

流量観測の無人化・省力化に向けた非接触型流速計設置の考え方と 既存 CCTV を活用した高度観測手法の提案

株式会社 建設技術研究所 ○ 芥田 直輝
株式会社 建設技術研究所 大屋 敬之
株式会社 建設技術研究所 能登谷祐一
株式会社 建設技術研究所 由良 英作
株式会社 建設技術研究所 守谷 将史

論文要旨

国土交通省は流量観測の無人化・省力化に向けた取り組みを進めている。本検討では、その取り組みの一環として、非接触型流速計(電波式・画像式)の特徴や、観測地点ごとの条件を踏まえた手法の選定方法および設置にあたっての留意点を検討した。また、既存の CCTV カメラを活用する観点から、AI 技術を用いた画像解析による水位観測手法、STIV による流速・流量観測手法を検討した。水位観測においては転移学習や Trimap 法などを適用した精度向上策を検討した。流速・流量観測においては STIV に使用する動画像のフィルタリングによる精度向上策を検討した。さらに、その適用性を確認するとともに、他地点への展開に関する留意点を確認した。

キーワード：流量観測，非接触型流速計，CCTV，画像解析，STIV，AI

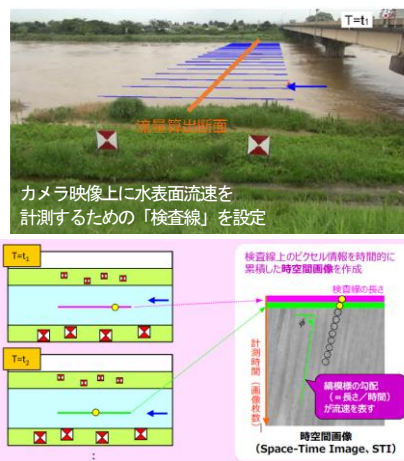
まえがき

これまで、日本国内における高水流量観測は浮子観測により実施されている。浮子観測は観測原理が単純であるとともに、急峻で流速が大きく、ごみや流木等が多い日本の河川においても流速を確実に計測することができる。しかし、洪水規模や観測所によっては、「大規模出水時に観測所へ到達できない」、「洪水の立ち上がりが速く観測が間に合わない」、「観測員の安全が確保できない」、「観測体制の構築が困難」、「所定の精度が確保できない」等の課題がある。

これに対し国土交通省では、民間企業が持つ技術を活用して、河川行政における技術課題や政策課題を解決するための機器やシステム等を開発する「革新的河川技術プロジェクト第4弾(令和2~3年度)¹⁾」として、無人化・省力化に向けた流量観測機器の開発を行っている。この成果をふまえて、国土交通省は全国の約100か所において非接触型流速計の設置を予定しており、精度が確認された地点から順次、浮子観測を廃止し、非接触型流速計による観測に移行していくことを想定している。また、高水流量観測の無人化・省力化に向けては、その有力な手法である画像式流速計(図-1)と電波式流速計(図-2)を対象として、「非接触型流速計測法の手引き(案)²⁾」(以下、手引き案)が令和5年3月に策定されている。

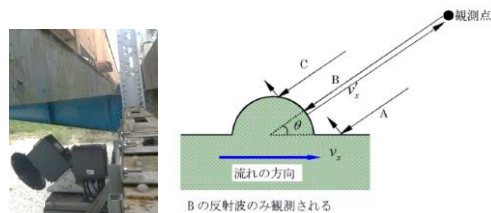
本検討では、国土交通省が管理する一級河川における非接触型流速計設置予定地点を対象として具体的な設置方法を検討した事例をもとに、手法(電波式 or 画像式)の選定や、機器配置等の考え方をとりまとめた。また、今後の非接触型流速計の設置に

あたって、既存施設の活用の観点から、既存 CCTV カメラ画像を用い、「水位を観測する手法」や「表面流速を観測する手法(STIV³⁾)」の適用性を検討した。



出典：非接触型流速計測法の手引き(案)

図-1 画像式流速計(STIV)の計測原理²⁾



出典：非接触型流速計測法の手引き(案)

図-2 電波式流速計とその計測原理²⁾

1. 非接触型流速計の手法選定・配置検討の考え方

手引き案²⁾では、「革新的河川技術プロジェクト第4弾¹⁾」などの最新の検討成果もふまえて、電波式および画像式流速計による観測手法の特性、観測手法の選定から、観測方法、観測時の留意点、データの記録、データの保管方法、観測機器の点検、観測データのチェックなど一連の仕様が示され、継続した観測、水文観測データとしての蓄積が可能なようにとりまとめられている(表-1)。

非接触型流速計の具体的な設置方法を検討した事例では、これらを参考にして検討が進められているが、手引き案²⁾に記載のない幾つかの留意すべき事項がみられたため、本検討ではこれらの事項を中心にとりまとめる。

(1) 電波式流速計のセンサー設置基数

電波式流速計は、センサーから電波が照射される狭い範囲(左右幅3~5m程度、長さ5~10m程度)の流速を測定する手法のため、浮子観測と同様にセンサーを横断方向に複数設置することが基本となる。手引き案²⁾において、センサーの設置基数は「経済性を考慮し、原則、やむを得ず流量観測を緊急かつ迅速に行わなければならない場合の最小浮子流速観測線数(表-2)と同様とする。」とされている。

しかし、実務においては、目標とする設置地点数があり、コスト面から1地点あたりに設置するセンサーの基数は、精度を確保するために必要な最低限としたいとの要望がある。このため、本検討では、センサー基数・設置位置を現状の浮子観測線と同様とした基本ケースと、基本ケースから測線を間引いたケースを比較して、誤差が一定範囲に収まる最小の測線数を設定した。

表-1 手法の選定・機器配置の考え方

項目	手法の選定・機器配置の考え方	
	電波式流速計	画像式流速計
設置位置	<ul style="list-style-type: none"> 流下方向に向けたセンサーを横断方向に複数設置(設置基数の検討) 橋梁への設置が基本 対水面間距離20m以内 	<ul style="list-style-type: none"> 流向に対して直交方向を撮影する画角(推奨) カメラ1台の有効距離150m(目安)(有効距離の考え方について整理)
水面の見通し	<ul style="list-style-type: none"> 電波照射範囲内に護岸、岩、河床、植生等がない 	<ul style="list-style-type: none"> 水面との間に植生等の障害物がない
流れ	<ul style="list-style-type: none"> 渦を巻くような流れや偏流がない(三次元流況解析により河道形状や横流入等の影響を確認) 	
水面	<ul style="list-style-type: none"> 鏡面状態にならない(電波が反射しない(電波式)、波紋の移動速度が読み取れない(画像式)ため) 	
濃霧等	-(影響なし)	<ul style="list-style-type: none"> 水面が視認不可にならない
その他	-(特になし)	<ul style="list-style-type: none"> 住宅がある場合プライバシーに配慮してカメラ画角を設定 光源(ライト)を設置する場合は生物等への影響に配慮

※上表は手引き案²⁾に記載されている「非接触型流速計測の検討、観測手法選定の流れ(案)」を参考に手法の選定・配置の考え方を整理したもの
赤字：手引き案²⁾記載内容以外の検討事項(赤字の項目については以降で詳述する)

表-2 やむを得ず流量観測を緊急かつ迅速に行わなければならない場合の水面幅と最小浮子流速観測線数の関係⁵⁾

水面幅	50m未満	50~100m未満	100~200m未満	200~400m未満	400~700m未満	700m以上
浮子流速観測線数	3	4	5	6	7	8

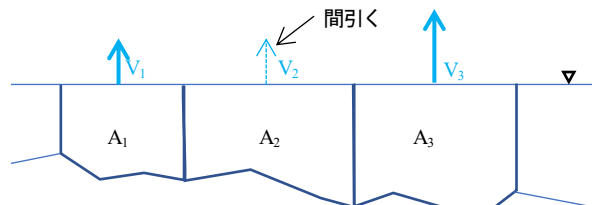
出典：河川砂防技術基準 調査編

以降では、基本ケース(全測線数)が9測線のA地点の例を示す。平成16年度から令和4年度までの計19年間の高水流量観測における浮子観測の記録(全31洪水)を整理し、検討した。間引く測線は、算出した流量(図-3)を基本ケースの流量と比較して最大誤差率が最小となる測線とした。

「浮子観測データとの整合性の確認の実施方針(案)⁴⁾」によれば、非接触型流速計による流量が浮子観測流量と比較して、「相対差-20%から0%の範囲に収まっている場合には、浮子観測データとの整合性が保たれていると判断」とされている。測線数の設定においては、測線を間引くことによる誤差に観測手法自体による誤差を加えた誤差が、-20~0%に収まる必要があることから、測線数による誤差は余裕を持たせて±5%を許容範囲とした。A地点においては6測線で最大誤差率が5%以下となることから、6測線を採用した(表-3、図-4)。A地点における機器配置の例を図-5に示す。

$$\text{通常の流量} : Q = V_1 \times A_1 + V_2 \times A_2 + V_3 \times A_3$$

$$\text{間引いた流量} : Q' = V_1 \times (A_1 + 0.5 \times A_2) + V_3 \times (A_3 + 0.5 \times A_2)$$



※Q, Q' : 流量, V : 各区分断面における流速, A : 河道区分断面
※間引いた区分の面積を両側に案分して算定
※中央部分の測線を間引いた場合を想定

図-3 測線を間引いた際の流量算定方法

表-3 各測線数の最適ケースと最大誤差率の例(A地点)

測線番号	1	2	3	4	5	6	7	8	9	最大誤差率(%)
(全測線観測)	●	●	●	●	●	●	●	●	●	0
8測線観測	●	●	●	●	●	●	●	●	●	0
7測線観測	●	●	●	●	●	●	●	●	●	3
6測線観測	●	●	●	●	●	●	●	●	●	5
5測線観測	●	●	●	●	●	●	●	●	●	8
4測線観測	●	●	●	●	●	●	●	●	●	21
3測線観測	●	●	●	●	●	●	●	●	●	41

※上表は各測線数の最適ケースと最大誤差率を整理したもの。
※最大誤差率が5%以下となった6測線観測(橙ハッチ)を本検討では採用。

● : 採用した測線番号

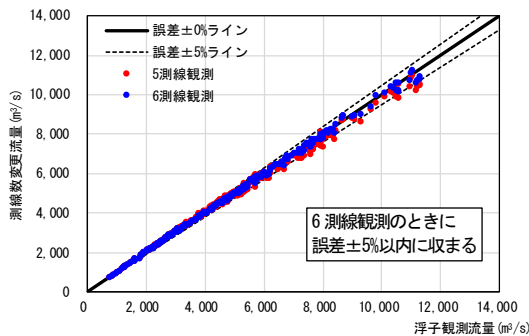


図-4 測線数に応じた流量精度のばらつき(例(A地点))

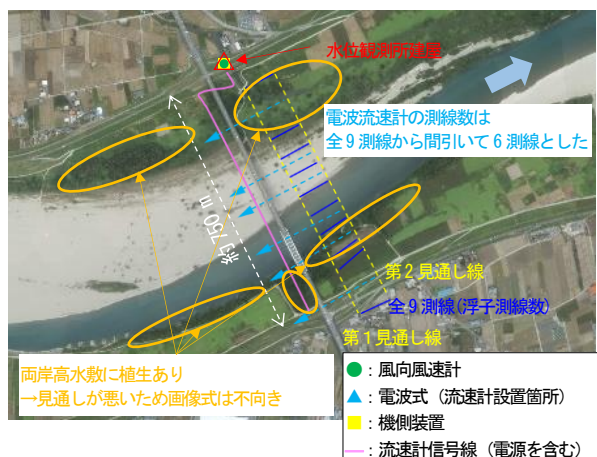


図-5 川幅が広い地点の手法選定・機器配置の例(A地点)

(2) 画像式流速計の観測可能範囲

画像式流速計はカメラで撮影した動画中の水面の波紋の移動速度を解析する手法であり、カメラから離れるほど波紋が不明瞭となるため精度が低下する。画像式流速計の観測可能範囲については藤田ら³⁾により、観測精度を確保するための撮影条件の目安として「解析範囲の1ピクセルが0.4m以下、俯角2°以上」が示されているほか、「革新的河川技術プロジェクト第4弾¹⁾」では、最大280mまで観測可能であることが確認されている。これらを参考に、手引き案²⁾ではカメラ1台の有効距離の目安を150m(両岸設置で300m)としているが、革新的河川技術プロジェクト¹⁾の事例は晴天時で見通しが良い好条件下の観測結果である。夜間や豪雨時等の悪条件下においても、高感度カメラや遠赤外線カメラを用いることで、精度よく観測が可能であることが確認されているが、現時点では、悪条件下での観測可能範囲が明確に示されておらず、距離によっては精度が低下する可能性も考えられる。

このため、手法の選定・機器配置の考え方(表-1)をもとに、地点特性から電波式流速計と画像式流速計の両方が適用可能と判断され、費用面についても同等となる場合は、距離による精度低下が生じない電波式流速計を選定することとした。

また、川幅が広く画像式流速計のみでは観測ができない地点については、今回は両岸高水敷に植生があり見通しが悪い、コスト面でメリットがない等の理由から、電波式流速計を選定したが、地点によっては河道の中央部を電波式流速計、両岸を画像式流速計で観測するなど、両者を複合した観測体制とすることも考えられる。

(3) 流況の確認

電波式流速計の設置位置について手引き案²⁾では、「河道が湾曲していない地点で、かつ大きく流速の異なる水域がなく、渦を巻くような流れにならない場所であること」、「橋脚の位置及び橋脚により流れに変化が及ぶ位置でないこと」、「電波の照射位置の河床に旧橋の基礎等が残る等、河床が極端に凹凸している場所でないこと」とされている。

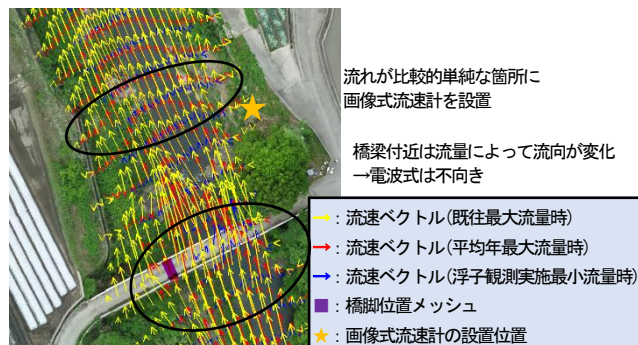


図-6 流況解析の例(B地点(河道形状が複雑))

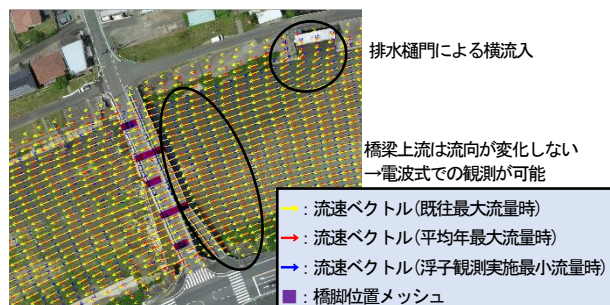


図-7 流況解析の例(C地点(上流に横流入))

また、画像式流速計については、「渦を巻くような流れや偏流がない位置」、「横断面の見通しがよく、流況に影響を与える植栽等の障害物がない位置」、「橋脚の影響を受けない位置」とされており、両者とも単純な流れとなる位置で観測することが求められているが、地点によっては適切な観測位置を判断することが難しい場合もある。

このため本検討では、河道形状や橋脚等の構造物、支川等の横流入等による流況への影響が想定される地点については、三次元流況解析により観測位置の妥当性を定量的に評価したうえで、精度を確保できる機器の配置を検討した。

河道形状が複雑なB地点については、電波流速計の設置が想定される橋梁付近の流向が、事前に想定されたとおりの流量規模によって複雑に変化することが確認されたため、設置位置の自由度が高い画像式流速計を適用することとし、計器を流れが比較的単純な箇所配置した(図-6)。

C地点については、上流側に排水樋門があり、洪水時には横流入により流向が変化することが想定された。しかし、洪水時における本川流量に対する樋門流出量の流量比は10%程度であり、流況解析の結果、橋梁の上流側においてもその影響が及ばないことが確認できた。このため、橋脚の影響がない橋梁上流側に電波流速計配置することとした(図-7)。

(4) その他(住宅、生物への影響)

画像式流速計を使用する際には、周辺住民のプライバシーや光源の設置による生物への影響に配慮する必要がある。本検討では、観測所周辺における住宅の存在状況や、河川環境情報図をもとに光源設置によって影響が考えられる生物の存在有無を確認した(図-8)。

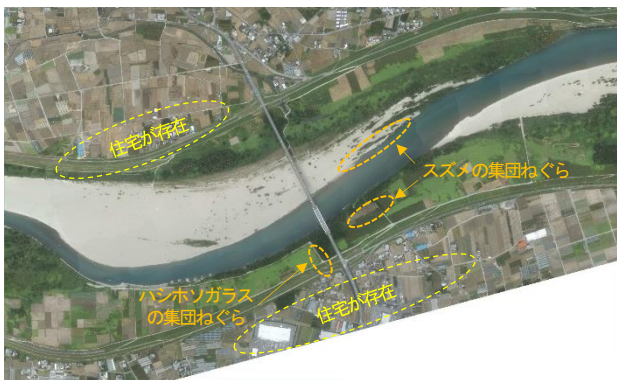


図-8 住宅や生物への配慮が必要な地点の例(A地点)

2. 既存 CCTV を活用した水位・流速算定手法の検討

既存の CCTV を活用した活用した高度観測として、CCTV 画像を 10 分毎にリアルタイム取得し、水位および流速を自動算定する水位・流速算定システムの運用を想定し、以下の検討を実施した。

(1) CCTV 画像の適用方法の検討

本検討の対象地点は、淀川水系桂川の渡月橋左岸上流部である。対象とした CCTV は、一級河川の河川監視用カメラとして一般的な性能のものである(表-4)。この CCTV は、設置箇所の下流左岸のバラペット区間の監視を目的としており、水位・流速算定には不向きな画像であった。そこで、以下を考慮して CCTV の画角を設定した。

- ・水位判読の目印となる構造物があること(仮想量水標を想定)
- ・流速観測を行う場合、左右岸全体を見通せる必要があること
- ・河川の流れに影響を及ぼす障害物が少なく、水面が確認しやすいこと

なお、設定した画角は、CCTV 管理システムのプリセット機能を用いて、観測時に呼び出せるようにした。図-9に対象 CCTV の観測画角を示す。

CCTV の観測画角内に、水位観測で必要となる仮想量水標を想定した。また、現場で簡易測量を実施し、流速観測で必要となる標定点を設定した。

(2) 画像解析による水位算定手法の検討

ここでは、水面検知を用いた水位算定手法を試みた。水面検知に関する画像解析として、水面と似た輝度を持つ他の構造物の誤検知や照明変化といった問題への頑健性の観点から、AI 技術の一つとして分類される CNN(畳込みニューラルネットワーク)を適用した。水位算定手法の概要を以下に示す。

- ・AI 画像解析モデルにより水面を検知し、検知した水面と構造物(仮想量水標)との対比により、水位を算定。今回は、対岸法面と水面との位置関係により水位を推定(図-10)。
- ・AI 画像解析モデルとして、高精度かつ高速処理が可能な「MobileNetV3」を採用。

表-4 対象 CCTV の諸元

項目	仕様
形式	旋回式カメラ
有効画素数	約 240 万画素
最大口径比	1 : 1.6(WIDE) ~ 1 : 4.7(TELE)
最低照度	カラー : 0.06lx, 白黒 0.02lx
ズーム比	30 倍(光学)
画角	水平 : 2.3° (TELE) ~ 64° (WIDE) 垂直 : 1.3° (TELE) ~ 38° (WIDE)



図-9 CCTV の観測画角

■水面検知

・水面画像を学習した AI モデルを作成し、適用地点の水面エリアを検知



■水位予測

・検知水面と測量箇所のピクセル数の差で水位を算定する



図-10 水面検知による水位算定手法の概要

表-5 水面検知の精度検証結果

出水ケース	転移学習前			転移学習後		
	正検知率	誤検知率	未検知率	正検知率	誤検知率	未検知率
No.1	0.74	0.04	0.22	0.97	0.00	0.03
No.2	0.72	0.04	0.24	0.99	0.01	0.00
No.3	0.00	0.02	0.98	0.96	0.01	0.03
No.4	0.76	0.05	0.19	0.98	0.01	0.01
No.5	0.00	0.01	0.99	0.95	0.01	0.04
No.6	0.80	0.04	0.16	0.97	0.01	0.02

- ・インターネット上に公開された河川監視カメラ画像 3,000 枚を用いて、水面の特徴を学習したモデルを構築。
 - ・上記モデルに、対象 CCTV の画像 150 枚を追加学習する「転移学習」を実施したモデルを構築。
- 出水時に得られた CCTV 画像を用いて、「転移学習前モデル」および「転移学習後モデル」を用いた水位観測手法を適用し、精度を検証した。転移学習前モデルでは、水位観測において実用上十分な精度が得られなかった。しかし、転移学習により水面の正検知率が 20%程度上昇し(表-5)、水位観測精度も上昇した(図-11)。

前述のAI画像解析モデルを用いる水位算定手法により実用性を確保するには、現地画像を用いた学習(転移学習)の適用が必要といえる。これに対し、転移学習を必要とせず、かつ水位算定の精度向上が期待できる方策として、Trimap[®]を活用した手法(Trimap法)を試みた。本手法は、AIを用いた検知プログラムに対し「元画像」と「Trimap画像(地点毎に手作業で作成)」を入力することで、検知対象とそれ以外の部分の境界を表す「出力画像」を得るものである。Trimap画像は白(検知対象)、黒(検知対象外)、灰色(不明)の3色で構成され、灰色に指定した不明な部分について、検知プログラムにより検知対象(水面)との境界が判定される(図-12)。前述の転移学習後モデルとTrimap法による水位算定の比較を図-13に示す。同図によると、昼間の水位だけでなく、水面の視認が困難な夜間においても1.0m以内の水位算定精度を確保しており、転移学習後モデルと比べて精度向上を確認した。

(3) STIVによる流速算定手法の検討

画像を用いる非接触流速観測手法として、Optical_Flow⁷⁾、LSPiV⁸⁾、STIV⁹⁾が挙げられる。これらのうち、CCTV画像を用いた手法として、様々な河川表面の映り方に対応できる汎用性の観点からSTIVを選定した(表-6)。

現地において標定点測量を行い、画像上のピクセル空間と実空間を対応付ける幾何補正処理モデルを構築した。また流速観測を横断的に行うため、左岸から右岸まで5m間隔で18本の測線を設定した。

対象CCTVにおいて出水時に撮影された動画にSTIVを適用し、流速分布の妥当性を検証した。また、観測地点の上下流に位置する水位観測所の流量をHQ式により算定し、CCTV地点の換算流量との比較を行った。検証結果を以下に示す。

・出水があった時刻における流速は、河道の中心付近で大きく、両岸に向かって小さくなり、自然な流速分布となっている(図-14)。

・CCTV地点の観測流量は、上下流の地点流量の中間に収まっており、概ね確からしい(表-7)。

STIVによる流速観測の精度低下の要因として、「出水規模が小さい場合に水面が波立ちにくい」、「夜間における照度不足」などが挙げられる。これに対して、動画へのフィルタリングの適用による精度向上策を試みる。同手法は、画像や動画の解像度を擬似的に上げる超解像度技術「Real-ESRGAN¹⁰⁾」を用いて水面検知動画に対しフィルタリング処理を行うものである。動画のフィルタリング手順を以下に示す。

- ①動画ファイルをフレーム単位(900フレーム)に分割し、jpg形式で保存
- ②分割したjpg画像をフィルタリング処理
- ③フィルタリング処理後の画像ファイルを動画ファイルに変換

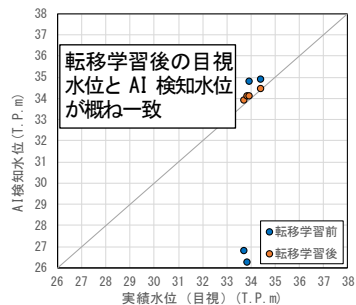


図-11 水位観測の精度検証結果

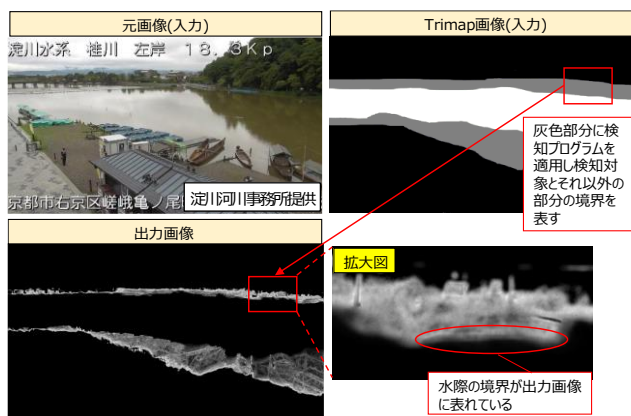


図-12 Trimap法による水位算定のイメージ

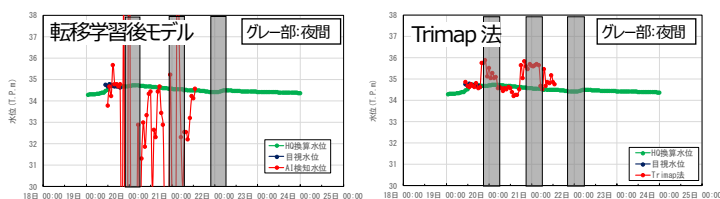


図-13 転移学習後モデルと Trimap 法の水位算定精度比較

※目視水位：水面位置を目視により判別しピクセル数をもとに水位に換算した値

表-6 流速算定手法の選定

手法	概要	採用
Optical Flow	・輝度の勾配、時間変化から動画中の特徴点を追跡する。 ・特徴点となるマーカー散布が必要となる。	
LSPiV	・ある大きさ一定の矩形を設定し、矩形内の河川表面に現れる波紋パターンを相関により、面的に追跡する。 ・波紋パターンの時間変化への頑健性が弱く、マーカー粒子の使用が推奨される。	
STIV	・検査線上の輝度を積み重ねたSTI画像に現れる波紋パターンの勾配から流速を得る。 ・波紋パターンの時間変化に頑健で、マーカー粒子を必要としない。	○

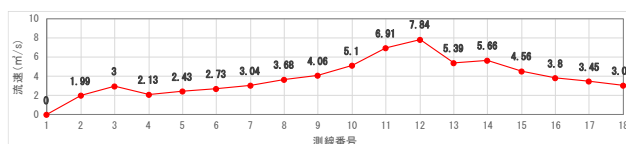


図-14 STIVを用いた流速算定結果

表-7 CCTV地点と上下流地点の流量の比較

位置関係	上流 ^{※1}	CCTV地点 ^{※2}	下流 ^{※1}
地点名	保津峡地点	18.3KP	天竜寺地点
流量(m³/s)	47.12	68.78	91.52

※1：当該地点のHQにより流量を算定

※2：CCTV地点の水位とSTIVによる流速より流量を推定



図-15 フィルタリング処理によるブロックノイズの軽減

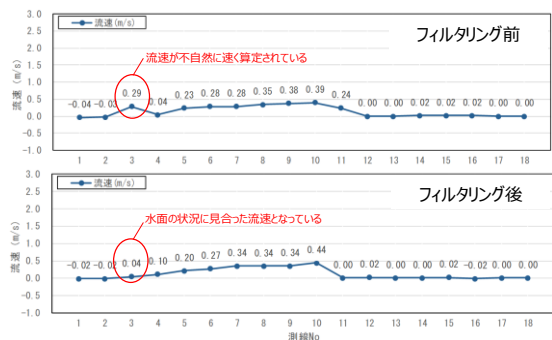


図-16 フィルタリング処理による流速観測の精度向上

表-8 水位・流速算定手法の他地点への展開に関する着目点

着目点	水位観測	流速観測
水面の視認性	日中・夜間に水面を確認できるか	日中・夜間に水面の波紋を確認できるか
夜間照明	夜間照明があるか	
左右岸の見通し	-(特になし)	左右岸を見通せるか
川幅	-(特になし)	川幅が概ね 100m 以内で遠方の水面を認識できるか
仮想量水標の想定	仮想量水標の想定が可能な構造物があるか	-(特になし)

動画のフィルタリングに関する考察を以下に示す。

- ・フィルタリング処理により水面のブロックノイズ(画像圧縮によるモザイク状のノイズ)が軽減され、不自然な流速の算定結果が是正された(図-15、図-16)。これより、流速観測の精度向上が期待できる。
- ・動画のフィルタリングは処理負荷が高く、高性能PCを用いた場合においても、30秒間の動画ファイルに対して90分程度を要する。このため、「水位・流速観測システム」の利用方法を踏まえて、同手法の適用を判断することが望ましい。

(4) 水位・流速算定手法の他地点への展開について

既存CCTVは、必ずしも水位や流速の観測に適した画角となっていない。そこで、これまでの検討を踏まえて、既存CCTVを活用した水位・流速算定手法を他の既存CCTVへ展開する場合の着目点を整理した(表-8)。

あとがき

本検討では、一級河川における非接触型流速計設置予定箇所の検討事例を踏まえ、電波式流速計および画像式流速計の設置に向けた検討における留意点を、手引き案に記載されていない

内容も含めてとりまとめた。この内容は、今後設置を検討する地点に対しても参考となる内容である。課題としては、画像式流速計の夜間・豪雨時等の悪条件下における観測可能範囲を明確にすることが挙げられる。今後、画像式を含めた非接触流速計の設置事例が予定されており、観測事例が一定程度蓄積された時点で、精度検証を行う必要がある。

既存CCTVの活用検討では、CCTV画像から水位及び流速を算定する手法の適用性を検討した。既存CCTVを水位・流速観測へ適用するには、いくつかの着目点を満足する必要があるが、夜間の観測には課題が残るが、一定の精度が得られることが確認できた。本手法を適用することで、新たな観測機器の設置が不要であり、コストを下げずに多くの地点で流量観測が可能となる。今後は、本手法の運用事例を増やし、さらなる高精度化・汎用化を目指す必要がある。

本検討に際し、国土交通省四国地方整備局、近畿地方整備局淀川河川事務所のご厚意により、論文執筆の許可をいただいたことに感謝の意を示す。

参考文献

- 1) 水管理・国土保全局:“革新的河川技術プロジェクト(第4弾)”, 国土交通省, 2023. https://www.mlit.go.jp/river/gijutsu/innovative_project/project4.html
- 2) 非接触型流速計測法の手引き(案), 国土交通省, R5.3, 228p.
- 3) 藤田一郎他:複数アングルの画像計測とラジコンボート搭載型ADCPによる融雪洪水流の空間計測, 土木学会論文集B1(水工学) Vol. 70, No. 4, I_613-I_618, 2014
- 4) 浮子観測データとの整合性の確認の実施方針(案), 国土交通省, R5.3.29, 6p.
- 5) 河川砂防技術基準 調査編, 国土交通省, R4.6.9改定, 第2章 第4節 -18
- 6) Hao Lu, et al.: Indices matter: Learning to index for Deep Image Matting, Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.
- 7) Musaab Khalid, Lionel Pénard, and Etienne Mémín: Application of optical flow for river velocimetry, 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). IEEE, 2017.
- 8) Ichiro Fujita, Marian Muste, and Anton Kruger: Large-scale particle image velocimetry for flow analysis in hydraulic engineering applications, Journal of hydraulic Research 36.3 (1998): 397-414.
- 9) Ichiro Fujita, Hideki Watanabe, and Ryota Tsubaki: Development of a non-intrusive and efficient flow monitoring technique: The space-time image velocimetry (STIV). International Journal of River Basin Management 5.2 (2007): 105-114.
- 10) Xintao Wang: Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data, <https://arxiv.org/pdf/2107.10833.pdf>