

車載マルチメディア Onboard Multimedia

前田 賢一
MAEDA Ken-ichi

赤嶺 政巳
AKAMINE Masami

加瀬 直樹
KASE Naoki

次世代の自動車(以下、車と略記)には、様々な情報を処理し、ドライバーを支援する機能が求められている。ここで述べる各技術は、その実現に大きく貢献するコア(核心)技術である。安全運転を支援する車載画像認識は、車載カメラによる画像の処理にオプティカルフローを利用して、後側方からの接近車両を検出する。更に平面投影ステレオの原理を利用し、前方の障害物を検出する。耐雑音音声認識、自然音声学習による音声合成は、機器操作のHMI向上に有効である。そしてエージェント技術を利用したドライバー情報支援システムにより、膨大な情報の中から必要な情報を容易に入手し、活用できる環境を整えることができる。

Functions that process a variety of information in order to assist the driver are required in the next-generation vehicle. The following technologies will contribute to the production of such an advanced vehicle. Image recognition that detects obstacles in front and vehicles approaching from the rear and side will be useful in supporting safe driving. This system will use an optical flow for processing images input from onboard cameras. Noise-resistant speech recognition and the text-to-speech synthesizer will be effective in improving the human-machine interface of the onboard equipment. The InfoMirror™ agent-based driver's information assistant will enable a driver to obtain required information easily from a huge amount of data.

1 まえがき

車は、単に“走る”、“止まる”、“曲がる”の基本機能を備えた移動のための道具としてではなく、知能化された移動空間へと進化しつつある。つまり、安全運転支援、利便性の向上を目的とするアプリケーションシステムから、車内にビジネス環境やエンターテインメント(音楽、映像、ゲームなど)を提供するシステムまで、様々な機能を備えた高機能車へと変貌(へんぼう)を遂げようとしている。

当社は、車の高機能化においてキーとなる要素技術の開発を推進しているが、ここでは、当社の車載事業の主軸になると期待される技術の中から、以下の三つについて述べる。

- (1) 車載画像認識 安全運転を支援する車載カメラの画像処理技術
- (2) 音声処理技術 車内における機器操作のHMI (Human Machine Interface)を改善し、安全性・利便性を向上させる音声認識/合成技術
- (3) エージェントを利用したドライバー情報支援システム 氾濫(はんらん)する情報を有効に活用できるような、次世代車載アプリケーションの統合技術

2 車載画像認識

従来、車載カメラは業務用大型車両の後方視界の悪さをカバーするために利用されてきた。これは、車がバックする際、ドライバーがモニタを介して後方を見るという利用法で

ある。最近では、ワンボックスカーとカーナビゲーション(以下、カーナビと略記)の普及に伴い、家庭用の車でもカメラを搭載したものが見られるようになってきている。ここでも画像を見るのは人間である。人間はミスをする恐れがあり、見るのを忘れてたり、見間違えたりする可能性があるため、これをカバーする意味で自動認識が必要とされてきた。

2.1 後側方画像監視

車載後側方画像監視は、車線変更の際に後側方からの接近車両を検出して、ドライバーに警告を出すことを目的とするものである。ドライバーは車線変更をする際に、ミラーによって後方を確認する決まりになっているが、ついっかり見忘れてたり、見間違えたりした場合のバックアップ手段として位置づけられるものである。

自車が動いているので、接近車両だけでなく背景全体も動いて見えるため、接近車両だけを認識するためにオプティカルフローと呼ばれる画面の動きベクトルを利用する。

オプティカルフローの計算は、複数フレームの画面全体に対して各部分の移動量を計算することになる。移動量を求めるには各部分をフレーム間でマッチングする必要があり、データ量が多いため多くの計算を必要とする。

しかし、後方から接近してくる車両は、自車との相対速度が比較的小さいので、探索範囲が小さくて済むという特徴がある。これを利用して後側方監視用の高速アルゴリズムを考案した。その結果、パソコン(PC)だけのシステムで従来のDSP(Digital Signal Processor)複数以上の計算速度を達成することができた。



図1. 後側方監視試作システム 後側方監視試作システムのディスプレイ画面を示す。

Prototype rear and side surveillance system

これを利用した後側方監視試作システムのディスプレイ画面を図1に示す。画面上部のバーは警報を表わす。

2.2 移動障害物認識

後側方の監視では、オプティカルフローの監視は比較的容易であった。その理由の一つは、後方から接近してくる車両との相対速度が比較的小さいことによる。しかし、一般の方向、特に前方では他の車両との相対速度は小さいとは限らない。したがって、ここでは従来どおりの本格的なオプティカルフロー計算が必要となる。

2.3 移動可能空間検出

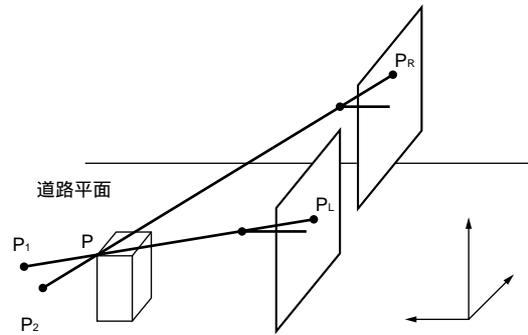
オプティカルフローの利用は動いている対象物を認識するのに有効であった。しかし、障害物としては、駐車中の車両や落下物など、静止したものも多々ある。こうした障害物を検出するためには、オプティカルフローを求めるだけでは不十分である。

障害物までの距離を測定する手段としてステレオ視技術がある。ステレオ視技術というのは、複数(最低2個)のカメラを用いて三角測量と同じ方式で距離を求める技術である。しかし、一般にステレオ視技術は異なるカメラの画像間での対応づけに計算時間が掛かるという問題と、カメラをセットする際の調整(キャリブレーション)が困難であるという問題を抱えている。

この問題を解決するために開発されたのが平面投影ステレオである。これは、当初、自律移動車の視覚技術として開発され、現在より一けた遅いIPCを用いても秒速1m程度で移動する試作車両を誘導することができた。

単一のカメラで見た場合、あるテクスチャ(画面中の模様、構造)が道路の模様起因するものか、障害物が存在することに起因するものかを決定するのは困難である。換言すれば“物がある”のか“物の絵が描いてある”のかを判定することが困難であるということである。

平面投影ステレオは、これを逆手にとって、すべてが地面に描かれた絵であるという見方をしてみる。そうすると本物の絵



P: 障害物上の一点
 P_R, P_L : 対応する右, 左のカメラ上の点
 P_1, P_2 : 道路面上に描かれた絵であると見なしたときに対応する点

図2. 平面投影ステレオの原理 平面投影ステレオの原理で、障害物を左右のカメラから道路面の絵と見なしたようすを示す。

Principle of planar projection stereopsis method

であれば別のカメラで見た画像が一致するのに対して、障害物の場合には画像は一致しない。このようすを図2に示す。

3 音声処理技術

カーナビの普及に伴い、車載機器と人間とのインタフェース、すなわちHMIが重要な要素となっている。運転中に機器を操作したり、ディスプレイを注視することは安全性の面から大きな問題があるため、音声によるインタフェースが必要とされている。今後、インターネットを用いた情報提供サービスの進展が予想され、音声によるインタフェースがますます重要になると考えられる。ここでは、音声によるHMIを実現するうえで必要になる音声認識と音声合成の技術について述べる。

3.1 音声認識

車載機器では、車内の走行雑音による誤認識が大きな問題であり、雑音抑圧の技術と雑音に強い認識方式の開発に注力している。

耐雑音性の向上のために、図3に示すように入力音声を認識用の音声特徴ベクトルに変換する特徴抽出処理と、認識処理の2段階で雑音に対処する。特徴抽出部では、音声分析結果を雑音に影響されない安定した特徴量に変換する。まず、雑音重畳音声を用いて雑音特有の変形を抽出する。次に、雑音に影響されにくい特徴空間を生成し、特徴ベクトルをその空間へ射影^(注1)する。更に、実際の雑音環境を模擬した音声を用いて認識部で用いる辞書を学習する。これにより、特徴抽出部で吸収しきれない変形に対処する。学習のポイントは、学習に用いる音声信号の信号と雑音の比(SN比)を変えながら特徴ベクトルを生成し学習を繰り返

(注1) 特徴ベクトルをその空間のベクトルに変換すること(マッピング)

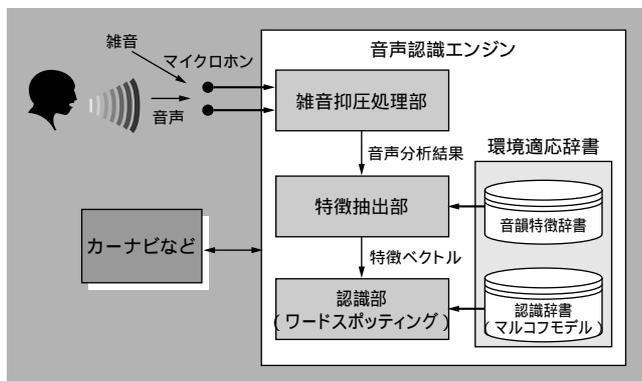


図3. 音声認識システムの全体構成 入力音声を確認用の音声特徴ベクトルに変換する特徴抽出処理と、認識処理の2段階で雑音に対処する音声認識システムの全体構成を示す。
Overview of speech recognition system

すことであり、これを雑音免疫学習法と呼んでいる。雑音に影響されにくい特徴ベクトルと雑音免疫学習法により、走行中の高騒音の中でも高い認識率を達成することができるようになった。

また、利用者の発声時の制限を緩和するため、キーワードスポッティング^{注2)}による話しことばを対象とした意図理解方式を開発した。ここでは、話しことばに含まれる不要語に対処するため、入力音声を一字一句認識するのではなく、あらかじめキーワードとして定められた単語だけを抽出し、キーワードのつながりから利用者の発話の内容を理解する。これにより、不要語だけでなく、語順の逆転や言いよどみなど、話しことば特有の現象にも対処でき、利用者の負担を大幅に軽減できる。

3.2 音声合成

走行ルートの音声案内や音声認識のトークバック^{注3)}のため、音声合成がカーナビで採用され始めている。車中では、走行中に情報を確実に違和感なく伝達する必要があるため、明瞭(めいりょう)で自然なイントネーションの合成音が望まれている。

当社の音声合成方式の全体構成を図4に示す。この方式は、あらかじめ収録した自然音声のデータベースから合成に用いるイントネーションやリズムのパターンと、音声素片辞書を自動的に学習する機構を持ち、学習で得られた音節素片のピッチや継続時間長を制御した後、これらを接続して音声信号を生成する。

音声素片辞書の学習は、合成音と自然音声とのひずみの総和を最小化する規範に基づいて行われる。合成音声と自然音声とのひずみを学習にフィードバックすることからこの方法を閉ループ学習と呼んでいる。これにより、従来、音質

(注2) 発声時のことばの中から、キーワードとして定められた単語だけを抽出する技術。
(注3) 音声認識システムが認識した単語を利用者に知らせるため、その単語を読み上げること。

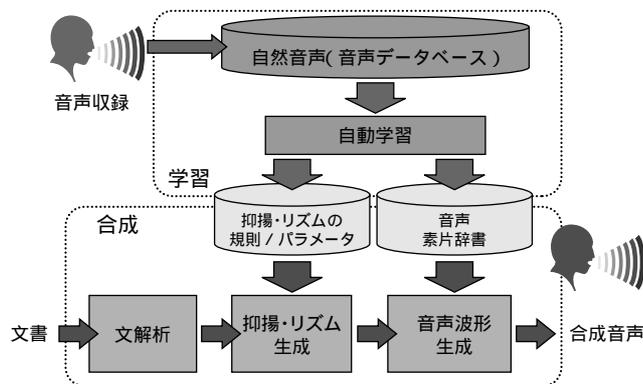


図4. 音声合成システムの全体構成 自然音声のデータベースから合成に用いるイントネーションやリズムのパターンと、音声素片辞書を自動的に学習する機構を持つ音声合成システムの全体構成を示す。
System overview of text-to-speech synthesizer

劣化の主要因であったピッチ変更に伴うひずみが激減し、明瞭で肉声感豊かな音質を実現できるようになった。イントネーションのパターンの学習も音声素片辞書の学習と同様に閉ループ学習法に基づいて行われる。これにより、自然なイントネーションを持つ合成音声が可能になったばかりではなく、学習に用いた話者の話し口調に類似した音声を合成できるようになった。

また、この方式では音声素片を辞書として格納する際、スペクトルパラメータと残差信号の組で音声素片を表現している。このため、スペクトルパラメータの操作により声質の変更が容易であるとともに、音声素片辞書のサイズを大幅に圧縮することができる。11.025kHzサンプリング出力の場合、一話者当たり約150Kバイト、22.05kHzサンプリングの場合で約300Kバイトと非常にコンパクトである。このため、カーナビなど車載機器への組込みも容易である。

4 エージェントを利用したドライバー情報支援システム

ドライバー情報支援システム InfoMirrorTMは、エージェントを利用することによって、動画像や音声など様々な形態で提供される情報に柔軟にアクセスし、車載の適切なデバイス(装置)での情報の再生を可能とするフレームワーク^{注4)}である。

インターネットに代表されるコンピュータネットワークの普及により、利用者に提供される情報が膨大になってきている。にもかかわらず、現状では、車内で扱うことのできる情報は、ユーザーインターフェースやネットワークの制約などから限られている。カーナビ上に表示される付加的な情報が文字だけであったり、動画像を扱うだけの無線通信帯域が確

(注4) システム構築のための開発ツール群やライブラリ群、それに構築方法論などを含めてフレームワーク(枠組み)と呼ぶ。

保できなかったりしている。しかし、今後は、車内で扱うことのできる情報の量や質がともに爆発的に伸びることが予測され、マルチメディアをいかに扱うかの技術開発が必要となってきた。

情報インフラストラクチャが整えば、膨大な情報をダイレクトに車内に伝送することが可能となる。しかし、提供される情報量が多くなるほどアクセスの操作は煩雑になるし、情報をフィルタリング(取捨選択)しなければならない。すなわち、利用者の求める情報を収集し、最適なネットワークチャネルを選択してデータを伝送し、利用者の扱いやすい情報に変換し、適切なデバイスで表示・再生しなければならない。これを行うのがエージェントである。

ドライバー情報支援システムInfoMirror™では、エージェントを利用することによって、ネットワーク上で提供された情報を車載の適切なデバイスで再生することができる(図5)。例えば、音楽情報に適用したInfoMirror™ for Ubiquitous Musicでは、音声入力されたリクエストから聴きたい音楽をネットワーク上で検索し、カーオーディオデバイスを利用して楽曲を再生することができる(図6)。

音声圧縮技術の進歩により、音楽データがネットワーク配信されるサービスが現実味を帯びてきている。となれば、車両の中で、これらのサービスを利用したいという要求は当然発生する。このような場合に、ネットワーク上の音楽配信サービスと車載のカーオーディオというデバイスの接続をエージェントによって実現することができるのである。

InfoMirror™ for Ubiquitous Musicでは、音楽ソースの存在場所は、ネットワーク上の音楽配信サイトに限定しない。購入したCD(Compact Disc)を、自宅でMP3(MPEG1 Audio Layer III)ライブラリとして、ネットワークアクセス可能なデジタルデータ化している場合もあり得るし、将来、ネットワーク回線の速度が向上すれば、音楽データ自体はレコード会社のサイトでダイレクトに提供されることになるかもしれない。

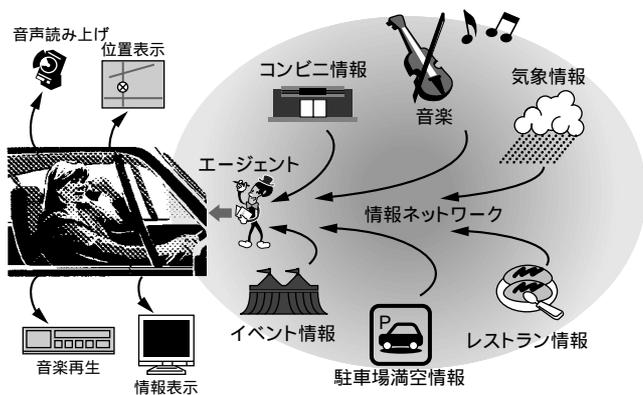


図5. ドライバー情報支援システムInfoMirror™の機能 様々な情報をネットワークから収集し、車載の適切なデバイスを選択して表示、再生する。

Functions of InfoMirror™ agent-based driver's information assistant

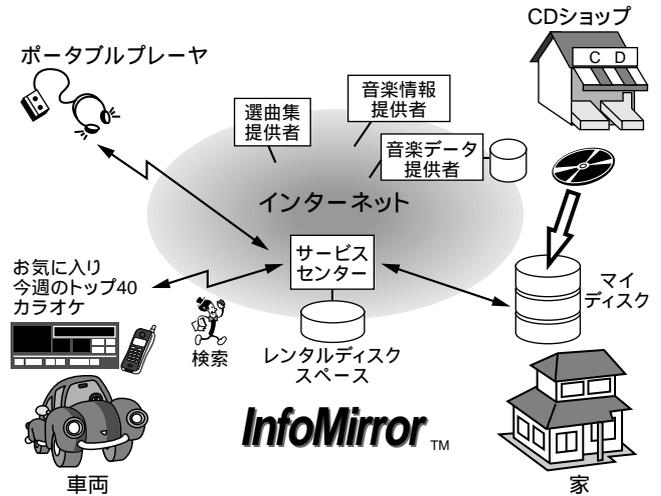


図6. 車両向け音楽情報システムへのInfoMirror™適用事例 家庭用PCやレンタルディスクでライブラリ化した音楽データをカーオーディオやポータブルプレーヤで再生する。

Example of application of InfoMirror™ to music information system

このように、様々な状況がありうる場合には、エージェントが音楽情報の提供方法を柔軟に変更できるような形態が重要になってくるのである。

5 あとがき

ここに述べた技術により、次世代の車に求められる機能のイメージが少なからず理解されたのではないだろうか。しかし、この分野は更に様々なアプリケーションシステム、新技術への展開の可能性が無限にある。また、一方、機能の統合化、車両への合理的搭載も進んでいくであろう。

今後、当社は、この車載技術分野の発展に貢献すべく更に研究・開発に注力していきたい。



前田 賢一 MAEDA Ken-ichi

研究開発センター マルチメディアラボラトリー研究主幹。画像認識の研究・開発に従事。電子情報通信学会、情報処理学会、IEEE会員。Multimedia Lab.



赤嶺 政巳 AKAMINE Masami, D.Eng.

研究開発センター ヒューマンインターフェースラボラトリー主任研究員、工博。音声符号化、認識、合成など音声処理技術の研究・開発に従事。電子情報通信学会、日本音響学会、IEEE会員。Human Interface Lab.



加瀬 直樹 KASE Naoki

研究開発センター コンピュータ・ネットワークラボラトリー研究主務。エージェント技術の研究・開発に従事。情報処理学会、人工知能学会、日本ソフトウェア科学会会員。Computer & Network Systems Lab.