第 4 章

調査に基づく修正のタグ付け手法

4.1　はじめに

本章では前章の調査を踏まえて、タグの設計と自動タグ付けの手法を提案する。

4.2　調査から考える修正のタグ設計

まず、調査を踏まえたタグ付けの利点を以下に 3 つ示し、それに応じたタグを考える。

1. 修正が多発している部分を分析し、同様の修正の発生を抑制する
   具体例として、担当者によらずシルエットの修正が多発している作品などでは、今後の制作でアニメータがシルエットを描く際に注意喚起することができ、以降の修正の削減に繋げることができます。このためのタグとして被写体の動きや姿勢、修正対象の部位のタグを用意する。
2. アニメータの得意不得意や修正者の癖を把握し、仕事の割り振りに活かす
   得意不得意を分析することで、アニメータにとって得意なカットを割り振りすることができ、制作進行を円滑に進めることができる。このためのタグとして、原画担当者、修正担当者、被写体の表情、動きと姿勢、修正部位を用意する。
3. 正確な仕事速度の見積もり
   例えば手のアップのカットがあったとして、どのくらいのスピードで修正をこなせるかを把握すれば、似たカットの仕事の依頼を出す際に、スケジュールの見積もりを高精度で出すことができます。このためのタグは被写体の映り方、カメラアングル、被写体の画面占有率を用意する。

これらのタグを整理すると以下の 7 つになる。

- 原画担当者
- 修正担当者
- 被写体の映り方・カメラアングル
- 被写体の表情
- 被写体の動き・姿勢
- 修正対象の部位
- 画面占有率

それぞれのタグの粒度や切り口についてここで述べる。

4.3　修正の自動タグ付けの流れ

タグ付けの流れを図 4.1 に示す。まず、渕上らのシステムから上がってきた成果物である彩色済み画像を VLM と物体検出に推論させる。その後、修正素材と対応する彩色済み画像に対し、物体検出の箱を重ね、修正の書き込みがあるかどうか判定し、画面占有率と修正対象の部位について得る。

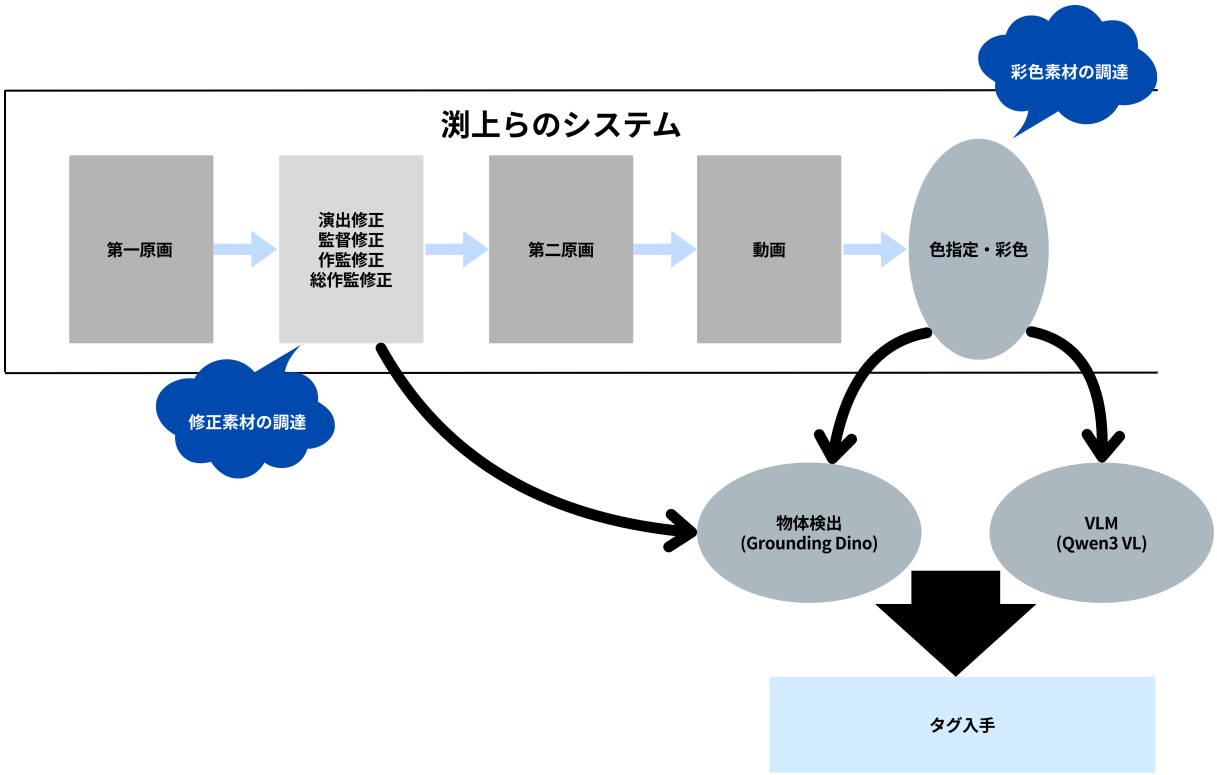

図 4.1　渕上らのシステムの情報伝達の流れ

なお、前述したタグにおいて、

- 原画担当者
- 修正担当者

については既存の渕上らのシステムで蓄積することができるため、その他のタグの入手を考える。

4.4　VLM による彩色画像のタグ付け

具体的な VLM によるタグ付けについて述べる。第二章で述べた通り、モデルとしては Qwen3 VL を採用する。

VLM に出力させたいタグは以下の通りである。

- 被写体の映り方・カメラアングル
- 被写体の表情
- 被写体の動き・姿勢

このためのプロンプトとして、以下の文章を採用する。

【目的】

入力画像（単体 or 複数フレーム）から、以下の 5 カテゴリだけを抽出して厳密に JSON のを出力してください。

※ 5 カテゴリ以外（outfit/attributes/colors/objects/style/background/text 等）は出力しない。説明文・コメントも出さない。

【抽出カテゴリ（固定語彙の例つき)】

1. 被写体の写り方（cropping）
   - 候補（例）：close_up（顔のアップ）, bust（胸上/上半身）, half（腰上/半身）, full（全身）
2. カメラ角度・視点（camera_angle）
   - 候補（例）：front（正面）, back（背面）, profile（側面/横顔）, low（仰角）, high（俯瞰）, three_quarter（3/4 視）
3. 姿勢（pose）
   - 例：立つ/座る/歩く/走る/指さす/手を振る/抱える/持つ/かがむ/跳ぶ 等（短い動詞句、日本語で OK）
4. 表情（expression）
   - 例：笑顔/怒り/悲しみ/驚き/困惑/無表情/緊張 等
5. 相互関係（interactions）※被写体が 2 人以上のとき必須
   - 例：conversation（会話）, handshake（握手）, gaze_contact（見つめ合い）, physical_contact（接触）, group_action（集団動作）

【人数に応じた出力ルール】

- 被写体（人物/キャラ/生物など）の人数が 0：空の subjects は出さず、代わりに summary のみ返す（interactions も出さない)。
- 1 人：subjects を 1 要素で出力。interactions はキーごと省略。
- 2 人以上：subjects に人数ぶんの要素を作成し、interactions を必ず出力して participants に subjects[].id を参照させる。

【信頼度（confidence)】

- 0.0〜1.0 の小数（小数第 2〜3 位程度）で付与。
- cropping / camera_angle はオブジェクト形式で値と信頼度を持たせる。
- pose / expression は配列（複数候補可、各要素に confidence)。
- interactions は各関係ごとに confidence を持たせる。

【不明時の扱い】

- 不明な要素は出力しない (キーごと省略)。空配列は避ける。
- 推測しすぎず、判断困難な場合は該当キーを省略する。

【最終出力フォーマット（JSON のみ/UTF-8/改行可/キー順任意)】

```
{
  "summary": "カット全体の要約（1〜2 文）",
  "subjects": [
    {
      "id": 1,
      "cropping": { "value": "bust", "confidence": 0.00 },
      "camera_angle": { "value": "front", "confidence": 0.00 },
      "pose": [ { "description": "手を振る", "confidence": 0.00 } ],
      "expression": [ { "description": "笑顔", "confidence": 0.00 } ]
    }
    // 2 人以上なら id=2,3,... を追加
  ],
  // 2 人以上のときだけ出力
  "interactions": [
    {
      "type": "conversation|handshake|gaze_contact|physical_contact|group_action",
      "participants": [1, 2],
      "description": "簡潔な要約（例：向かい合って会話）",
      "confidence": 0.00
    }
  ]
}
```

(注)上記以外のキー(outfit/attributes/colors/objects/style/background/text_in_image/nsfw/uncertainties 等）は出力しない。

このように出力させた理由は、

4.5 物体検出による彩色画像のタグ検出

4.6 修正部位のタグ付け

4.7 おわりに