

撮影環境に基づく確率分布切り替えによる画像向き判定方式

Image orientation detection based on subject distances

高橋 祐介
Yusuke Takahashi

平田恭二十
Kyoji Hirata

1. まえがき

近年、デジタルカメラ、カメラ付き携帯電話の普及により、一般ユーザがデジタル写真を撮影する機会が増えている。カメラを縦に回転して撮影した場合、画像ファイルの向きは画像の内容と一致しないことがあり、現状ではユーザが手動で回転補正を施さなくてはならない。一部の最新デジタルカメラでは重力センサなどにより撮影時にカメラの縦横方向を判定し自動で画像の向きを補正するものがあるが、携帯電話などでは構造的に新たなセンサの導入が困難となるため、ソフトウェア的な自動判定方法が望まれている。[1]では、画像の Low-level 特徴量の SVM 向き判定結果と、意味的特徴(顔・青空)による向き判定結果とを統合し MAP 推定による方向判定により 90%程度の精度を達成している。一方、精度向上のために、視覚的特徴に「草・天井・壁・曇り空」の追加が提案されているが、1%程度の改善にとどまっている。計算コスト増を勘案すると、効率的な精度向上方法が望まれる。

本研究では、Exif 情報から取得可能な低演算量の撮影環境を表す特徴を用いることによる画像向き判定方式の改善提案を行う。

2. 撮影環境に基づく画像の向き判定

デジタルカメラ等で一般人が撮影した画像において、その構図と被写体距離などの撮影環境とは密接な関係があると考えられる。例えば、1 m離れた人物を撮影する場合は、顔が中心の構図となり、3 m離れた人物の場合は、顔の位置は下がり逆に背景が多く写り、空が映る確率も高くなるなどの特徴が挙げられる。このように、被写体距離は画像の構図に大きく関わる特徴を示すと考えられ、構図は Low-level 特徴量および意味的特徴と相関が高い。このため、撮影環境を向き判定の 1 特徴として利用することを考える。しかしながら、撮影環境単体では画像の向きを判定する情報を有していないため、意味的特徴として[1]に示される確率統合手法の枠組みに組み入れることができない。

我々は、撮影環境に応じて Low-level 特徴量や意味的特徴量の出力結果と画像向きとの条件確率が変動するという特徴を基に、図 1 に示すように各方向に対する事後確率計算に用いる条件確率を切り替える指標として撮影環境を用いる方式を提案する。各特徴に基づき判定結果を算出し、撮影環境に基づき各特徴の判定結果に対する出現事前確率を切り替えた上で、式(1)に示すように、最終的な各画像の向き $\omega = \{\text{North, East, West, South}\}$ における確率を算出し、最大確率を示す方向を画像の向きとして出力する。

$$\hat{\omega} = \arg \max P(\omega|S, L, C) = \arg \max P(S, C|\omega)P(L, C|\omega)P(\omega) \quad (1)$$

ただし、L は Low-Level 特徴量、S は意味的特徴、C は撮影環境特徴である。条件確率 $P(S, C|\omega)$ 、 $P(L, C|\omega)$ と事

前確率 $P(\omega)$ は学習データから取得し、CPM (Conditional Probability Matrix) として予め記録されるものとする。CPM は学習画像を撮影環境に基づきカテゴリ分類し、カテゴリ毎に S.L 特徴量の出現確率を算出することにより作成される。判定時では、入力画像の撮影環境のカテゴリに対応する CPM を切り替えて各画像の向きにおける確率を算出することで向き判定を行う。

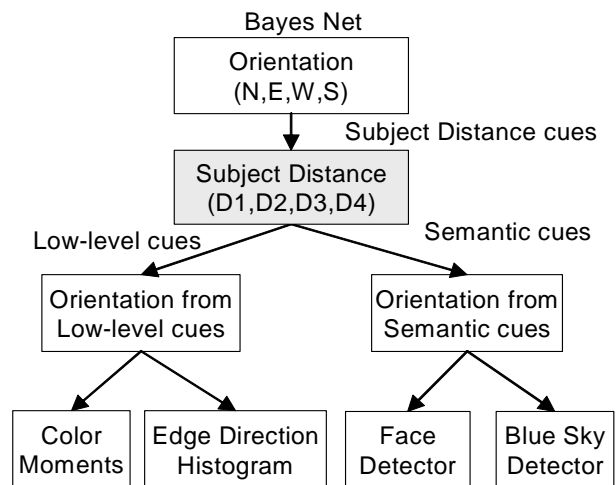


図 1 Low-level および意味的特徴の統合方式

撮影環境として被写体距離(dist)を採用し、本稿では閾値(Th1, Th2, Th3)により 4 つのカテゴリに分類する。

D1: $\text{dist} < \text{Th1}$

D2: $\text{Th1} \leq \text{dist} < \text{Th2}$

D3: $\text{Th2} \leq \text{dist} < \text{Th3}$

D4: $\text{Th3} \leq \text{dist}$

撮影環境情報を取得する方法として、デジタルカメラで撮影された画像に付随する Exif 情報から算出する。Exif とは、JEITA で標準化された画像フォーマットであり、現在多くのデジタルカメラで採用されている。被写体距離はカメラレンズからピントが合っている被写体までの距離として Exif に記録されている特徴量である。

被写体距離は、Exif の被写体距離(SubjectDistance)タグの値を用い、取得できない場合は、過焦点距離を利用する。過焦点距離とは、無限遠が被写界深度にはいるもっとも近い距離のことである。一般のデジタルカメラにおけるオートモードでは、写真の被写体のみならず背景もボケないように絞りなどを自動選択するため、過焦点距離を被写体距離と近似する。過焦点距離は Exif の焦点距離、35mm 換算焦点距離と F 値を用いて式(2)により算出する。過焦点距離 = (焦点距離 × 35mm 換算焦点距離 × 許容錯乱円定数) ÷ (35mm フィルム対角サイズ × F 値) [mm] (2)

ただし、許容錯乱円定数=1300、35mm フィルム対角サイズ=43.27 とする。

3. 実験

3.1 実験条件

実験に用いる画像として、個人が撮影した写真を対象とする。3363枚の個人撮影写真を2つのグループに分け、グループ1(1682枚)を学習用に、グループ2(1681枚)を評価用に利用する。これらの写真は4家族が撮影した写真であり、子供の遊び風景・家族旅行・結婚式・小物撮影など屋内外の様々なジャンルの写真で構成される。これら全ての写真にはExif情報が付与されており、被写体距離(過焦点距離)を算出できるものとする。従来法[1](Conventional)と提案手法(Proposed)において画像のLow-Level特徴量にはColor MomentとEdge Direction Histogram、意味的特徴として顔と青空を用いる。向き判定の正解率として、評価用の入力画像が特定の出現率に基づいた方向を向いていると仮定した際の平均正答率を算出し、正解率と定義する。

画像向きの事前確率と評価用画像の向き(N, E, W, S)の出現確率は以下の3ケースを設定する。

- ・ 出現確率ケース1 N=0.72, E=0.14, W=0.12, S=0.02
- ・ 出現確率ケース2 N=0.25, E=0.25, W=0.25, S=0.25
- ・ 出現確率ケース3 N=0.81, E=0.18, W=0.01, S=0.0

ケース1は、[1]の評価実験で用いられている出現確率であり、中間的な出現確率である。ケース2は、事前知識が全く無い場合(ランダム)での出現確率である。ケース3は、1家族が撮影した個人撮影写真(1315枚)における向きの出現確率であり、NとEに大きな偏りがある場合の出現確率である。

予め、学習画像をその撮影環境に応じて4カテゴリ分類し、各カテゴリに対してCPMを作成し記録しておき、判定時は入力画像の撮影環境に応じたカテゴリのCPMを用いて画像の向きの確率を算出を行う。この分類手法について、実験により効果を確認する。

カテゴリ分類に用いる閾値を決定するために、画像向きの事前確率と入力画像の各向き(N, E, W, S)の分布確率が共にケース1である場合において、正解率が最大となる閾値を求め、以後の実験における閾値として設定する。求めた閾値は、 $Th_dist1=2.4$, $Th_dist2=3.9$, $Th_dist3=12.0$ である。

3.2 正解率

画像向き判定時の事前確率と評価用画像の向き出現確率のそれぞれにケース1～3の確率を用いて、従来法および提案手法により正解率を算出した結果を図2に示す。

提案手法では、設定と実験における出現確率の事前知識が無い場合では、2%程度精度改善が見られる。その他の場合においても0.5%～1.5%の改善が見られる。これは[1]で報告されている「草」「天井・壁」「曇り空」などの特徴量抽出の追加と同程度の精度向上を実現していることが判る。

3.3 処理速度

提案手法における画像1枚あたりの向き判定の処理時間は20秒程度であり、1サンプル画像における主要なプロセスでの処理時間を表1に示す。用いたPCはPentium4 3.6GHz(1.0GB RAM)である。

処理時間のほとんどがSVMに費やされているが、SVM判別器を外部のプログラム[2]で実行しているためにファイルI/Oなどにおける時間ロスによるものが大部分である。

SVM以外の処理時間では、Low-level特徴量、顔特徴量、青空特徴量の抽出処理に数百msかかっていることがわかる。一方、Exif情報取得を含んだ被写体距離算出は無視できる程度の処理コストであることがわかる。つまり、「草」「天井・壁」「曇り空」などの特徴量抽出の代わりに、被写体距離算出を導入することにより大幅な処理コストが低減できることを意味する。

4. むすび

デジタルカメラで撮影した個人撮影画像の向き判定方式を提案する。撮影環境ごとの視覚特徴の出現特性を考慮するため、複数の視覚特徴量に基づき向きを判定する際に、Exifから低演算量で取得可能な撮影環境に応じて確率分布を切り替えることを特徴とする。撮影環境特徴として被写体距離を用い、画像向きに関するさまざまな事前確率条件での実験において73.2～91.4%の正解率であることを示した。Low-Level特徴量と意味的特徴量による判定を基にMAP推定する従来法と比較し0.5%～2%程度の精度向上を確認した。これは複雑な視覚的特徴を追加した場合と同程度の改善効果である。さらに、撮影環境情報を取得する処理コストは無視できる程の低演算であることを確認した。

参考文献

- [1] Luo Jiebo, M. Boutell, Automatic Image Orientation Detection via Confidence-Based Integration of Low-Level and Semantic Cues, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, No. 5, pp. 715-726, 2005.
 [2] T. Joachims, "Making large-Scale SVM Learning Practical. Advances in Kernel Methods - Support Vector Learning," B. Schölkopf and C. Burges and A. Smola (ed.), MIT-Press, 1999.

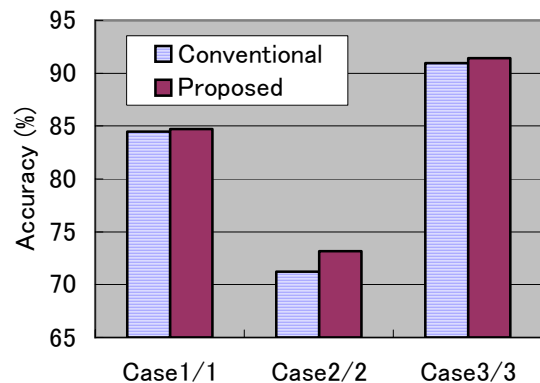


図2 確率設定における正解率

表1 主要プロセスの処理時間

プロセス	時間 (ms)
Low-level 特徴量抽出	312
SVM 判別	20890
顔抽出による判定	656
青空抽出による判定	125
被写体距離算出(Exif取得を含む)	0.0017