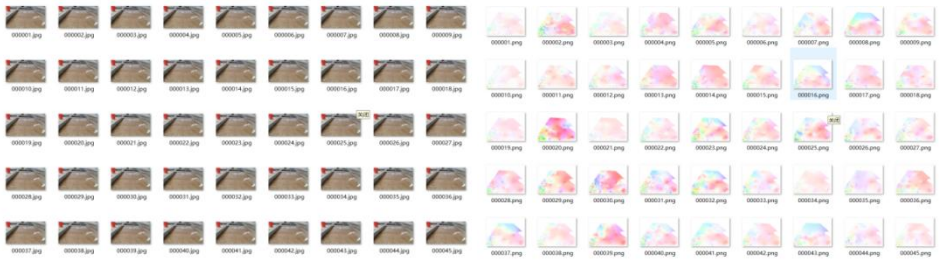

Fig.2 Raw Images with Optical Flow Labels

2.2 Model Development

To address the problem of flow velocity estimation without tracers and with full water-surface coverage, a dense optical flow-based velocity measurement method built on the

RAFT (Recurrent All-Pairs Field Transforms) optical flow network is proposed, which achieves global flow-field measurement through pixel-wise motion estimation[3-4].

The optical flow velocimetry algorithm based on the RAFT model is employed for water-surface velocity detection[5-7]. It computes the motion of corresponding pixels between consecutive frames to obtain a dense optical flow field, thereby realizing global velocity measurement. RAFT is a deep neural network that follows the fundamental principle of optical flow—estimating the motion of corresponding pixels in an image sequence—and represents a deep-learning-based implementation of optical flow. It consists of several key modules, including feature extraction, correlation volume construction, and iterative refinement, with its structure illustrated in the figure below.

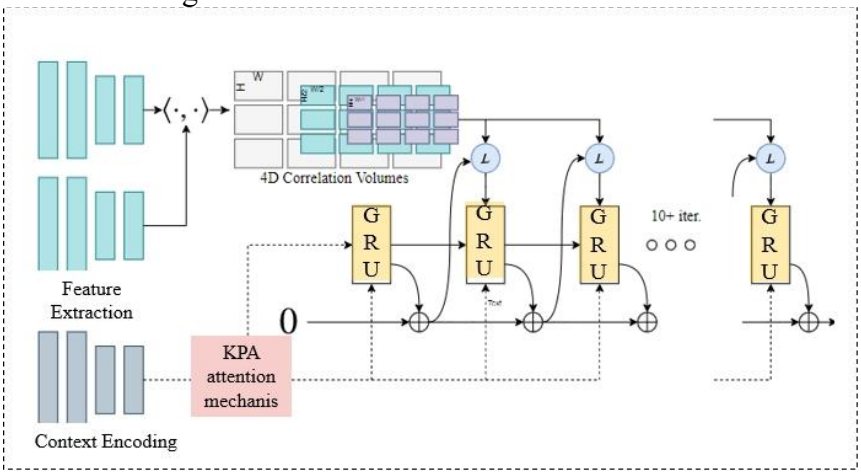

Fig.3 RAFT Architecture Diagram

Its working principle is as follows: The spatial feature maps of two adjacent frames are obtained separately through a feature extraction CNN network, the correlation matrix between the feature maps is calculated, and a 4D correlation volume pyramid at different scales is constructed. The recurrent update module utilizes the feature extraction CNN network to extract the context feature map of the current frame. This context feature map, along with the correlation query results from the 4D correlation volume pyramid and the initialized optical flow field, is then fed into the GRU-based RNN module to derive pixel motion information with temporal characteristics. The network is described in detail below.

Feature Extraction Module: The structure used is illustrated in the figure below. RAFT employs a CNN network composed of six residual blocks to extract both spatial feature maps and context feature maps. After feature extraction, the image size is downsampled to 1/8 of its original resolution. The input image first passes through a Conv layer containing convolution, batch normalization, and activation functions. This layer uses a 7×7 kernel to expand the network's receptive field. Three residual layers with different numbers of channels sequentially downsample the input feature map by a factor of two, ultimately achieving 8× downsampling, followed by a Conv layer operation with a 3×3 kernel.The context feature extraction network and the spatial feature extraction network share an identical structure. They are distinguished by their input: both the previous and current frames are processed by the spatial feature extraction network, while only the previous frame is fed into the context feature extraction network.

Computing Visual Similarity: In the feature extraction network, we obtain the 8× downsampled feature maps of the two consecutive frames. A 4D correlation volume is constructed by computing the similarity between these two feature maps. Given the image features $g_0(I_1) \in \mathbb{R}^{H\times W\times D}$ and $g_0(I_2) \in \mathbb{R}^{H\times W\times D}$ correlation volume is formed by taking the dot product between all pairs of feature vectors.

The lookup operator LC is employed to generate feature maps of "query correlation" by indexing into the correlation pyramid. Given the current estimate of optical flow (f_1, f_2), each pixel $x = (u, v)$ in I_2 is mapped to its estimated correspondence

$$x' = (u + f_1(u), v + f_2(v))$$

A local grid $N(x')_r$ is then defined around x' as follows:

$$N(x')_r = \{x' + d_x | d_x \in \mathbb{Z}^2, ||d_x||_1 \leq r\}$$

The correlation lookup is performed by querying and sampling from the correlation pyramid using the local neighborhood $N(x')_r$. Since $N(x')_r$ is a real-valued grid, the bilinear sampling method is employed.The lookup is executed across all levels of the pyramid, where the correlation volume C_k at level k is indexed using the grid $N(x'/2^k)_r$. A constant radius across levels implies larger contextual information at lower levels. The values from each level are then concatenated into a single feature map.

Iterative Update Module: The iterative update operation starts from an initial flow field $f_0 = 0$ and sequentially estimates a series of flow fields $\{f_1 \ldots f_n\}$. At each iteration, the module produces a refinement update ΔfΔf applied to the current estimate:

$$\Delta f: f_{k+1} = \Delta f + f_k + 1$$

The iterative update operation takes the context features, the queried correlation results, and the current hidden state as inputs, and outputs both the update ΔfΔf and the updated hidden state. The structure of the iterative update module is designed to mimic the steps of an optimization algorithm. Its core component is a gated activation unit based on the GRU (Gated Recurrent Unit), where fully connected layers are replaced with convolutional layers. The structure of the GRU is illustrated in the figure below. Through the iterative update module, a global flow estimation at 1/8 of the original image resolution is produced.

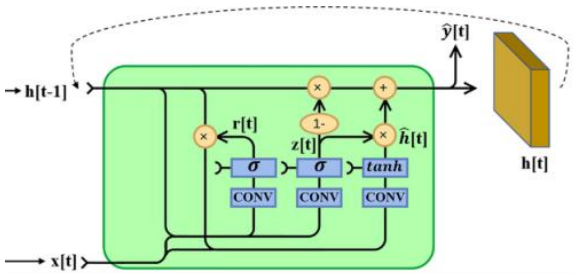

Fig.4 GRU Structure

Upsampling: As described above, the network outputs an optical flow at 1/8 resolution. The flow is then upsampled to full resolution by converting the full-resolution flow at each pixel into a convex combination of a 3×3 grid of its coarse-resolution neighbors. The final high-resolution flow field is obtained by permuting and reshaping into an H×W×2 dimensional flow field. This layer can be implemented directly in PyTorch using the unfold function.

Loss Function: The L1 loss function is used to calculate the L1 distance between the optical flow predicted by the network and the ground truth flow, thereby supervising the training and optimization of the network. Let the sequence of flows predicted by the iterative update module be $\{f_1 \ldots f_n\}$, with γ as the corresponding loss weight for each prediction. Given the ground truth flow f_{gt}, the loss function is defined as:

$$\mathcal{L} = \sum_{i=1}^{N} \gamma^{N-i} \left\| f_{gt} - f_i \right\|_1$$

2.3 Model Application

The figure below shows an example of the detection results from the RAFT-based OFV water surface velocity estimation model. Multiple flow velocity scenarios were simulated, and the actual surface velocities were measured to validate the model's performance. As can be seen from the figure, the average flow velocity over 30 video frames is 0.625 m/s, which is close to the reference velocity of 0.585 m/s obtained by radar velocimetry. The average velocity errors for other test videos are also small, resulting in an overall mean error of only 0.04 m/s across all test videos, with an average deviation rate of 7.57% and an accuracy of 92.43%. Although the accuracy is slightly lower compared to the PIV method, the detection process does not rely on tracer particles at all, enables full-field velocity measurement over the entire water surface, shows smaller fluctuations in results, and provides more stable and reliable detection outcomes.

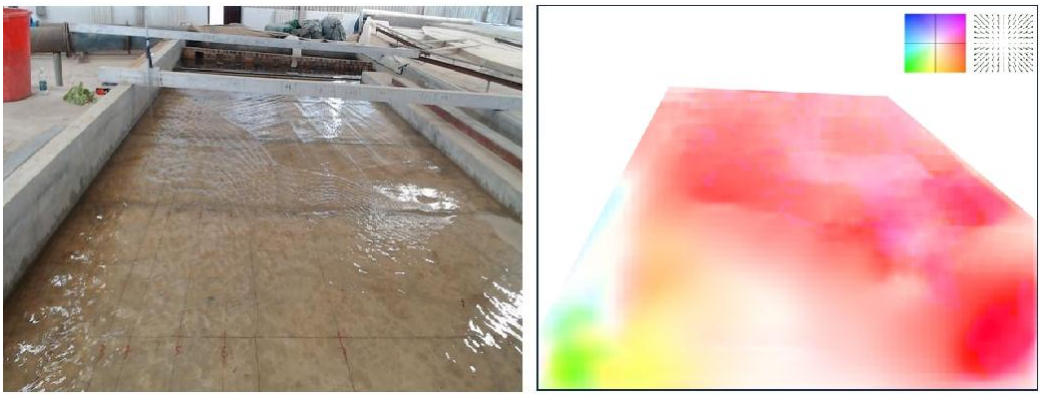

Fig.5 Detection Results

Tab.1 Comparison of Flow Velocity Test Results

	Model-Estimated Average Flow Velocity	Measured Average Flow Velocity	Deviation Rate
Scenario 1	0.495m/s	0.491m/s	0.82%
Scenario 2	0.420m/s	0.525m/s	20%
Scenario 3	0.625m/s	0.585m/s	6.84%
Scenario 4	0.451m/s	0.500m/s	9.8%
Scenario 5	0.538m/s	0.540m/s	0.37%

3. Research Conclusions and Significance

This study proposes an intelligent perception technology for river surface flow velocity based on video images, incorporating the RAFT model. Tests conducted on both indoor simulated channels and outdoor natural river channels demonstrate that the average velocity recognition error is less than 0.1 m/s, achieving non-tracer-based surface flow velocity estimation. The model results can effectively address the issue of acquiring hydrological data when conventional monitoring equipment fails, thereby enhancing the pervasive sensing capability of flood information.

REFERENCES

[1] Revaud J , Weinzaepfel P , Harchaoui Z ,et al.DeepMatching: Hierarchical Deformable Dense Matching[J].International Journal of Computer Vision, 2016, 120(3):1-24.

[2] Sorkine O , Alexa M .As-Rigid-As-Possible Surface Modeling[J].Eurographics Symposium on Geometry Processing, 2007.

[3]Chao Wang;Xiucheng Dong;Shifu Gu;Zhengyu Zhang;Hongjiang Qian.Global velocity measurement of fluorescent oil film based on deep learning optical flow method[J].Journal of Aerospace Power,2022,37(07):1539-1549.

[4]Tiantian Du;Xiaolong Wang;Jin He.Optical-flow-based Waterway Velocity Detection Algorithm Under Complex Illumination Conditions[J].Computer Engineering, 2024, 50(04):60-67.

[5]Teed, Z., & Deng, J. (2020). RAFT: Recurrent All-Pairs Field Transforms for Optical Flow. In A. Vedaldi, H. Bischof, T. Brox, & J.-M. Frahm (Eds.), Computer Vision – ECCV 2020 - 16th European Conference, 2020, Proceedings (pp. 402-419). (Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics); Vol. 12347 LNCS). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-3-030-58536-5_24

[6] Brox T , Bregler C , Malik J .Large displacement optical flow[J].Proceedings / CVPR, IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2009:41-48.

[7] Kroeger T , Timofte R , Dai D ,et al.Fast Optical Flow using Dense Inverse Search[C]//European Conference on Computer Vision.Springer International Publishing, 2016.

BIOGRAPHICAL NOTES

Ke Zhang, Ph.D. in Science, is a Senior Engineer at Beijing Water Science and Technology Institute. She primarily engages in video image-based recognition of river and lake hydrological conditions and intelligent supervision of aquatic ecological spaces.

CONTACTS

Dr. Ke Zhang
Beijing Water Science and Technology Institute
No.21,ChegongzhuangWest Road, Haidian District
Beijing
China
Email:zke@bwsti.com