

形状情報量の分布に基づくシーンの撮影の自動化

Automatic Production of Video Images Using Spatiotemporal Configuration of Surface

山下 拓也[†]
Takuya Yamashita青木 茂樹[‡]
Shigeki Aoki大西 正輝[‡]
Masaki Onishi福永 邦雄[†]
Kunio Fukunaga

1. まえがき

近年、TV放送のデジタル化や通信技術の発達により多種多様な映像の配信が可能になり、遠隔講義やTV会議といったシステムが実用化されている。しかし、講義や会議の様子を撮影するためにはカメラマンの経費が必要となることから撮影の自動化を目的とした研究が行われている。また、撮影の自動化によってTV番組の制作費が削減できれば、これまで撮影に費用がかげられなかったため放送できなかった番組の放送が可能になる。

撮影を自動化する場合、視点を自動的に決定することで、視聴者が見たい領域を映像化する必要がある。遠隔講義を対象とした撮影の自動化を目指した研究[1]では、映像中の近傍領域における色情報の複雑さから各画素が持つ情報量を定義し、情報を多く発生している領域を受講者が注目する領域であると考えている。一方、物体には色情報以外にも形状情報が存在し、物体の見え方と3次元形状の知覚のしやすさとは密接な関係があることから、物体の形状に注目することで3次元形状を知覚しやすい視点での映像化が可能になると考えられる。

本研究では、物体の形状に関する特徴を定量的に定義し、3次元形状を効果的に表現できる視点の決定手法を提案する。さらに、物体の動きを時空間における形状として捉えることで、形状だけではなく動きの特徴も考慮した視点の決定手法に拡張する。

2. 形状情報量に基づく視点の探索

人間が物体の形状を知覚するとき、物体を構成する面の形状特徴や面の占める領域および接続関係などに注目すると考えられ、このような特徴を多く捉えることができる視点が物体形状を知覚しやすい視点である。本研究では、車のような人工物が動くシーンを撮影対象としており、シーンに登場する物体は滑らかな面で構成されていると仮定する。この仮定の下では、凹凸の激しい面や面の境界部分である頂点や稜線部分に物体の形状に関する特徴が強く現れる。このような形状に関する特徴の強度を情報量として定量的に定義し、多くの情報を受け取ることができる視点を探索することで、視点を決定する。

2.1 形状に関する情報量

法線ベクトルは面の形状と連動して変化するため、法線ベクトルの分布を用いた面の特徴表現が可能で、面の境界部分は法線ベクトルが不連続に変化する部分として表現できる。本研究では、凹凸の激しい面や面の境界部分に形状に関する情報が多く分布すると考えており、このような部分における法線ベクトルの変化が激しいことを考慮して、形状に関する情報量 $I_s(x)$ を定義する。

[†]大阪府立大学大学院 工学研究科, Graduate School of Engineering, Osaka Prefecture University

[‡]理化学研究所 バイオ・ミメティックコントロール研究センター, Bio-Mimetic Control Research Center, RIKEN

本手法で扱う空間は、2値ボクセルの集合で表現され、式(1)のように、座標 $x = (x, y, z)^T$ に対応するボクセルが物体を構成する場合は1、それ以外の場合は0の値を持つものとする。

$$f(x) = \begin{cases} 1 & \text{a part of object} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

座標 x におけるボクセルが持つ法線ベクトル $N(x)$ (以下 N) は、式(2)のように $n=3$ とした3次元のガウス関数 $G(x)$ を用いて求める。ただし、 \otimes は重畳積分、 ∇ はベクトル微分演算子を表す。

$$N(x) = (\nabla G \otimes f)(x) \quad (2)$$

次に、物体は滑らかな面で構成されているという仮定から、近傍領域における法線ベクトルの変化は小さく、同じ方向を向いている可能性が高いと言える。このことから、近傍領域における法線ベクトルの分布から座標 x における法線ベクトル N の発生確率 $P(N)$ を $n=3$ とした3変量正規分布の確率密度関数を用いて式(3)で定義する。ただし、 μ , Σ は、座標 x を中心とした半径 r ボクセルの近傍領域における法線ベクトルの平均と分散共分散行列を表す。

$$P(N) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{(N - \mu)^T \Sigma^{-1} (N - \mu)}{2}\right) \quad (3)$$

座標 x における形状に関する情報量 $I_s(x)$ は、Shannonの情報理論から法線ベクトル N の発生確率 $P(N)$ を用いて式(4)で表す。

$$I_s(x) = -\log_2 P(N) \quad (4)$$

2.2 動きに関する情報量

人間は物体の動きを観察するとき、物理法則から物体は同じ動きを続けるものと予測し、予測が困難な動きを行った瞬間を特に注目すると考えられる。本研究では、物体の動きを時空間における形状と捉えることで、動きの特徴を表現する。物体が等速度や等加速度で動いている場合、時空間における形状は平面や放物面となり、面の形状で動きの特徴表現が可能で、動きが変化した瞬間は面の境界として特徴表現が可能である。注目すべき動きに対応する形状は、2.1で考えた特徴的な形状と一致するので、形状に関する情報量 $I_s(x)$ を時間軸方向に拡張することで、動きに関する情報量 $I_{st}(x)$ を定義する。

まず、2.1で扱ったボクセル空間を拡張してボクセル時空間とする。ボクセル時空間の座標を $x = (x, y, z, t)^T$ で表し、座標 x における法線ベクトル N は、式(2)で $n=4$ とした4次元のガウス関数 $G(x)$ を用いて導出する。次に、法線ベクトル N の発生確率 $P(N)$ は、2.1

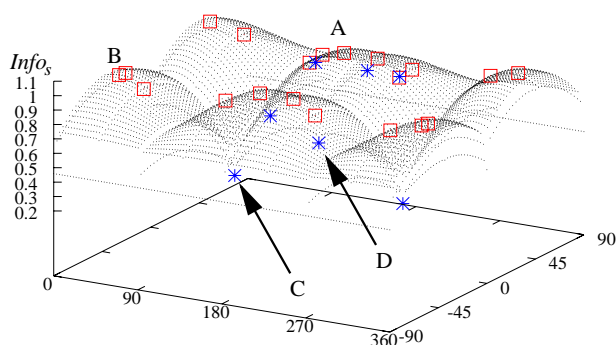
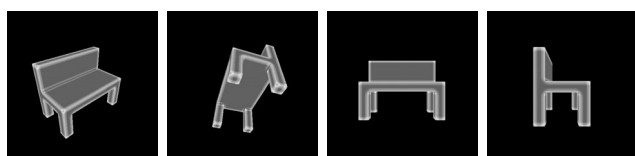


図 1: 形状に関する情報量のヒストグラム



A:極大 B:最大 C:最小 D:極小

図 2: 情報量が最大 (極大) と最小 (極小) の視点

と同様に座標 x を中心とした半径 r ボクセルの時空間の近傍領域を考え、式 (3) で $n = 4$ とした 4 変量正規分布の確率密度関数で表す。物体の動きに関する情報量 $I_{st}(x)$ は、式 (4) と同様にして定義する。

2.3 視点の決定

視点の決定は、2.1 2.2 で得られた情報量 $I_s(x)$, $I_{st}(x)$ (以下 $I(x)$) を多く受け取ることができる視点 $V = (x, y, z)^T$ を探索することで行う。このとき、視点 V で受け取ることができる情報量 $Info_s(V)$, $Info_{st}(V)$ (以下 $Info(V)$) は、式 (5) のように座標 x_i の情報量 $I(x_i)$ に座標 x_i における法線ベクトル N_i と視点の方向 $(V - x_i)$ のなす角 θ_i に応じた重み $\cos \theta_i$ を加えた加重平均で求める。ただし、重み $\cos \theta_i$ が負となる点は視点 V から見えない点であることから無視し、残りの m 個の可視点に対する加重平均を考える。また、動きを考慮した情報量 $I_{st}(x)$ の場合には、法線ベクトル $N_i = (x, y, z, t)^T$ の時間軸成分 t は無視する。

$$Info(V) = \frac{1}{m} \sum_{i=1}^m \begin{cases} I(x_i) \cos \theta_i & (\theta_i < \pi/2) \\ 0 & (\theta_i \geq \pi/2) \end{cases} \quad (5)$$

3. 実験

3.1 形状に関する情報量に基づく視点

動きを考慮しない場合の情報量について椅子の形状をした物体のボクセルデータ $f(x)$ を与え、物体から十分離れた視点 V で受け取ることができる形状に関する情報量 $Info_s(V)$ を求めた。横軸および縦軸には物体に対する視点の方向を極座標系で表現した角度 θ, ϕ 、高さ軸に情報量 $Info_s(V)$ をとったヒストグラムを作成すると、図 1 のようになった。このヒストグラムにおいて、受けとれる情報量 $Info_s(V)$ が最大 (極大) および最小 (極小) となる視点から物体を見ると、図 2 のようになった。図

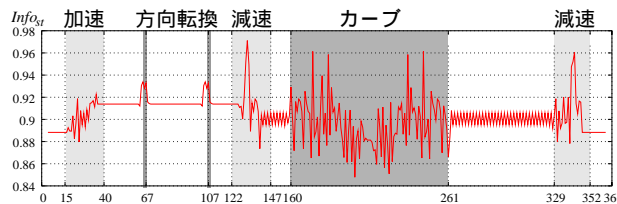


図 3: 動きに関する情報量の変化

2 において輝度値は各ボクセルが持つ情報量を表しており、最大 (極大) となる視点では、物体の特徴的な形状を多く含んでいた。

次に、形状に関する情報量 $Info_s(V)$ と 3 次元形状の知覚のしやすさについて物体像エントロピー H を用いて評価する。物体像エントロピー H は物体の可視面の数 N と可視面 i が物体全体に占める面積の割合 r_i から式 (6) で与えられる一次元の値で、心理実験により 3 次元形状の知覚のしやすさを表せることが示されている [2]。

$$H = - \sum_{i=1}^N r_i \log_2 r_i \quad (6)$$

視点 V において受け取ることができる情報量 $Info_s(V)$ と物体像エントロピー H との相関係数は、椅子の場合では 0.44 であったが、立方体のような単純な形状の場合では 0.82 となり、強い相関があることから、情報量が多く受け取れる視点ほど、物体の 3 次元形状を知覚しやすい視点であることがわかった。

3.2 動きに関する情報量に基づく視点

加速や減速およびカーブや方向転換といった動きをする立方体に対して動きに関する情報量 I_{st} を求めた結果、受けとれる情報量 $Info_{st}$ の最大値の変化は図 3 のようになった。この結果から、静止状態よりも動いている状態、特に、速度が変化している瞬間に情報量が高くなり、動きの特徴を反映した情報量が得られた。また、動きの方向に対して垂直な方向、カーブに対して内側と外側で情報量が高くなっており、物体の動きの特徴を効果的に表現できる視点の決定ができた。

4. まとめ

本研究では、物体の形状および、動きに関する情報量を定義し、シーンを効果的に表現できる視点を決定する手法を提案した。今後は、テクスチャから得られる色に関する情報量と統合し、舞台上で行われる演劇などの複雑なシーンの撮影の自動化に応用することが考えられる。また、視聴者が自由に視点を設定できる放送システムにおいて視点決定サポートの働きができると考えられる。

参考文献

- [1] 大西正輝, 泉 正夫, 福永邦雄, “情報発生量の分布に基づく遠隔講義撮影の自動化,” 信学論 (D-II), vol.J82-D-II, no.10, pp.1590-1597, Oct. 1999 .
- [2] 石川尋代, 佐藤幸男, “多面体認識における心理量と物体像エントロピーの関係について,” 信学論 (D-II), vol.J86-D-II, no.1, pp.111-118, Jan. 2003 .