

# 低画質画像及び実環境下を対象としたドライバの顔向き推定

伊藤優司<sup>†1</sup> 小島真一<sup>†1</sup> 倉橋哲郎<sup>†1</sup>

ドライブレコーダでドライバの日常運転を計測する際、低画質、見えの多様性、照明変動等の厳しい条件下での顔向き推定が要求される。本研究ではそれら課題に対して、顔の局所特徴の有無（見える/見えない）の情報に着目する事で低画質向けの顔向き推定手法を構築する。さらに、オンラインで逐次的に局所特徴検出器を学習する事で実環境下の対応を試みる。実際のドライブレコーダ画像を用いて、本提案手法の有効性評価を行う。

## Driver Head Pose Estimation under Real-world in Low Resolution Video

YUJI ITO<sup>†1</sup> SHINICHI KOJIMA<sup>†1</sup> TETSUO KURAHASHI<sup>†1</sup>

In order to analyze daily driving from a drive recorder, we are required to estimate the driver head pose in low resolution videos captured by the drive recorder. Moreover, various appearances of the head are caused by many factors in real-world such as change of the head orientation, individual variation, wear of a sunglass or hat, illumination change and so on. We propose the head pose estimation method in low resolution videos based on the information that each local feature in the face is visible or not. For adaptation of the local features detection to various appearances, we learn the detector in online. The proposed method is validated through the experimental results.

### 1. はじめに

近年、ドライブレコーダで記録された車両挙動に関する情報を用いてドライバの運転を常時診断するためのシステムが開発されてきているが、より有効かつ的確に診断するためには、車両挙動情報だけでなくドライバの行動にも着目すべきである。特に、周囲に対する確認行動は安全性を診断する上で重要と考えられ、顔の向きや視線の情報が必要となる。そのため、実路走行中にドライブレコーダカメラで撮影された動画像に対する顔向き推定手法が求められている。

実路走行中の環境では、ドライバの確認行動時に顔向きの変化が大きい場合や、太陽光等外乱の影響が大きい場合が存在する。一般に普及しているドライブレコーダでは、仕様上、画像処理に有利な高フレームレート及び高画質な画像を取得できる事は少ない。

顔向き検出手法は既に数多く提案されており[1]、例えば、画像から抽出した特徴量を用いて多クラス識別や回帰によって向きを推定する手法がある。これらは事前に収集した多数の顔画像を用いて学習を行う必要があるが、顔の向き、サングラスや帽子等の装着物、環境変動等によって生じる見え方の多様性を考慮すると、膨大な学習サンプルが必要となる可能性がある。

また、顔の局所的な特徴とそれを結合する形状モデルで構成された手法も多く、検出された顔の局所特徴に基づき顔形状モデルのフィッティングを行う事で、モデル形状の幾何学的情報に基づき向きを推定している[2]-[4]。これら

の多くも局所特徴の検出には事前に学習された検出器を使用しているが、局所的な特徴の見え方が顔向きに依存しないものと仮定すれば、向きの情報とは独立に検出器を生成出来るため[4]、学習サンプル数を大幅に削減出来る可能性がある。さらに、検出器をオンラインで学習する事により環境変動等への対応も可能と考えられる。一方で、特徴検出器を用いずに、画像のフレーム間での局所特徴のマッチングから姿勢を追従するトラッキングベースの手法も存在する[5][6]。しかし、顔形状モデルを用いたこれらの手法は、局所特徴の位置情報から幾何学的に向きを算出するため、局所特徴位置や形状モデルの精度が要求される。低画質画像に対して適用する際には、局所特徴位置や形状モデルの誤差にロバストな手法が必要となる。

したがって本研究では、局所特徴と顔形状モデルで構成された、低画質画像に対してロバストな顔向き推定手法の構築を目的とする。さらに、特徴検出器に関しては事前に膨大な学習サンプルは用いず、オンラインで学習する事で環境変動等への対応を試みる。

### 2. 従来手法

以下に関連する従来研究の一部を紹介する。

[2]に記載の Active Appearance Model (AAM)では、まず事前に計測した多数の顔形状を用いて顔の形状変形を再現出来るような可変形状モデルを生成する。その可変形状モデルと共に顔のアピアランスを変化させながら顔モデルのフィッティングを行う。また、[3]に記載の Constrained Local Model (CLM)では、まず事前に生成した可変形状モデルで複数の局所特徴を拘束した条件下で、全局所特徴の類似度の合成値に基づき顔形状モデルのフィッティングを行う。

<sup>†1</sup> 株式会社 豊田中央研究所  
TOYOTA CENTRAL R&D LABS., INC.

これらの手法はフィッティングした際の変換パラメータから顔姿勢の推定が可能であるが、モデルフィッティング時には各局所特徴位置や形状モデルの精度が要求されるため低画質環境下等での適用は困難と考えられる。

[4]では、Latent SVM [7]に基づき局所特徴検出器及び可変形状モデルを事前に学習し生成する。その可変形状モデルで複数の局所特徴を拘束した条件下で、全局所特徴の類似度の合成値に基づき顔の位置及び向きを検出する。[2][3]の手法よりも局所特徴位置の拘束条件が緩和されている形状モデルになり得るが、位置拘束を緩和する事で向きの判別性能が低下する恐れがある。すなわち特徴点位置の誤差には頑健であってもモデルが本来有する姿勢の判別性能自体も低下してしまう。

[5]では、顔形状を円筒と仮定し Lucas-Kanade 法によるアピアランスのマッチングから姿勢変化を推定しているが、低画質環境下や顔の動作が素早い状況等では Lucas-Kanade 法の適用は困難と考えられる。

[6]では、顔形状を円筒と仮定し SIFT 特徴や正規化相関を用いて局所特徴間のマッチングを行うことで姿勢変化を推定しているが、SIFT 特徴は画像平面に平行な軸周りの回転には不向きであり、低解像度画像において顕著となる。

### 3. 提案手法

本章では、まず 3.1 節で本提案手法の顔探索モデルを記述する。次に、3.2 節以降で探索モデルに必要な各要素の詳細を記述する。

#### 3.1 局所特徴の有無に着目した顔探索モデル

低画質画像にロバストな顔向き推定手法を実現するため、本提案手法では以下の2点に着目する。

- (i) 局所特徴位置情報に関する拘束条件の緩和
- (ii) 局所特徴の有無情報に基づく拘束条件の付加

(i)を行う事で、局所特徴位置や形状モデルの誤差による影響を低減できると考えられる。しかし、局所特徴位置に関する拘束を緩和する事で幾何学的に向きを推定するという本来の目的達成が困難になるため、新たに(ii)を導入する事で顔向き推定を試みる。ここで、局所特徴の有無というのは、局所特徴が見えているか否かという1ビットの情報であり、低画質画像においてもロバストと考える。例えば図1のように、画像平面と平行な軸に対して頭部の回転を考えた際、顔の局所特徴の有無は顔の向きによって一意に与えられる。そして、全局所特徴の有無の組み合わせに対して顔の向きは一意に決まるため、全局所特徴の有無を観測する事で顔向きの推定が可能となる。

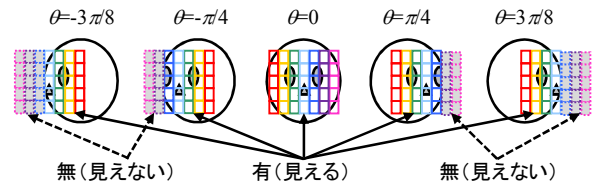


図1 顔向きと局所特徴の有無の関係

これを基に、(i)と(ii)を達成する探索手法を確率モデルとして定式化する。まず、画像内で任意の位置 $(x, y, z)$ 及び向き $\theta$ に対する顔の存在確率は以下で与えられる。

$$p_{obj}(x, y, z, \theta) = p(c_{1,x,y,z,\theta} = \delta_{1,\theta}, c_{2,x,y,z,\theta} = \delta_{2,\theta}, \dots, c_{n,x,y,z,\theta} = \delta_{n,\theta}) \quad (1)$$

ここで、 $c_{i,x,y,z,\theta}$ は顔の重心位置 $(x, y, z)$ に対応する局所特徴 $i$ の有無であり、見えている場合は1、見えていない場合は0とする。同様に $\delta_{i,\theta}$ は、顔の向きによって一意に決まる局所特徴の有無の条件であり、以下のように0あるいは1の値として顔形状モデルと共に事前に定義されている。

$$\begin{cases} c_{i,x,y,z,\theta} = \{0,1\} & (i=1,2,\dots,n) \\ \delta_{i,\theta} = \{0,1\} \end{cases} \quad (2)$$

次に、画像内の任意位置 $(x, y, z)$ を中心とした参照領域に対する局所特徴 $i$ の存在確率を $p_i(x, y, z)$ と定義すると、顔重心位置に対応する局所特徴 $i$ の存在確率 $p(c_{i,x,y,z,\theta})$ は、局所特徴位置に関する拘束条件を考慮した上で $p_i(x, y, z)$ と関連付けられ、形状モデルの定義に該当すると共に(i)と深く関わる。ここで、ある顔向き $\theta$ の状態において見えないはずの局所特徴 $i$ について考える。ある顔向き $\theta$ の状態において局所特徴 $i$ が見えていない事を観測するには、局所特徴 $i$ が存在し得る位置に対して存在しない事を観測する必要がある。すなわち、見えないはずの局所特徴 $i$ の位置は、向き $\theta$ に依らず局所特徴 $i$ が存在し得る位置として定義されていなければならない。これは、顔形状に従って単に向き $\theta$ だけ回転させた場合の局所特徴位置とは異なり、他の顔形状モデルとの相違点である[2]-[4]。図1に示すように、見えないはずの局所特徴 $i$ の位置関係はその局所特徴が見えている時の位置関係が維持されており、結果として、顔向きに対して大きな位置関係の変動が無い形状モデルが要求される。これに基づき、本研究では顔向きに依らず位置関係が一定である顔形状モデルを採用し、4.1節にてその詳細を述べる。また、これを考慮すると、[4][7]の様に局所特徴位置に対して形状モデルからの誤差に応じたペナルティを適切に与える事は困難になると考えられる。本手法では、形状モデルからの誤差に応じたペナルティは与えず、局所特徴位置に対する許容誤差の範囲のみを定義する。具体的には、以下のように $p(c_{i,x,y,z,\theta})$ を定義する。

$$p(c_{i,x,y,z,\theta} = 1) \equiv \max p_i(x + d_{i,\theta,x}, y + d_{i,\theta,y}, z + d_{i,\theta,z}) \quad (i=1,2,\dots,n) \quad (3)$$

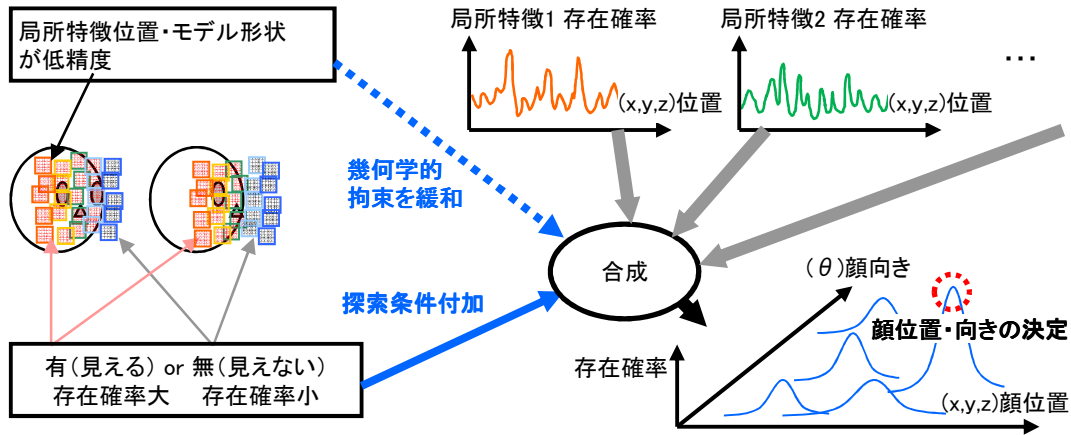


図2 局所特徴の有無に着目した顔探索モデルの概略図

$$p(c_{i,x,y,z,\theta} = 0) = 1 - p(c_{i,x,y,z,\theta} = 1) \quad (i=1,2,\dots,n) \quad (4)$$

ここで、 $(d_{i,\theta,x}, d_{i,\theta,y}, d_{i,\theta,z})$ は顔重心位置 $(x, y, z)$ から局所特徴  $i$  までの相対位置を示しており、以下の式を導入する事で、局所特徴位置に関する拘束条件が付加される事になる。

$$\begin{cases} |d_{\theta,i,x} - d_{0,\theta,i,x}| < \delta_x \\ |d_{\theta,i,y} - d_{0,\theta,i,y}| < \delta_y \\ |d_{\theta,i,z} - d_{0,\theta,i,z}| < \delta_z \end{cases} \quad (i=1,2,\dots,n) \quad (5)$$

ここで、 $(d_{0,i,\theta,x}, d_{0,i,\theta,y}, d_{0,i,\theta,z})$ は事前に定義した顔形状モデルにおいて、顔重心位置から局所特徴  $i$  までの相対位置である。すなわち、 $(\delta_x, \delta_y, \delta_z)$ は誤差の許容量を意味しており、それぞれ0である場合には誤差が一切許容されず、 $(\delta_x, \delta_y, \delta_z)$ の値が大きい程、局所特徴位置や形状モデルの誤差が許容される。これにより、(i)の局所特徴位置情報に関する拘束条件の緩和を行う。

以上で(i)(ii)を達成する確率モデルの定式化が整ったため、これを基に入力画像に対して  $p_{obj}(x, y, z, \theta)$ が最大となる  $(x, y, z, \theta)$ の組を探索する事で、顔の位置及び向きを同時推定を行う。推定した顔の位置及び向きを  $(x_{est}, y_{est}, z_{est}, \theta_{est})$ とした時、それらは以下で与えられる。

$$(x_{est}, y_{est}, z_{est}, \theta_{est}) \equiv \arg \max_{x,y,z,\theta} p_{obj}(x, y, z, \theta) \quad (6)$$

ここで、複数の局所特徴の同時確率分布である  $p_{obj}(x, y, z, \theta) = p_{obj}(c_{1,x,y,z,\theta} = \delta_{1,\theta}, c_{2,x,y,z,\theta} = \delta_{2,\theta}, \dots, c_{n,x,y,z,\theta} = \delta_{n,\theta})$ の値を直接観測する事は困難であるため、単独の局所特徴存在確率  $p(c_{i,x,y,z,\theta})$ を用いて顔の位置及び向き  $(x_{est}, y_{est}, z_{est}, \theta_{est})$ を推定する問題に帰着させる。図2は本提案手法の概略図であり、(i)(ii)の条件下において局所特徴存在確率を合成し、最終的な顔の位置及び向きを探索している。また、 $p(c_{i,x,y,z,\theta})$ を用いるためには、 $p_i(x, y, z)$ を算出する局所特徴検出器が必要である。これらの詳細に関しては以降の節で記述する。また、本提案手法の精度向上のための存在確率の補正、及び探索効率向上のための顔位置の事前トラッキングに関して

も後述する。

### 3.2 局所特徴存在確率を用いた顔の位置及び向きを推定

本節では、式(1)(6)より  $p(c_{i,x,y,z,\theta})$ を用いて  $(x_{est}, y_{est}, z_{est}, \theta_{est}) = \arg \max p_{obj}(c_{1,x,y,z,\theta} = \delta_{1,\theta}, c_{2,x,y,z,\theta} = \delta_{2,\theta}, \dots, c_{n,x,y,z,\theta} = \delta_{n,\theta})$ を推定する手法を提案する。前述した通り  $p_{obj}(x, y, z, \theta) = p_{obj}(c_{1,x,y,z,\theta} = \delta_{1,\theta}, c_{2,x,y,z,\theta} = \delta_{2,\theta}, \dots, c_{n,x,y,z,\theta} = \delta_{n,\theta})$ を直接算出するのは困難である。これに対して、本提案手法では条件付き確率  $p_{obj}(x, y, z, \theta)$ 及び  $p_{obj}(\theta | x, y, z)$ に着目し、これらを算出する事で2つの条件付き確率から  $(x_{est}, y_{est}, z_{est}, \theta_{est})$ の推定を試みる。4つの確率変数  $(x, y, z, \theta)$ の内の一つでも定まった条件下であれば、同時確率  $p_{obj}(x, y, z, \theta)$ を直接算出するよりも容易、且つ何らかの近似を行った際の誤差も減少すると考える。ここで、 $p_{obj}(x, y, z, \theta)$ 及び  $p_{obj}(\theta | x, y, z)$ を次のように算出する。条件付き確率  $p_{obj}(x, y, z, \theta)$ の算出に関しては、向き  $\theta$ が与えられているためどの局所特徴が見えるべきかが既知である。これを基に、見えるべき局所特徴間の従属性から、平均確率として以下のように算出する。

$$p_{obj}(x, y, z, \theta) = \frac{1}{n_{v,\theta}} \sum_{k=1}^{n_v} p(c_{v(k),x,y,z,\theta} = 1) \quad (7)$$

ここで、添字  $v(k)$ は向き  $\theta$ によって定められた見えるべき局所特徴を意味し、 $n_{v,\theta}$ はその総数である。条件付き確率  $p_{obj}(\theta | x, y, z)$ の算出に関しては、任意の向き  $\theta$ に対して有無の条件  $\delta_{i,\theta}$ が等しい局所特徴の群  $j$ を定義する。そして、同一群内では従属性を、群間では独立性を仮定する事で、以下のように算出する。

$$p_{obj}(\theta | x, y, z) \equiv \prod_{j=1}^J \left\{ \frac{1}{n_j} \sum_{k=1}^{n_j} p(c_{s_j(k),x,y,z,\theta} = \delta_{s_j(k),\theta}) \right\} \quad (8)$$

ここで、局所特徴群  $j$ の総数を  $J$ 、各局所特徴群  $j$ に含まれる局所特徴を添字  $s_j(k)$ 、各局所特徴群  $j$ に含まれる局所特徴の総数を  $n_j$ とする。

以上のようにして  $p_{obj}(x, y, z, \theta)$ 及び  $p_{obj}(\theta | x, y, z)$ が算出されたため、最後にこれらの条件付き確率から  $(x_{est}, y_{est}, z_{est}, \theta_{est})$ を推定する。ここで、条件付き確率  $p_{obj}(x, y, z, \theta)$ 及び  $p_{obj}(\theta | x, y, z)$ が同一の  $(x, y, z, \theta)$ に対して最大値をとった場

合に、それは $(x_{est}, y_{est}, z_{est}, \theta_{est})$ である事に着目すれば、 $p_{obj}(x, y, z | \theta)$ 及び $p_{obj}(\theta | x, y, z)$ が最大となる $(x, y, z, \theta)$ を探索すれば良い。ただし、同一の $(x, y, z, \theta)$ に対して両者が最大となる場合は常に保障されているわけではないため、本提案手法では微少閾値 $\delta_p$ を用いて、以下の式で $(x_{est}, y_{est}, z_{est}, \theta_{est})$ を推定する。

$$\begin{cases} (x_{est}, y_{est}, z_{est}, \theta_{est}) \equiv \arg \max_{x, y, z, \theta} p(x, y, z, \theta) \\ \equiv \arg \max_{x', y', z', \theta'} p_{obj}(\theta'_{x, y, z} | x', y', z') \\ p_{obj}(x', y', z' | \theta'_{x, y, z}) \geq \max_{x, y, z | \theta_{x, y, z}} p_{obj}(x, y, z | \theta_{x, y, z}) - \delta_p \\ \theta_{x, y, z} \equiv \arg \max_{\theta} p_{obj}(\theta | x, y, z) \end{cases} \quad (9)$$

式(9)により、 $p_{obj}(x, y, z | \theta)$ が殆ど最大値に近い値を有する範囲内で $p_{obj}(\theta | x, y, z)$ の最大値を求める事になる。以上のようにして、条件付き確率 $p_{obj}(x, y, z | \theta)$ 及び $p_{obj}(\theta | x, y, z)$ から $(x_{est}, y_{est}, z_{est}, \theta_{est})$ を推定する。

### 3.3 局所特徴存在確率の算出のための検出器

局所特徴の存在確率 $p_i(x, y, z)$ を算出するための局所特徴検出器に関しては、[8]を参考に図3で示されるようなHaar-like特徴に似た1ビットのバイナリパターンを使用する。図3において、 $S_{p_j}$ 及び $S_{n_j}$ はそれぞれの領域内の輝度値の総和であり、この大小関係が特徴量となる。参照領域に対するバイナリパターンの種類(図3の(a)または(b))、大きさ、縦横比、位置は全てランダムとする。本研究ではこれらの特徴量に対して[8]のRandom forestを構成したものを検出器として使用する。ここで、ノード数を $N_{node}$ 、ツリー数を $N_{tree}$ とする。また、検出器の学習時には局所特徴が存在する場合とそうでない場合の事前確率が等しくなるように予め正規化を行う。

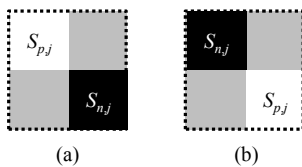


図3 1ビットバイナリパターン

また、[8]に基づき検出器のオンライン学習を行う。入力画像に対して最終的に決定された局所特徴画像を新たなポジティブサンプルとし、アフィン変換を行う事でサンプルを増幅する。一方で、背景と判定された画像領域の中で $p_i(x, y, z)$ の値が高くなった領域を優先的に抽出し、それらをネガティブサンプルとする。これらを新たな学習サンプルとしてオンラインで学習を行う事で環境変動等にロバストになり、事前学習時にアピランスに対する多様性の網羅が不十分であったとしても、オンライン学習によって補う事が期待できる。

### 3.4 局所特徴存在確率計算時の補正

式(3)(4)に基づき局所特徴存在確率 $p(c_{i, x, y, z, \theta})$ を算出する

際に、ノイズやオクルージョン等によって極端に $p(c_{i, x, y, z, \theta})$ が下がると探索精度に悪影響を及ぼす可能性がある。本提案手法では、以下のように $p(c_{i, x, y, z, \theta})$ の下限値を設けた補正を行い、補正後の局所特徴存在確率を $p'(c_{i, x, y, z, \theta})$ と定義する。ここで、 $p_{low}$ は $p(c_{i, x, y, z, \theta})$ の下限値である。

$$p'(c_{i, x, y, z, \theta}) \equiv \max\{p(c_{i, x, y, z, \theta}), p_{low}\} \quad (i=1, 2, \dots, n) \quad (10)$$

### 3.5 顔位置の事前トラッキング

式(9)に基づき顔の位置及び向き $(x_{est}, y_{est}, z_{est}, \theta_{est})$ を探索する際、画像内のすべての領域に対して探索を行わず、事前におおよその位置をトラッキングしておき、その周囲のみを探索する事で探索の精度や効率の向上を試みる。本提案手法では、以下の式(11)のように前局所特徴の存在確率の総和が最大となる状態 $(x_{pre}, y_{pre}, z_{pre}, \theta_{pre})$ を、 $(x_{est}, y_{est}, z_{est}, \theta_{est})$ の探索前の事前のトラッキング結果として定義する。尚、以下の式に示すように事前トラッキングは前回の検出位置 $(x_{est-1}, y_{est-1}, z_{est-1}, \theta_{est-1})$ から定数 $(x_{range}, y_{range}, z_{range})$ を用いて一定範囲内で行う。

$$(x_{pre}, y_{pre}, z_{pre}, \theta_{pre}) \equiv \arg \max_{x_{pre}, y_{pre}, z_{pre}, \theta_{pre}} \sum_{i=1}^n p(c_{i, x, y, z, \theta} = 1) \quad (11)$$

$$\begin{cases} x'_{pre} \in \{x | x_{est-1} - x_{track} \leq x \leq x_{est-1} + x_{track}\} \\ y'_{pre} \in \{y | y_{est-1} - y_{track} \leq y \leq y_{est-1} + y_{track}\} \\ z'_{pre} \in \{z | z_{est-1} - z_{track} \leq z \leq z_{est-1} + z_{track}\} \end{cases} \quad (12)$$

ここで、 $(x_{pre}, y_{pre}, z_{pre})$ から正負の両方向に各定数 $(x_{range}, y_{range}, z_{range})$ だけ拡大した範囲を $(x_{est}, y_{est}, z_{est}, \theta_{est})$ の探索範囲として以下に定義する。ただし、向き $\theta_{pre}$ に関しては画像平面上の探索領域とは独立なので考慮しない。

$$\begin{cases} x \in \{x | x_{pre} - x_{range} \leq x \leq x_{pre} + x_{range}\} \\ y \in \{y | y_{pre} - y_{range} \leq y \leq y_{pre} + y_{range}\} \\ z \in \{z | z_{pre} - z_{range} \leq z \leq z_{pre} + z_{range}\} \end{cases} \quad (13)$$

## 4. 実験結果

図4は、本提案手法を動画像に対して適用した際のフローチャートである。以下の節では、本実験で用いた顔モデル及びドライブレコーダ画像に対する顔向き推定の評価結果に関して記す。

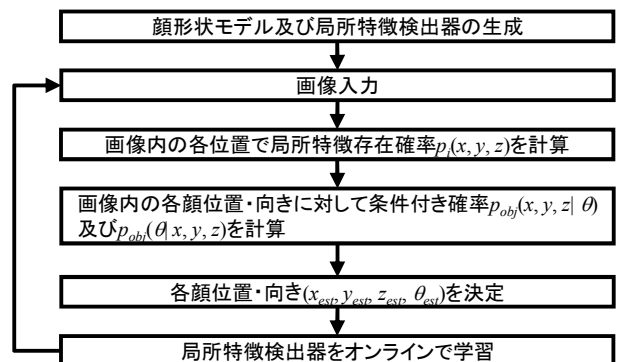


図4 提案手法のフローチャート

#### 4.1 顔形状モデル及び局所特徴検出器の生成

本実験においては、顔形状モデル及び局所特徴検出器は、動画の初期フレームのみを用いて生成する。ここで、初期フレームにおける顔領域及び向きは既知であるとする。また、このようにする事で、事前の学習サンプルによる学習が不必要となるため、本手法を容易に適用可能となる。

顔形状モデルに関しては、図5に示すように  $R$  行  $C$  列の局所特徴を構成し、 $r$  行目及び  $c$  列目の局所特徴  $i$  に対して、添字  $i$  と  $r$  及び  $c$  の対応関係は以下の式の通りとする。

$$(i = cR + r) \quad (14)$$

さらに、顔領域内に局所特徴が収まるように、横方向は等間隔に、縦方向は間隔を可変にして配置する。尚、局所特徴の大きさは全て等しく、顔重心位置に対する局所特徴の位置関係は 3.1 節で述べた通り顔向きに依らず一定とする。

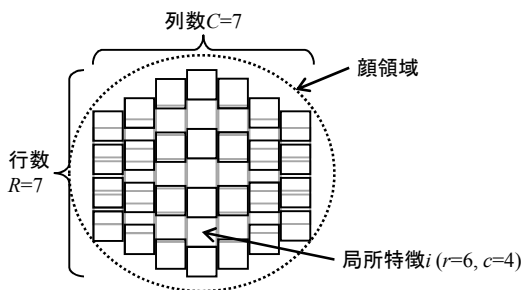


図5 顔形状モデル

また、3.2 節の条件付き確率  $p_{obj}(\theta | x, y, z)$  の算出において式(8)で用いられる局所特徴群  $j$  を定義する必要がある。今回の顔モデルでは、図1のように局所特徴の列方向は顔向きの回転軸と平行であると仮定し、各列をそれぞれ一つの局所特徴群とする。すなわち、 $s_j(k)$  を以下のように定義する。

$$\begin{cases} s_j(k) = jR + k \\ n_j = R \\ J = C \end{cases} \quad (k=1,2,\dots,n_j, j=1,2,\dots,J) \quad (15)$$

最後に、今回の顔モデルにおいて向き  $\theta$  によって定められた  $\delta_{i,\theta}$  の値を表1に示す。尚、向き  $\theta$  は -4 から 4 までの整数で定義し、 $\delta_{i,\theta}$  は局所特徴の同一列内で等しいと仮定しているため、 $\delta_{i,\theta}$  は列  $c$  毎に変化する値である。向き  $\theta$  と列  $c$  に対する  $\delta_{i,\theta}$  を表1に記す。

表1 向き  $\theta$  と列  $c$  に対する  $\delta_{i,\theta}$  の設定値

	$c=1$	$c=2$	$c=3$	$c=4$	$c=5$	$c=6$	$c=7$
$\theta=-4$	0	0	0	0	1	1	1
$\theta=-3$	0	0	0	1	1	1	1
$\theta=-2$	0	0	1	1	1	1	1
$\theta=-1$	0	1	1	1	1	1	1
$\theta=0$	1	1	1	1	1	1	1
$\theta=1$	1	1	1	1	1	1	0

$\theta=2$	1	1	1	1	1	0	0
$\theta=3$	1	1	1	1	0	0	0
$\theta=4$	1	1	1	0	0	0	0

顔形状モデルの生成後、[8]を参考に初期フレーム画像を用いて検出器の学習を行う。生成した顔形状モデルに従って初期フレーム画像から局所特徴の画像領域をポジティブサンプルとして抽出し、アフィン変換を行う事でサンプルを増幅する。また、背景領域からネガティブサンプルを抽出する。これらを学習サンプルとして検出器の学習を行う。

#### 4.2 ドライブレコーダ動画による顔向き推定評価結果

本節では、複数の高齢ドライバーの運転中の車内映像をドライブレコーダで記録した動画の中から、首振り動作が比較的多いと予想される交差点通過前後のシーンを合計37シーン抽出し、本提案手法を用いて左、正面、右の3方向の顔向き推定を行った。ここで、推定された顔向き  $\theta$  が -4 から -2 の時を左向き、 $\theta$  が -1 から 1 の時を正面向き、 $\theta$  が 2 から 4 の時を右向きと定義し、顔領域と眉間位置のラベルから近似的に算出した顔向き角度に対して、 $-30^\circ$  未満の範囲を正面向き、 $-30^\circ$  から  $30^\circ$  の範囲を正面向き、 $30^\circ$  以上の範囲を右向きと定義した。尚、図6はドライブレコーダで記録された画像のサンプルであり、画像サイズは  $720 \times 240$  pixel、カメラのフレームレートは 10Hz である。



図6 ドライブレコーダによる記録画像

表2及び表3は交差点37シーンを平均した場合の顔向き推定結果を示している。尚、表2は検出器をオンラインで学習せず初期フレームのみの情報で学習した場合の推定結果であり、表3はオンラインで学習した場合の検出結果である。3方向いずれの向きに対しても、オンライン学習を行った場合の方が良い推定結果が得られた事から、その有効性が確認できる。表3においては、3方向とも80%前後の正解率が得られた。推定に失敗したシーンの主な例としては、顔の一部が撮像領域外に出てしまった場合や、顔のサイズが著しく変化した場合等が多かったが、これらは閾値処理による未検出判定や、顔の  $z$  方向座標を可変にした探索を行う事で、解決できるものと考えられる。



表2 検出器をオンラインで学習しない場合の交差点37シーンの顔向き推定平均正解率 (%)

		推定値			フレーム数
		左	正面	右	
正解値	左	<b>74.1</b>	4.2	21.7	614
	正面	20.7	<b>70.6</b>	8.7	4184
	右	19.7	2.7	<b>77.6</b>	737

表3 検出器をオンラインで学習した場合の交差点37シーンの顔向き推定平均正解率 (%)

		推定値			フレーム数
		左	正面	右	
正解値	左	<b>78.5</b>	3.1	18.4	614
	正面	9.0	<b>86.0</b>	5.0	4184
	右	15.9	1.6	<b>82.5</b>	737

## 5. 結論

本研究では、低画質画像に対してロバストな顔向き推定を実現するため、局所特徴の有無に着目した顔探索モデルを提案し、その有効性を検証した。また、検出器をオンラインで学習する事により顔向き推定性能が向上し、実環境に対してよりロバストな推定が確認された。今後は、推定に失敗したシーンの課題に対して着手し、本提案手法の性能向上を目指す。

## 参考文献

- 1) E. Murphy-Chutorian and M. M. Trivedi, "Head Pose Estimation in Computer Vision: A Survey", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 31, No. 4, pp. 607-626, 2009.
- 2) T. Cootes, G. Edwards, and C. Taylor, "Active Appearance Models", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 23, No. 6, pp. 681-685, June 2001.
- 3) J. M. Saragih, S. Lucey and J. F. Cohn "Deformable Model Fitting by Regularized Landmark Mean-Shift", *International Journal of Computer Vision*, Vol. 91, Issue. 2, pp.200-215, 2011.
- 4) X. Zhu, D. Ramanan, "Face Detection, Pose Estimation, and Landmark Localization in the Wild", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.2879-2886, 2012.
- 5) J. Xiao, T. Moriyama, T. Kanade and J. Cohn, "Robust Full-Motion Recovery of Head by Dynamic Templates and Re-registration Techniques", *International Journal of Imaging Systems and Technology*, Vol. 13, pp. 85 - 94, 2003.
- 6) J. S. Jang and T Kanade, "Robust 3D Head Tracking by Online Feature Registration", *IEEE International Conference on Automatic Face and Gesture Recognition*, 2008.
- 7) P. F. Felzenszwalb, R. B. Girshick, D. McAllester and D. Ramanan, "Object Detection with Discriminatively Trained Part-Based Models", *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 32, Issue. 9, pp. 1627-1645, 2010.
- 8) Z. Kalal, J. Matas and K. Mikolajczyk, "P-N Learning: Bootstrapping Binary Classifiers by Structural Constraints", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.49-56, 2010.