

# 都市の安全と利便性向上に寄与する 視覚情報処理

—中国映像ソリューション—

Peng Yang

今田 努  
Imada Tsutomu

Liu Kai

松原 大輔  
Matsubara Daisuke

中国のスマートシティにおいて高画質映像装置とそのプラットフォーム、および関連サービスが急速に普及しており、高度な映像解析・検索を用いたアプリケーションや事業に大きな可能性をもたらしている。しかし、映像符号化および視覚空間の特性（容量、複雑性、多様性）により、映像コンテンツに含まれている情報や知見を完全に把握するには障害も多い。それらを克服するため、WiseFinderや関連プラットフォームとともに中国映像ソ

リューションを構築した。超高速の映像検索、高度な映像解析、強力な組み込みフロントエンドプラットフォーム、システム負荷の小さい映像処理、柔軟な映像プラットフォームといった、重要な技術が用いられている。中国における技術およびビジネス慣習は著しく顧客指向であることを踏まえると、高度なセキュリティを確保する、市街地向けの高品質なサービスと、付加価値のあるアプリケーションといった形での貢献に大きな可能性を持っている。

## 1. はじめに

中国の社会全体の急速な発展を背景に、マルチメディアコンテンツ、特に映像コンテンツを制作できる関連システムおよびプラットフォームが多数登場している。それに伴い、制作される映像コンテンツの品質も向上し、高ビットレート化やデータサイズの増大を招いている。映像コンテンツの情報エントロピーは、他のデータ形式（例えば文章やログデータ）よりも大幅に高い。そのため、映像装置とプラットフォームを利用して構築できるサービスおよびアプリケーションは多岐にわたる。

CCWが2015年に発表した予測<sup>1)</sup>によれば、この分野は今後5年にわたって1兆米国ドル規模の事業を支えたとみられている。映像を活用する市街地インフラは飛躍的な進歩を見せている（表1参照）。関連分野に関する市場およびサービスの可能性も非常に大きいと期待されているが、こうした急速な発展はシステム性能および技術力の面に重いプレッシャーをかけている。何をおいても、映像コンテンツのエンコード、デコード、ストレージ、分析、検索、および管理といった各機能を大幅に強化する必要がある。映像コンテンツから情報を最大限に引き出すために、映像・画像情報の分析、取得に際しては、マルチメディア情報を記述、格納、および整理することが求められる。そう

するためには、先進的なアルゴリズムおよびプラットフォームを使用して、出来事（イベント）をリアルタイムで把握し、映っている対象物およびリソースの検索を簡単かつ迅速に行えるように支援することが必須である。

中国映像ソリューションは、スマートシティで視覚情報処理を行う際の技術的課題の克服を目的に設計・開発された。映像分析・検索プラットフォーム WiseFinder<sup>※1)</sup>、高性能な映像情報処理組み込みプラットフォーム、および大規模な映像プラットフォームで構成されている。スマート

※1) 中国で開発された、大規模かつ高度な映像解析および検索を行う製品。

表1 | 映像関連市街地インフラの移行

中国では映像を活用する市街地インフラが飛躍的に進歩している。

	従来	現在
用途	安全のみ	安全、交通、社会サービス、企業向けなど
範囲	区域およびカメラの台数を限定	市街地の全体に、多数のカメラ
画質	CIF/4CIF	720 P/1,080 P。近々4Kに。
データの量（1日）	カメラ1台当たり2~5 Gバイト	カメラ1台当たり20~50 Gバイト、市中部のビデオ映像3,500万時間
帯域幅	ナローバンド	有線/無線ブロードバンド
情報処理	なし	広範な処理
プラットフォーム	単独/個別	統合/クラウド

注：略語説明 CIF (Common Intermediate Format)

シティの安全、QoS (Quality of Service: サービスの質)、および付加価値の高いアプリケーションを改善するために、高度な映像検索、分析、管理、および最適化といった先進的な機能が多く盛り込まれている。

## 2. 現状および課題

### 2.1 現状

中国の大都市では、3つの分野において高度な映像活用およびサービスが顕著な発達を見せている(図1参照)。

最も急速に伸びている分野は、公共安全、より具体的に述べるならば高度な映像監視である。テロの脅威を踏まえて、中央政府、各州、および地方自治体は映像監視の強化に力を入れている。企業や非営利団体も政府自治体と同様の問題に直面している。すなわち、産業スパイ、怠業、および窃盗などへの対策である。小売業界でも、損失を防ぐために店内セキュリティカメラへの投資が行われている。政府自治体や企業は、映像監視をセキュリティ施策の中心分野と位置づけている。この分野では、都市の保安部署のように大規模なシステムを利用する組織のほうが、映像の分析および検索の機能を十全に活用する可能性が高い。その一方で、セキュリティ部門からの他の情報を映像処理結果と統合し、セキュリティ機能の強化につなげている。

もう一つの重要な分野は、都市生活におけるサービスの質を高めるための映像処理活用である。交通、銀行、環境保護などが、巨額な投資の対象となっている。公共サービスのリソースを調整するうえで、都市環境をリアルタイム

に把握することの重要性は高まる一方である。こうした要望を満たすには、映像分析と検索が好適な手段といえる。映像解析の機能の多様化(顔、車両、複雑な動作、環境変化など)が、顕著な現象である。

都市環境の非営利的な分野に限らず、高度な映像利用は、営利的な分野でも大きな可能性を示している。さまざまな分野の事業主(商業、不動産、報道、交通など)が、映像情報から得られるデータを利用して新たな事業を起したり、既存の事業の業績改善に利用したりしている。例えば、新規サービスを直接販売したり、すでに利益を生んでいるサービスを間接的に強化したりできるのである。上述のとおり、機能の多様化は顧客の側からも求められており、高度な映像機能はインターネット/モバイル通信と組み合わせられて、一般市民にとっての利便性向上や普及推進に貢献している。

高度な映像技術の長所としては、設置が容易で、保守の手間もかからず、豊富な情報と機能が単独で得られるということもある。各サービスに広範に適用でき、サポートも強力であることから、利用例の数は急速に増加している。

他方、高度な映像技術を十分に活用するためには、技術面に多くの課題があることも確かである。

### 2.2 技術的な課題

#### (1) 大量の映像コンテンツを検索する際の効率

関連する機材の性能が飛躍的に向上したため、生成される映像コンテンツの量も大幅に増えている。それにつれ

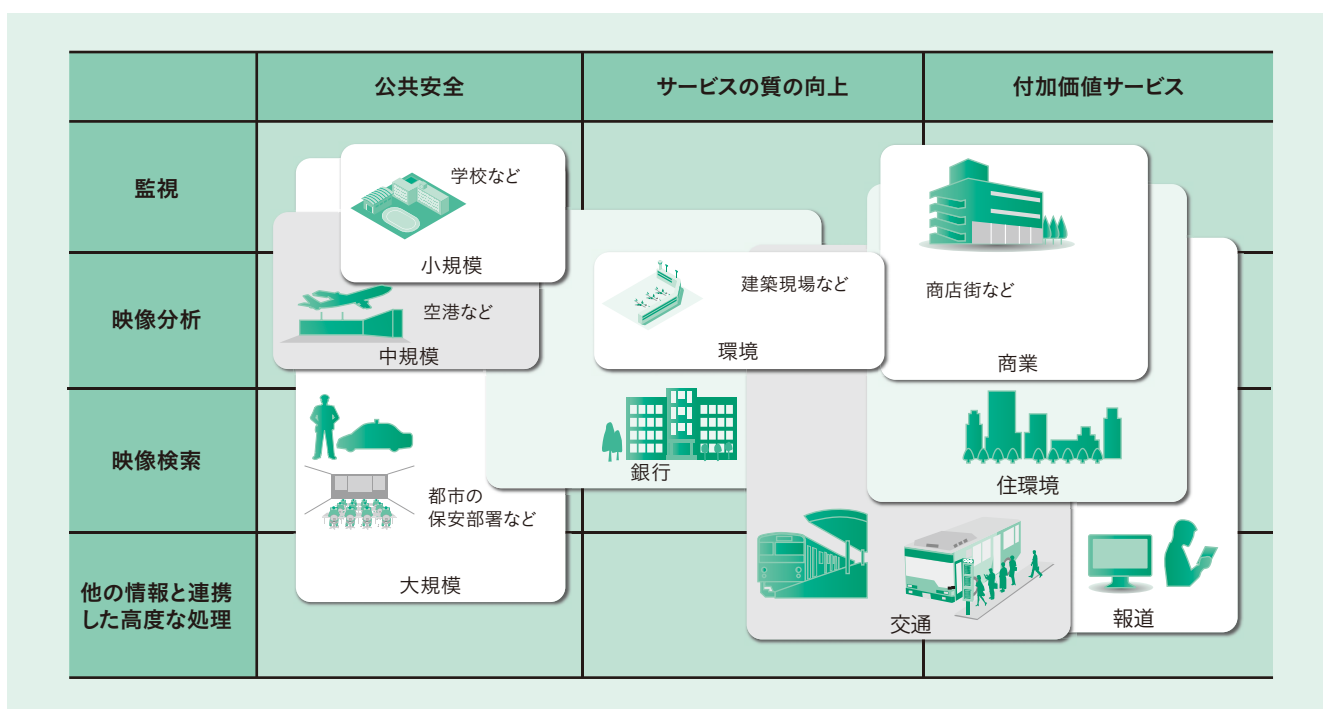


図1 | 技術と適用分野の対応

スマートシティの要素に応じて、サービスや技術の要件も異なる。高度な映像技術に関しても幅広い多様性がみられる。

て、膨大な量のコンテンツの中から、目的の対象物や特定のビデオクリップをどうやって探し出すかが大きな問題となっている。例えば、都市レベルの大規模な映像監視システムでは、個人の顔を検索するために1,200人/月を要する場合があり得る。これはシステム運用者やエンドユーザーにとって多大な負担であり、サービスの利便性と品質を大きく損なっている。中には、映像の内容をテキストで記述するために、複数のタグを追加するシステムを作成した利用者もいる。こうした方法は、小規模なシステムであれば問題はないが、大規模なシステムにおいては、タグを作成すること自体にも膨大なリソースを消費することになる。また、大量の映像コンテンツに説明を付与することは非常に困難である。そのうえ、コンテンツが同じであっても、付与するタグに個人差が生じることがある。つまり、人それぞれの違いのために、テキストベースの映像検索の使い勝手には限界がある。

そこで、高速かつ高精度な、コンテンツベースの映像検索技術が必須となる。

## (2) ITリソースの課題

高度な映像設備を大規模に展開する際の主要な技術課題の一つに、多大なIT (Information Technology) リソースの消費がある。すなわち、CPU (Central Processing Unit) などの情報処理リソース、メモリやディスクなどのストレージリソース、およびネットワークリソースである。例えば、HD (High-definition) 映像の送信ストリーム1本は、ネットワーク帯域6 Mビット/s、エンタープライズ向けCPUの1コア、デコード処理用のメモリ512 Mバイトを消費する。さらに、映像解析では、1つの特徴につきCPUの1~2コア、メモリ1 Gバイトが使用される。1日分のストレージとして、ハードディスク50 Gバイトと、サービスの信頼性を確保するための相応のバックアップリソースが必要である。1つの都市でビデオカメラ数十万台が稼働する大規模システムともなれば、こうした関連コストも膨大な負担となると考えられる。

また、高性能なITハードウェアを展開できない場合も多い。動きの多い場所、例えば公共交通機関の車両、工事現場などでは、設置のスペース、電力供給、およびサーバやストレージの冷却などに制約が生じる。しかし、こうした設備における顧客も、映像情報処理を必要としている。

これを踏まえると、映像のシステムおよびアプリケーションは、リソース節約のために最適化されるだけでなく、現場環境のあらゆる状況に耐えられるようにカスタマイズされるべきである。

## (3) 複数の視点からの多様な要件

映像コンテンツは、テキストや音声と比較して極めて多

くの情報量を含んでいる。映像に情報コードが付加されていれば、さらに豊富な情報を探することも可能となる。ただし、産業分野が異なれば、映像コンテンツで探す対象となる出来事や事物も大きく異なったものになると考えられる。

人間一人ひとりが視覚的に空間をどう認識しているかは、それぞれ大幅に異なっており、これを記述あるいは描写することは不可能である。この同じ原則が、高度な映像解析の分野にも該当する。どのような特徴を映像監視で把握したいかについて、顧客を対象にした市場調査を実施したところ、非常に幅広い回答が得られた。例えば、人物の生物学的特徴、個人あるいは集団の行動の特徴、車両の特徴、都市環境の特徴、およびこれらの特徴の組み合わせや遷移である。こうした機能すべてを単一の企業で開発することは、コストや期間の点で莫大な負担となりうる。

その一方で、都市環境においては、さまざまなタイプの映像システムがすでに展開されている。それらの多くは、高度な処理機能を備えているわけではない。新たに高度な映像設備を展開する際、こうした旧来のシステムとの統合は極めて困難になる。あらゆる種類の端末機器、既存の映像管理システム、多数の顧客が使用している多種多様なアプリケーションやサービスシステム、既存のIT設備、およびさまざまな管理プラットフォームに対応しなければならない。

これまでに述べたとおり、高度な映像技術には、スマートシティにおけるセキュリティ、生活の質、および価値の向上に関する大きな可能性がある反面、技術的な課題も多い。こうした状況の中、高度な視覚情報処理によって都市のセキュリティ、利便性、および価値を高めるため、中国映像ソリューションは設計・開発された。

## 3. 中国映像ソリューション

中国映像ソリューションの主要なアーキテクチャは、WiseFinderの各機能と、強力な組み込みフロントエンドプラットフォーム、そして負荷の小さい解析機能である(図2参照)。

WiseFinderの機能は以下のとおりである。

### (1) データ処理エンジン

都市環境の静的な場所にある装置およびシステムから集められた映像コンテンツについて、ETL (Extract, Transform, Load) を主に行う。

### (2) 映像解析エンジン

各種の映像解析を実施する主要なコンポーネントである。データ処理エンジンから、バッファされたキーイメージを取得する。高度な解析をJPEG (Joint Photographic Experts Group) フレームベースで行い、その結果として特徴ベクトルを抽出する。高度な解析の機能は、アプリケー

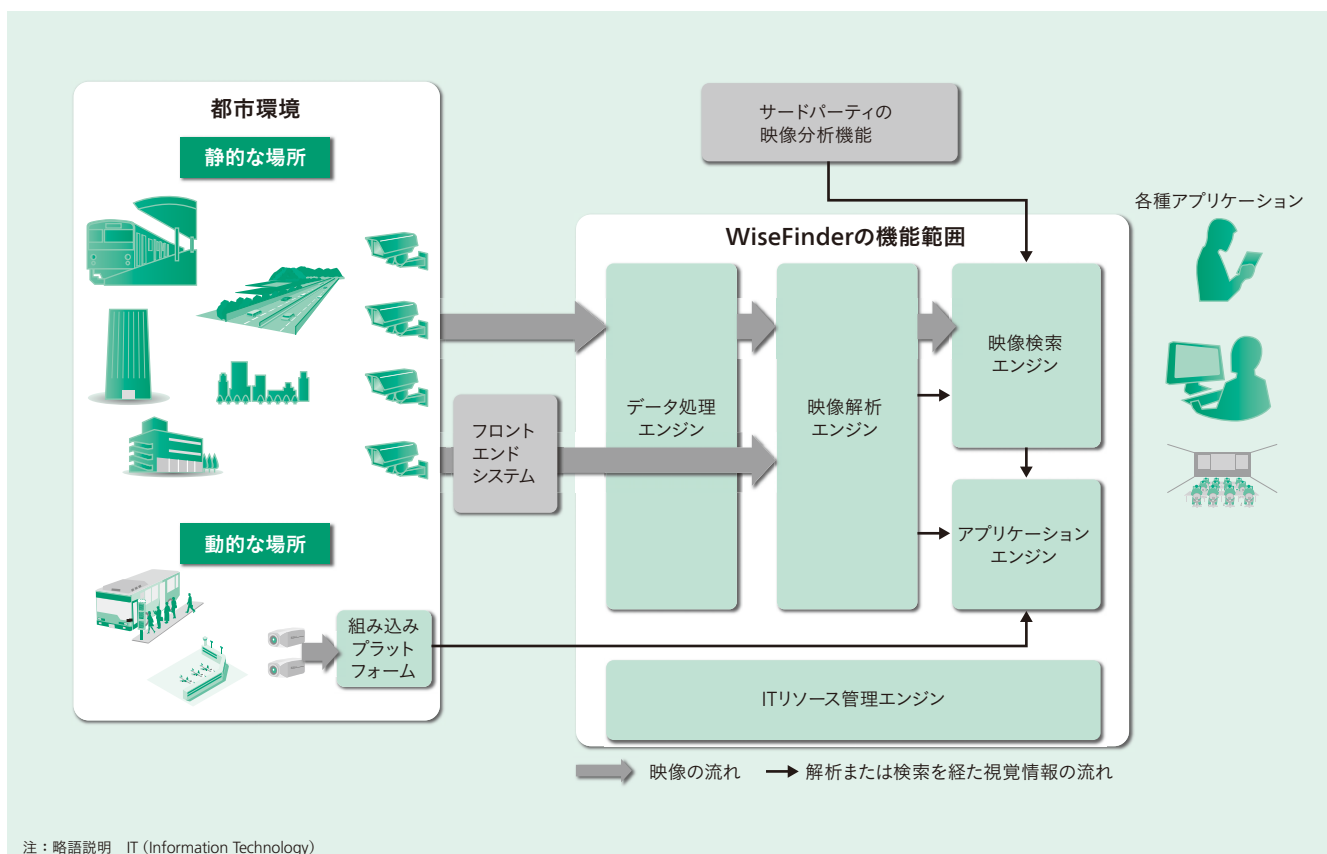


図2 | ソリューションアーキテクチャの概要

静的な場所では、ビデオカメラが設置され、信頼できる有線ネットワークと、サーバを保護する環境が用意されている。動的な場所は車両や建築現場などであり、移動する可能性があるため、無線ネットワークを用い、稼働環境は厳しい。

シヨンのシナリオに沿って設定・保守するものである。例えば、公共安全の分野では、人間の顔、自動車、自動車のナンバープレートなどを検出する必要がある。

各種の特徴は、解析モジュールと関連する数理モデルによって作成され、特徴ベクトルによって記述されたうえで、映像検索の主要コンポーネントである映像検索エンジンの特徴データベースに登録される。

### (3) ITリソース管理エンジン

静的な場所にあるすべてのITリソースを集中管理する。アプリケーションエンジンは、外部アプリケーションおよびサービス向けのポータルである。

一方、組み込みフロントエンドプラットフォームは、動的な場所で映像解析を実施するために作成された。2.2で言及した技術的な課題を解決するために、多くの重要な技術が開発された。

## 3.1 コンテンツベース映像検索の機能強化

大規模な映像検索システムの初期バージョンは、日立製作所研究開発グループが開発した<sup>2)</sup>。このシステムの強みは、大規模なシステムで高速な映像検索を処理することである。一般的なPC (Personal Computer) サーバを用いて、キーイメージ1億点を対象にした単一のコンテンツベース

検索を1秒以内に完了することができる。実地試験では、大規模な映像データベースで複数のユーザーが対象物を検索した場合でも、高速映像検索の効率が99%以上向上するという結果が得られている。

中国では、超大規模の映像監視環境に対応することを目的に、以下に述べる新技術を開発した。

### (1) イタレイティブサーチ

超大規模な映像システムでは、映像解析のノイズとなる要素が非常に多岐にわたる。その結果、大量の特徴ベクトルを対象にした検索の精度が損なわれる。そこで、多段階の検索を行って検索精度を維持するイタレイティブサーチを開発した(図3参照)。

### (2) 場面ベースフィルタ処理機能

検索結果では、同じ場面の連続したフレームにある同じ事物が一緒に示される可能性が高い。このため、複数の場面に存在する事物を示したい場合には、有用な情報が減ってしまうことになる。そこで、アプリケーションエンジン内にフィルタ機能を組み込んで、1つの場面から1つの例だけを抽出し、重複する部分はフィルタできるようにした。

### (3) クラウドにおける動的なリソース調整

映像システムの規模は拡大することが予想されるため、顧客はIT基盤としてプライベートクラウドを選ぶ。2.2で

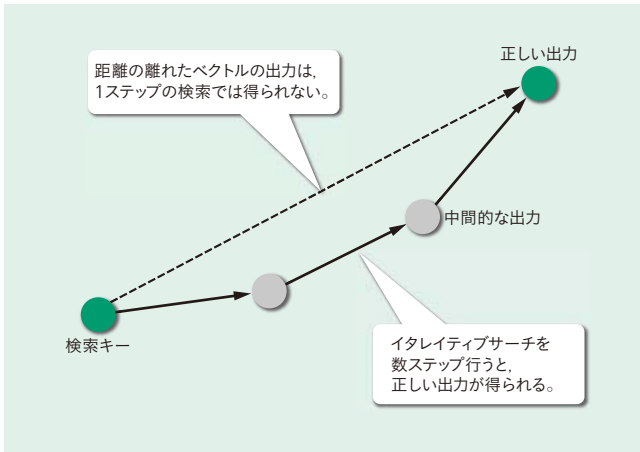


図3 | イタレイティブサーチの原理

大規模な映像検索システムでは、映像解析の結果として生成される特徴ベクトルも多数となり、莫大なノイズが生じる。イタレイティブサーチは、ノイズの影響を回避し、数ステップで正しい結果を得ることを目的として開発した。

述べたとおり、高度な映像技術は大量のITリソースを消費する。このため、仮想マシン管理によるリソース調整が必要となる。実験では、IT調整に高度なアルゴリズムを用いることで、リソースを20%節約することができた。

### 3.2 強力な組み込みフロントエンドプラットフォームおよび低負荷の映像解析

前章で示したとおり、動的な場所の多くでは、従来の解析に使用する装置(サーバなど)を展開することができない。

この問題に対処するために、中国で初(日立調べ)のオールインワン装置を開発した。組み込み用Linux<sup>※2)</sup>をベースにしたオープンプラットフォームと、それに見合ったコンピュータリソース [CPU+DSP (Digital Signal Processor)], ビデオ録画, および高性能ネットワーク (Wi-Fi<sup>※3)</sup>, 2G, 3G)を組み合わせるとともに、厳しい環境に耐える仕様としている。

一方、コスト抑制のため、複数の映像チャンネルに対する複数の解析を1台でサポートしている。このため、従来は多大なITリソースを消費していた解析アルゴリズムに関して、低負荷のバージョンが開発された。

市場に投入するプロセスを大幅に短縮し、かつ早めるため、プラットフォームはどのような低負荷アルゴリズムにも対応できるように設計されている。

### 3.3 多様なインタフェース

前章で述べた多様性に起因する障害に対処するため、多様なインタフェースを開発した。

※2) Linuxは、Linus Torvalds氏の日本およびその他の国における登録商標あるいは商標である。

※3) Wi-Fiは、Wi-Fi Allianceの登録商標である。

(1) フロントエンドデバイス/システムとのインタフェース  
データ処理エンジンは、中国にある多数のカメラベンダーの製品とのインタフェースをサポートしている。また、さまざまなストレージからのオフラインコンテンツのアップロードにも対応している。

(2) サードパーティによる映像解析とのインタフェース  
この業界の急成長により、多くの企業が映像解析のアルゴリズムを開発している。映像検索エンジンのインタフェースでは、外部の高度な処理機能による結果を登録することが可能である。これにより、業界で使用されている各種の機能を十分に活用することができる。

(3) 多様なアプリケーションのためのWeb API(Application Programming Interface)

このインタフェースを用いて、あらゆるアプリケーションが短期間でこのソリューションの機能を活用することが可能である。

## 4. 中国における適用事例

### 4.1 公共安全のための大規模映像検索

公共安全は規模が大きいため、膨大な映像データベースから目的の人物や自動車を検索する方法が大きな問題となった。サービスの典型的な流れを図4に示す。基本的に、検索機能は検索対象の絞り込みと、目的となる事物を素早く正確に見つけることに利用される。

中国の2都市において顔検索と自動車検索を実験した結

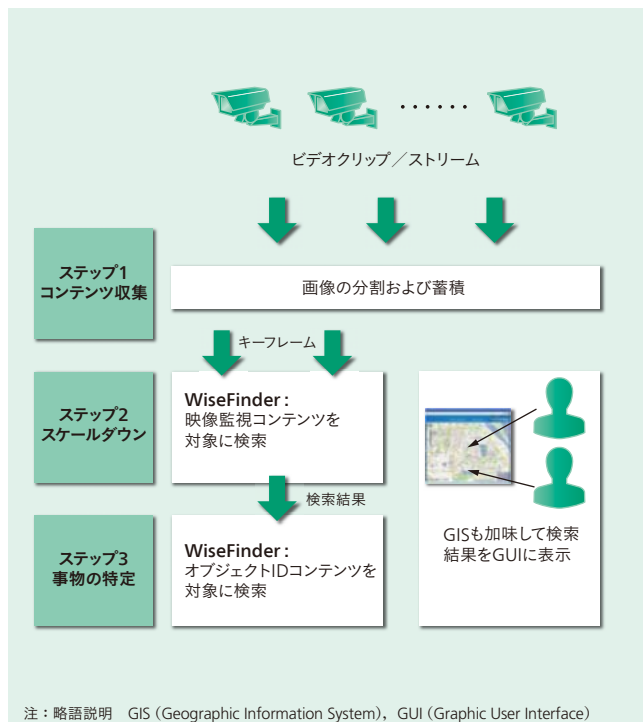


図4 | 公共安全のための映像検索の一般的なワークフロー

映像・画像検索を使用し、目的の事物(人や自動車など)を絞り込む。大量の映像コンテンツから、素早く正確に検索する。

## 表2 | 公共安全のための映像検索の結果

クイック検索の精度は90%を上回り、実際の場面では検索速度の向上率が99%を超えた。

カメラ	都市1	都市2
1920×1080	100% (繰り返し1~2回)	該当無し
1280×720	100% (繰り返し1~2回)	該当無し
720×576	100% (繰り返し2~3回)	90% (繰り返し1~3回)

果を表2に示す。このソリューションを用いたクイック検索の精度は90%を上回った。実際の場面では、検索速度の向上率が99%を超えた。

### 4.2 高度な車両プラットフォームによる公共交通機関の 利便性と付加価値の向上

公共交通機関の調整は長年にわたって実施されてきたが、スマートシティにおける交通リソースの利便性は最適化されているとは言い難い。主な問題は、現行の調整の仕組みが車両の外部にある情報に依存していて、車内にいる乗客の状況を考慮していない点にあった。

組み込みフロントエンドプラットフォームを用いた映像解析による乗客の混雑状況や人数の検出が、中国のいくつかの都市において実現され、運用中である。バスや地下鉄の車両ごとの乗車状況に合わせて、バス、各種車両、地下鉄、幹線道路などの公共交通リソースの利用をさらに最適化することが可能となる。

## 5. 将来的な方向性

中国映像ソリューションは中国の市場において有望と評価されているが、さらなる改善を追求すべき分野も多数残されている。例えば、このソリューションをスマートシティの各分野を縦断する形ではなく、1つの分野の各階層で利用するように機能強化することも可能である。この目標を達成するには、将来的にさらなる解析機能を開発したり、あるいは統合したりすることが必要となる。また、動的な環境を有する多くの分野あるいは現場では、耐久性のある高度な組み込みフロントエンドプラットフォームを使用できる。それにより、より低負荷の映像解析機能を、このプラットフォームに移行することも可能である。

中国における高度な映像アプリケーションの利用においては、ITリソースの過度な消費が依然として大きなボトルネックの1つである。将来的には、アルゴリズム、プラットフォーム、および管理機能をさらに最適化することが焦点となる。

## 6. おわりに

1兆ドル規模に達する中国の高度な映像処理技術市場をターゲットにして開発した中国映像ソリューションは、スマートシティ環境における安全、利便性(サービスの質)、およびアプリケーションの価値を高めることをめざしている。世界を見渡しても先進的な映像検索・解析、ハイブリッド映像プラットフォーム、多様なインタフェース、および高度な最適化機能を通じて、こうしたソリューションが人間の視覚を活用する可能性をさらに高めてくれることを期待したい。

技術面における友好的な提携関係は、産業界の全体を次の段階へ進める前向きな貢献につながる。本稿で述べたようなソリューションに基づき、さらに有用なアプリケーションおよびサービスが中国の都市部で生まれていくと考えられる。

### 参考文献

- 1) China Intelligent Video Market and Necessary Technologies, CCW research report 2015
- 2) 廣池, 外: 大規模に蓄積された画像および音声を対象とする情報検索, 日立評論, 95, 2, 200~205 (2013.2)

### 執筆者紹介



#### Peng Yang

日立(中国)研究開発有限公司 中国社会イノベーション協創センター 所属  
現在、都市セキュリティプロジェクト向けの高度なマルチメディアソリューションの開発に従事



#### 今田 努

北京日立北工大情報システム有限公司 事業戦略室 所属  
現在、中国およびアジア太平洋市場における画像検索ソリューションの事業開発に従事



#### Kai Liu

北京日立北工大情報システム有限公司 製品ソリューション本部 映像交通ソリューション部 所属  
現在、中国市場における画像検索ソリューションの事業開発に従事



#### 松原 大輔

日立製作所 研究開発グループ  
東京社会イノベーション協創センター 顧客協創プロジェクト 所属  
現在、都市セキュリティプロジェクトに従事  
電子情報通信学会会員