景観理解における普遍性と特殊性について
On Generic Property And Specific Property of Scene Understanding

笹木 美樹男 †
Mikio Sasaki

1. はじめに

近年，DNN による画像理解が主流となり，多くの高速動作例や人工知能を彷彿とさせる研究が報告されている．だが，いまだに人間に匹敵する理解は成されていない．ニューラルネットや人工知能のブームは過去にも何度も繰り返され，スパイラル的に着実に進歩は遂げてはいるものの近年の過剰な期待には静観せざるを得ない．一方で，IEEE の PAMI などでは着実に研究が推進され[1][2]，必ずしも景観理解では DNN 最優先ではない．その背景には DNN の学習能力が過去の経験 0 からスタートして無駄な労力に費やされている実情もある．全世界の DNN がつながっていて過去の失敗を共有できるならばこんな無駄は起こりえない．

本稿ではこの観点から，景観理解における普遍性と特殊性を多元的に考察し，集合知により認識率と認識そのものを改善する手法を提案する．

2. 景観理解における普遍性と特殊性

2.1 普遍性

近年，人の流れや出現の予測において学習事象の普遍性に着目する研究がなされている [3]．これは画像理解やその他の予測・推定問題にも拡張でき，画像では集合知による概念獲得の結果が普遍性と考えられる．だが，これさえも観測側の視点と個々のシーンの判別容易性に左右される．

2.2 特殊性

一方で，従来の画像理解では問題が設定される前から問題解決に不要な特殊性を排除せず，いきなりすべての属性を学習させるため，非常に高度な学習機能が要求されてきた．それが DNN のもてはやされる理由である．ところがこの不要な特殊性を除去すれば，景観理解の問題は線形の枠組みでも十分高速に解ける．

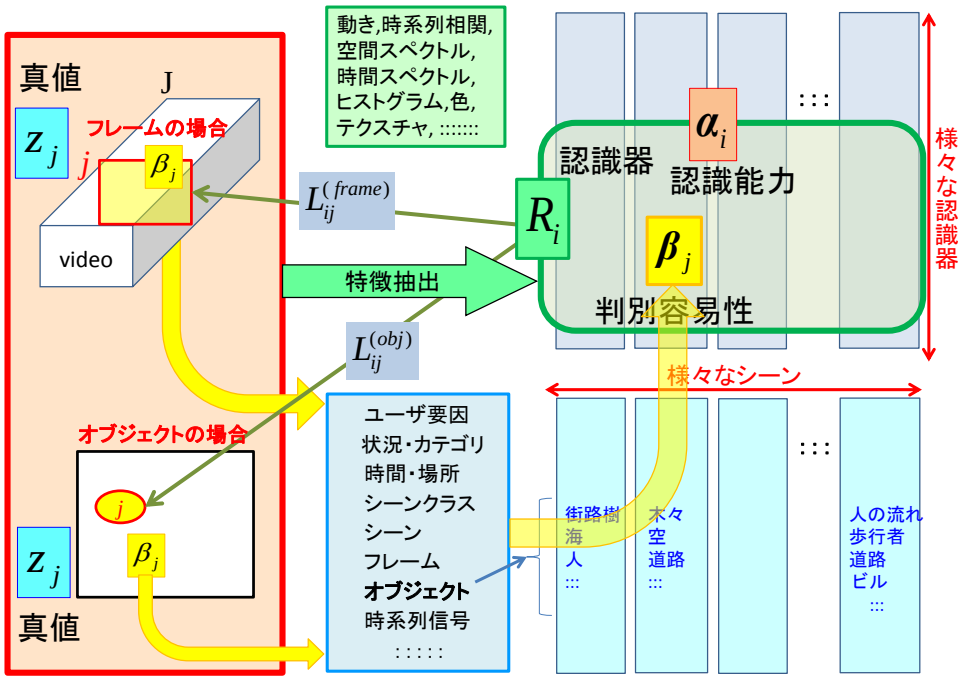

図 1　映像と認識器の関係
Fig. 1　Relation between images and recognizers.

2.3 普遍性変換

空間情報科学における予測・データ同化の分野では問題次元の最適化において注目すべき特殊性のみ残し，それ以外の特殊性を段階的に捨象する普遍性変換という概念が提案された[4]．これはまったく異なる時空の事象を残された特殊性（すなわち普遍性）のみで関連付け，学習対象にすることができる．これにより予測に必要な時空に観測対象がない場合でも過去の事例や類似する事象から予測を実施できる．この考えは画像理解における解の推定（たとえば歩行者予測やテレビの次の場面の予測など）において圧倒的に確信度の推定効率をあげるであろう．

† Music Scene Research
（個人研究のため公的所属は割愛）

普遍性変換で無駄な特殊性を捨象し，共通の普遍性を有する他の画像（あるいは事象）に重要な特殊性を付加することで新たな推論を誘発することも可能であろう．たとえば日本の街並みから学習した結果をスイスに適用しても基本オブジェクトを識別できる例が示されている[5].

3. 社会への浸透性

すでに社会で得られたルールや過去の試行で獲得した結果を共有し，常に認識器群が進化していく手段を考える．

特別な性能が要求されない限り，メディアや通信に適した景観理解の手法が普及には有利である．そこで，MPEG ベースの景観認識[6]を起点とし，社会への浸透性に力点を置く．すでに社会で得られたルールや手段をインターネットを介して適用できれば，認識器は 0 からスタートして無駄な学習プロセスをたどる必要はない.

4. 景観理解と集合知

集合知による確信度の共有について論ずる．初期の設定以外は教師なし学習を前提としている．

4.1 GLAD から集合知へ

複数の認識器を巡回的に協調させるという手法をより高度なものにした例として，EM アルゴリズムを導入した GLAD (Generative model of Labels, Abilities, and Difficulties)手法[7]がある．この手法を参考にして認識に関連する属性要因が拡張され，MBV に適用された[5][8] .

いま，K 個のブロック画像 $BLK(k),(k=1,...,K)$ の集合を考える．各ブロック画像は N 個の対象カテゴリの 1 つに属するとする．

このとき，ブロック画像 $BLK(k)$ のオブジェクトインデックス z_k を決定したい．観測したインデックスは主としてブロック画像の判別容易性，真のオブジェクトインデックス，認識器の認識能力の３つに依存する．

いま，ブロック画像 $BLK(k)$ の判別容易性を次式で表現する：

$$\beta_k \in [0,+\infty) \tag{1}$$

また，画像 $BLK(k)$ の真のインデックスを z_k とする．また，認識器 L_i の認識能力のパラメータを次式で表現する：

$$\alpha_i \in (-\infty,+\infty) \tag{2}$$

α_i は認識対象画像 $BLK(k)$ の 8 階層の属性で定義できる：ユーザが求める認識タスクとユーザの嗜好性（Level-7），シーン撮影時の状況と分類カテゴリ（Level-6），シーンの撮影時刻と場所（Level-5），シーンクラス（Level-4），シーン（Level-3），フレーム（Level-2），オブジェクト（Level-1），$BLK(k)$ の特徴量と時系列（Level-0）

L_i が $BLK(k)$ に与えたインデックスを l_{ik} とし，それが真のインデックスに等しい確率を次式のモデルで表現する．

$$p(l_{ik}=z_k|\alpha_i,\beta_k)=\frac{1}{1+e^{-\alpha_i\beta_k}} \tag{3}$$

このモデルのもとで獲得したインデックスで正しいものに関する対数尤度は認識能力と判別容易性の双一次関数として次式で表現される．

$$\log\frac{p(l_{ik}=z_k)}{1-p(l_{ik}=z_k)}=\alpha_i\beta_k \tag{4}$$

z_k,α_i,β_k は既知の先験的分布でサンプルされる．これらは式(3)によって観測インデックスを決定する．観測インデックスの集合 $L=\{l_{ik}\}$ が与えられたとき，求められるタスクは $Z=\{z_k\}$, α_i,β_k について最も尤度の高い値を同時に推定することになる．そこで，これらの最尤推定を行うために期待値最大化（EM）手法を適用する：

4.2 認識能力の適応化

集合知によりシーンに応じた認識能力のベクトルが適応化され，改善される．

4.3 判別容易性の適応化

判別容易性に関しても集合知により事前確率の分布が適応化し，改善される．これは特に様々なシーンでの適用結果のフィードバックによる．

5. おわりに

画像理解における普遍性に着目することで，より人間に近い結果をめざす枠組みを提案した．普遍性変換の具体的な獲得を DNN で行い，末端の認識処理は蓄積や通信に適した従来手法（たとえば MBV）で構成することで適材適所の景観理解が実現できると考えられる．さらに確信度の予測推定に拡張すればより汎用的な応用が期待される．

参考文献

[1] S. Hong, J. Choi, J. Feyereisl, B. Han, and L. S. Davis, ”Joint Image Clustering and Labeling by Matrix Factorization”, *IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI)*, vol.38, no.7, pp.1411-1424, Jul. 2016.

[2] Z. Akata, Z. Harchaoui, and C. Schmid, “Label-Embedding for Image Classification”, *IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI)*, vol.38, no.7, pp.1425-1438, Jul. 2016.

[3] M. Sasaki, Y. Sekimoto, “An Investigation on Generic Properties of People Flow Database”, *IEICE Technical Report*, ITS2013-69 (2014-3).

[4] 笹木， 関本，「人の流れにおける普遍性と特殊性について」，*CSIS DAYS 2014*. B10 (2014-10).

[5] M. Sasaki (2014). The Estimated Truth will Evolve on Neuro-ITS, *ITS World Congress 2014*, Detroit, Sep. 2014.

[6] 笹木美樹男：“MPEG 映像に適した景観認識手法の提案”,画像電子学会誌, vol.38, no.6, pp.890-899 (2009-11).

[7] B. Zhong, et al., “Visual Tracking via Weakly Supervised Learning from Multiple Imperfect Oracles”, *CVPR2010*.

[8] M. Sasaki, “A Proposal for Neuro-ITS over the Connected Vehicles Network”, *SSI2014*.