

インテリジェントな監視カメラネットワークを実現する画像解析ボックス

Add-on Type Image Analysis Box for Realization of Intelligent Surveillance Camera Network Systems

小坂谷 達夫

丸山 昌之

岡田 隆三

■KOZAKAYA Tatsuo

■MARUYAMA Masayuki

■OKADA Ryuzo

監視カメラの台数は年々増加しているが、多数のカメラ映像をクラウドサービスを利用して解析するためには、膨大な通信容量と計算資源をクラウドサービス側で確保しておく必要があり、現実的ではない。

そこで東芝は、画像認識プロセッサ Visconti™2を搭載したアドオン型の画像解析ボックスを開発した。既設の監視カメラネットワークに画像解析ボックスを接続するだけで、防犯及びセキュリティや、マーケティング、公共エリア監視、防災など次世代の街づくりに役立つ画像処理結果をクラウドサービス側に送信する、インテリジェント機能を容易に付加できる。画像解析ボックスは処理結果だけを送信するため、ビデオ映像を直接送信する従来システムと比べて、ネットワーク通信量を1/7~1/2,000に削減できることを実データにより確認した。

With the increase in the number of surveillance cameras in recent years, it has become necessary for cloud computing providers to set up large-capacity networks and computing resources for this purpose. However, analysis of a large number of video streams from distributed cameras using the cloud is not a realistic approach.

As a solution to this issue, Toshiba has developed an add-on type image analysis box equipped with an image recognition processor. When this image analysis box is connected to an existing camera network by a simple procedure, it offers the intelligent function of sending useful image recognition results for next-generation urban development, including in the areas of security, marketing, observation of public areas, and disaster prevention, to the cloud. We have conducted verification tests using actual video data and confirmed that the image analysis box can reduce network traffic volumes to between one-seventh and one two-thousandth in comparison with the case of directly sending H.264 video streams to the cloud, by sending only low-capacity processing results instead of the video data of all cameras.

1 まえがき

街中に設置されている監視カメラの台数は年々増加しており、2009年～2015年の年平均増加率は約22%と予測されている⁽¹⁾。監視カメラシステムの主な目的は映像の録画と有事の際の確認であるが、近年、映像データを画像認識技術により解析することで、セキュリティ用途のほかに省エネやマーケティングなどに活用する事例が出てきている。

監視業務のクラウドサービス化を実現するために、画像認識のシステム構成として、現地にパソコンを設置して処理する方法と、映像データをクラウドサービス側に送信して中央で集中的に処理する方法が考えられる。しかし、24時間365日の稼働が要求される監視カメラシステムには産業用パソコンが使用されるが、非常に高価であり設置及び運用のコストが高い。また、カメラの映像データは容量が大きく非常に冗長なため、クラウドサービスに多数のカメラ映像が集中すると、膨大な通信容量と計算資源をクラウドサービス側で確保する必要があり、現実的ではなかった。

そこで東芝は、画像認識処理を高速に実行できるプロセッサ Visconti™2^{(2), (3)}を搭載し、画像処理結果だけを出力するアドオン型の画像解析ボックスを開発した。これまでにカメラ型の

デバイスを試作しているが⁽⁴⁾、監視カメラに接続するアドオン型にすることで、新たにカメラを設置することなく、既存の設備を有効活用できる。画像解析ボックスは、画像認識プロセッサと制御用マイコンを搭載し、LANポート経由で、ネットワークカメラからの映像入力と画像処理結果の送信を行う。画像解析ボックスを追加することで、防犯及びセキュリティや、マーケティング、公共エリア監視、防災など次世代の街づくりに役立つ画像センシング結果をクラウドサービス側に送信する、インテリジェントな監視カメラネットワークを実現可能である。

2 画像解析ボックスの概要

開発したシステムは、図1に示すように、ネットワークカメラ、ネットワークハブ、及び録画機で構成される既存の監視カメラシステムと、そのローカルネットワークに追加した画像解析ボックスから構成される。画像解析ボックスは、指定したネットワークカメラから映像を取得し、画像処理結果を、例えばクラウドサービス側のデータセンターなどに送信する。

システムの機能ブロック図を図2に示す。ネットワークカメラからReal Time Streaming Protocol (RTSP) で入力されたH.264^(注1)ビデオ形式で圧縮された映像は、画像解析ボックス

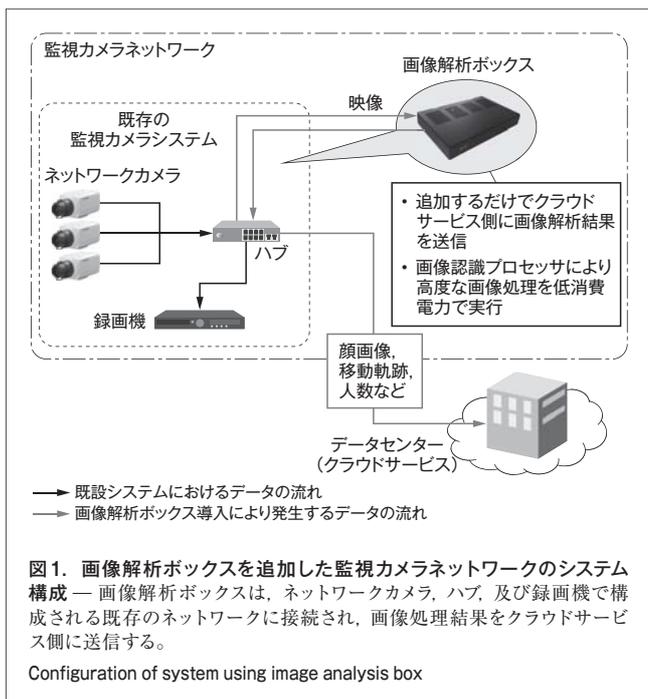
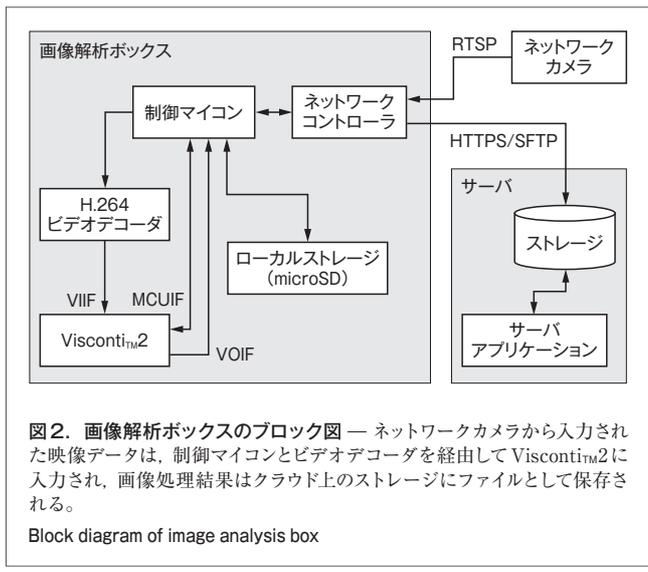


表1. 映像解析ボックスのハードウェア仕様
Specifications of image analysis box

項目	仕様	
入力画像	接続カメラ台数	1～4台 (カラー)
	圧縮 (動画)	H.264-30 フレーム/s (最大)
	解像度	1,280×1,024画素 (SXGA) ~ 640×480画素 (VGA)
処理チップ	画像処理	Visconti _{TM} 2
	制御マイコン	ARM Cortex TM (注2) A9 MP 667 MHz
インタフェース	ストレージ	SDカード (SDHC対応)
	有線LAN	100/1000Base-TX: 1ポート RJ45コネクタ
環境	電源	・PoE対応 ・DC12V, 1.5A以下
	温湿度対応	・温度: -10 ~ 50℃ ・湿度: 20 ~ 80%RH, 結露なし
外形	寸法	190 (幅) × 126 (奥行き) × 28 (高さ) mm
	質量	500g以下

DC: 直流 RH: 相対湿度

ルとして保存される。サーバアプリケーション ソフトウェアは、前述のストレージを監視しており、新たなデータを検知するとそのデータ処理を行う。ここで、メタデータをファイルベースで送信するのは、大量のデータが送られた場合にサーバ側が過負荷状態に陥ることを防ぐためと、サーバアプリケーション ソフトウェアがなんらかの不具合で停止や再起動した場合に、メタデータが欠落しないようにするためである。



の制御マイコンを経由してビデオデコーダで復号され、ビデオインプットインタフェース (VIF) から Visconti_{TM}2 に入力される。Visconti_{TM}2 と制御マイコンは、マイクロコントローラユニットインタフェース (MCUIF) とビデオアウトプットインタフェース (VOIF) により接続されており、容量が小さい情報は MCUIF から、また、画像などを含む容量が大きなデータは VOIF から出力される。

映像解析ボックスが送信する画像処理結果 (メタデータ) は、基本的には Secure Shell File Transfer Protocol (SFTP) により、クラウドサービス側にあるサーバのストレージにファイ

映像解析ボックスでは、ネットワークカメラからの映像は基本的には解析用で外部には出力しないが、ローカルストレージとして microSD カードスロットに処理結果の画像などを保存しておくこともできる。映像解析ボックスは HTTPS (Hypertext Transfer Protocol Secure) サーバの機能を備えているので、SDカードに保存したファイルには HTTPS 経由でアクセス可能である。

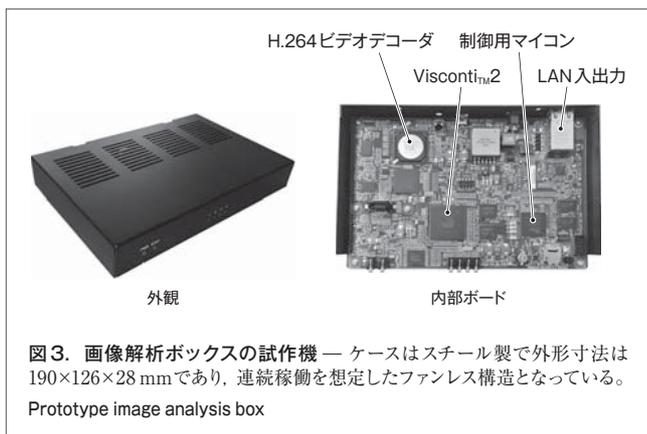
映像解析ボックスの主要なハードウェアの仕様を表1に示す。最大4台のネットワークカメラを接続可能であり、H.264形式のビデオ映像を入力できる。電源は、LANコネクタからの PoE (Power over Ethernet)、又は AC (交流) アダプタが利用でき、映像解析ボックス全体の消費電力は 10 W 程度である。

試作した映像解析ボックスの外観と内部ボードを図3に示す。ケースはスチール製で、外形寸法は 190 (幅) × 126 (奥行き) × 28 (高さ) mm であり、24時間365日の連続稼働を想定したファンレス構造となっている。

また、産業用途を考えて、このボックスを設置後には可能なかぎり遠隔でソフトウェアのメンテナンスができるように、Webベースの設定用画面を持たせている。設定用画面では、接続先の監視カメラのアドレスや接続状況などの情報、及び制御マイコンや Visconti_{TM}2 のパラメータ設定だけでなく、ファーム

(注1) ITU-T (国際電気通信連合-電気通信標準化部門) により策定された動画圧縮規格の一つ。

(注2) ARM, Cortex は、英国 ARM Limited の商標。



ウェア自体のアップデートもネットワーク経由で遠隔地から実行できる。

3 画像解析処理の概要

画像解析ボックスは、ファームウェアを書き換えることで様々なプログラムを実行することができるため、まったく同じハードウェアでありながら、用途の目的に応じた様々な機能を実現できる。

Visconti™2は、CoHOG (輝度勾配方向共起ヒストグラム) 特徴量⁽⁵⁾の抽出や、線形SVM (Support Vector Machine)⁽⁶⁾による識別など、様々な画像処理に適したハードウェアアクセラレータを搭載しており、これらを有効に活用することで最大のパフォーマンスが得られる⁽⁷⁾。また、全ての処理をボックス内で実行するのではなく、容量が大きなデータベースなどはクラウドサービス側にあるデータセンターのサーバ上に置くなど、処理を適切に分割することも重要である。

ここでは、監視カメラシステムの応用場面ごとの画像解析処理例について述べる。

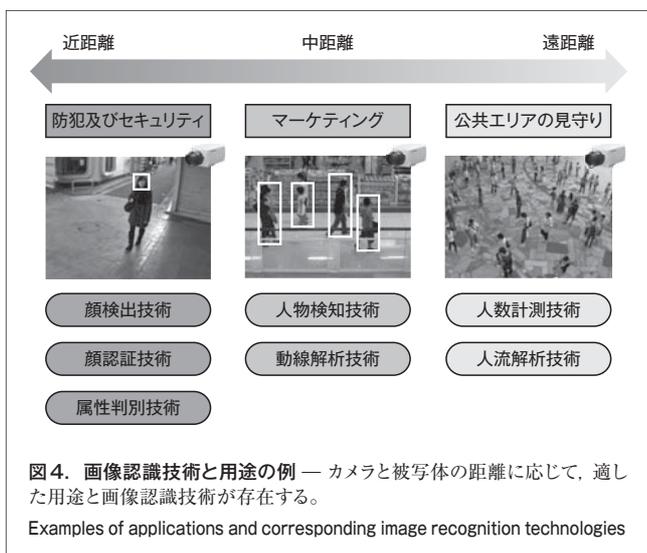


図4. 画像認識技術と用途の例 — カメラと被写体の距離に応じて、適した用途と画像認識技術が存在する。
Examples of applications and corresponding image recognition technologies

監視カメラは目的に応じて設置されており、被写体との距離や被写体のサイズはその用途ごとに異なる。カメラと被写体の距離に応じた用途と画像認識技術の例を図4に示す。

セキュリティ用途では、有事の際に顔から個人を判別する必要があるため、対象を大きく撮影する必要がある。この場合の画像処理は、まず画像解析ボックスが顔の検出や追跡を実行し、検出された顔部分の画像をメタデータとしてクラウドサービス側に送信する。クラウドサービスでは顔部分の画像から特徴を抽出し、事前に登録した顔データベースと照合することで、特定人物や迷子を発見できる。

また、マーケティング用途では、店舗内に設置されたカメラごとの人物の移動軌跡を画像解析ボックスが出力すれば、どの棚の前にどのくらい滞在したか、買い物客はどの商品に興味を示したか、といった情報を得ることができる。画像解析ボックスは人物の動線情報を抽出し、クラウドサービス側でカメラの設置位置に基づいて店内の絶対座標に変換し、動線情報を得る。買い物客の動線情報をデータベースに蓄積し、統計的に解析することで、マーケティングに必要な情報に変換できる。

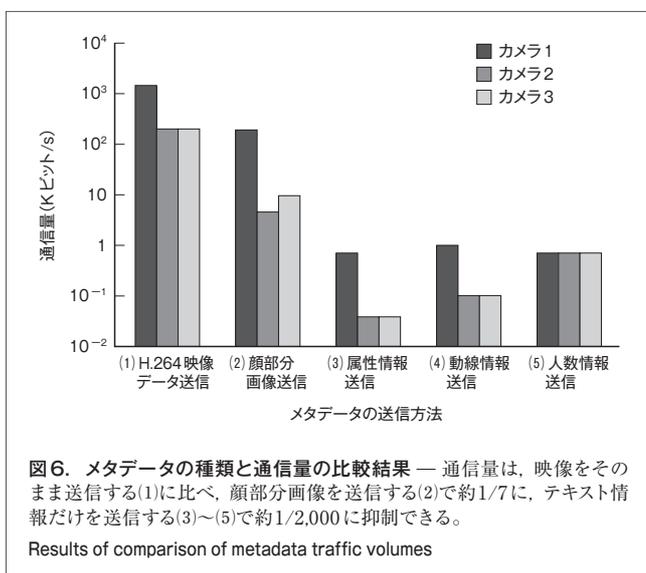
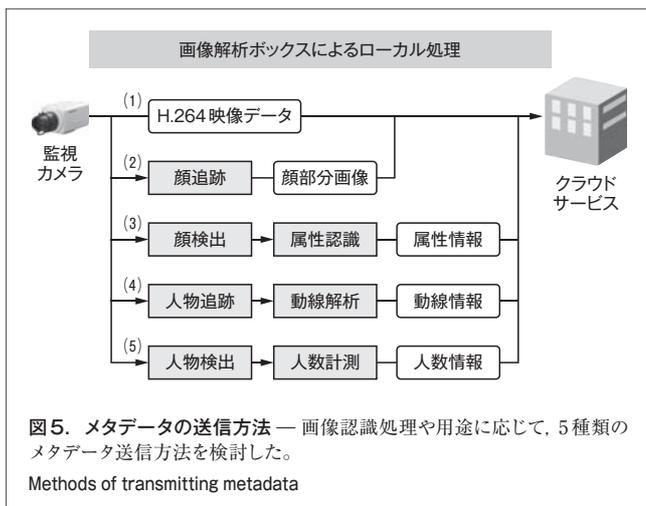
公共エリアの見守り用途では、駅や公園などの状況を把握することに主眼が置かれるため、見下ろすように撮影されることが多い。画像解析ボックスは、画像中の人数や人流情報を解析し、メタデータとして定期的送信する。人数の増減や人流情報が定常状態とかけ離れた値となっていれば、公共エリアになんらかの異常が発生している可能性がある。更に、災害発生時に混雑しているエリアを避けるなど、避難誘導への活用も期待できる。

4 通信量の削減効果

監視カメラシステムが送信するメタデータの種類や通信量は、用途や撮影環境に依存する。例えば、顔認識によるセキュリティ用途の場合は、画像解析ボックスは顔の検出と追跡までを実行し、検出された部分画像はクラウドサービス側に送信されて認証される。このとき、通行人数が多ければ多いほど通信量は増大する。

画像解析ボックスによる通信量の削減効果を検証するため、実際の通路とビルの入り口にカメラを設置して映像データを収集した。設置場所は、通行量が非常に多い通路 (カメラ1) と、通行量が比較的少ない部屋の入口 (カメラ2, カメラ3) の3か所である。各カメラは、10フレーム/秒のスピードと800×450画素の解像度で90分間撮影し、約15万枚の画像データを得た。これらの映像に対して顔の検出・追跡処理を行い、カメラ1で4,532個、カメラ2で270個、及びカメラ3で276個の人物検出結果を得た。

比較するメタデータの送信方法は、複数の用途を想定し、次の5種類を検討した (図5)。



- (1) H.264による映像データをそのまま送信
- (2) 顔の追跡処理を行い、顔部分画像を送信
- (3) 属性情報(年齢、性別)だけを送信
- (4) 人物追跡処理を行い、動線情報を送信
- (5) 人物検出を行い、人数だけを送信

通信量の比較結果を図6に示す。グラフの縦軸は通信量をKビット/sの対数目盛で表している。映像ストリームを直接送信する場合が通信量が多くなる。次に多いのは顔部分の画像を送信する場合で、テキスト情報だけを送信する属性情報、動線情報、及び人数情報は通信量が1Kビット/s以下であった。顔部分の画像を送信する場合でも映像データの送信と比較して約1/7に、テキスト情報だけを送信する場合には約1/2,000になり、大幅に通信量を抑制できることが確認できた。

5 あとがき

高い信頼性を誇る画像認識プロセッサ Visconti™2を搭載したアドオン型の画像解析ボックスを開発した。既存の監視カメラネットワークに接続することで、画像認識によるインテリジェント機能を容易に追加できる。

メタデータだけを送信することで通信の負荷を削減し、次世代の街づくりに役立つクラウドサービスと連携した大規模な監視カメラネットワークの実現を目指す。

文 献

- (1) 矢野経済研究所. 世界のネットワークカメラ市場に関する調査結果2013. <<http://www.yano.co.jp/press/press.php/001103>>, (参照2014-02-10).
- (2) Tanabe, Y. et al. "A 464GOPS 620GOPS/W heterogeneous multi-core SoC for image-recognition applications". ISSCC 2012. San Francisco, CA, USA, 2012-02, IEEE, 2012, p. 222-223.
- (3) 東芝. 画像認識プロセッサ Visconti™ファミリー. <<http://www.semicon.toshiba.co.jp/product/assp/automotive/infotain/visconti/>>, (参照2014-02-10).
- (4) 小坂谷達夫 他. 画像認識LSI "Visconti™2"を搭載したインテリジェントカメラ. 東芝レビュー. 68, 2, 2013, p.12-14.
- (5) Watanabe, T. et al. Co-occurrence Histogram of Oriented Gradients for Human Detection. IPSJ Trans. on Computer Vision and Applications. 2, 2010, p. 39-47.
- (6) Vapnik, V. The Nature of Statistical Learning Theory. Berlin, Springer-Verlag GmbH, 1995, 188p.
- (7) 岡田隆三 他. "次世代車載画像認識LSIによる安全運転支援". Vision Engineering Workshop (ViEW)2011. 横浜, 2011-12, 精密工学会. 2011, p.268-274.



小坂谷 達夫 KOZAKAYA Tatsuo

研究開発センター インタラクティブメディアラボラトリー研究主務。画像処理及びパターン認識の研究・開発に従事。情報処理学会会員。

Interactive Media Lab.



丸山 昌之 MARUYAMA Masayuki

研究開発センター インタラクティブメディアラボラトリー研究主務。画像処理技術及び画像認識技術の研究・開発に従事。

Interactive Media Lab.



岡田 隆三 OKADA Ryuzo, Ph.D.

研究開発センター インタラクティブメディアラボラトリー主任研究員、博士(工学)。画像認識技術の研究・開発に従事。電子情報通信学会会員。

Interactive Media Lab.