

目撃証言にもとづく正面顔似顔絵スケッチの生成: CLISOTS

CLISOTS: CLIP based Semantic Oriented Testimony to portrait Sketch

○ 阿見翼[†], 小森政嗣[†] ([†]大阪電気通信大学 情報通信工学部)

概要

- 「似顔絵捜査」はリアルな写真と比べて目撃者の記憶を歪める可能性が低いが似顔絵捜査官の育成は容易ではない
- CLIPを用いたスケッチ生成手法 (CLIPDraw およびCLIPasso) をベースに, 目撃証言から正面顔スケッチ画像を生成するシステム (CLISOTS) を構築した
- 意味的損失と幾何学的損失を考慮して正面顔スケッチ生成する

提案手法

- 言語的に表現された顔特徴をベジェ曲線でスケッチ化する
- (1) 初期スケッチ生成, (2) 証言からのプロンプト生成と意味的損失の計算, (3) 幾何学的損失の計算, (4) 似顔絵スケッチ生成, (5) スケッチの更新の5ステップから構成される

(1) 初期スケッチ生成

- 平均顔画像から MediaPipeにより顔の主要な標識点座標を取得し, その周辺領域を着目領域として46本のベジェ曲線で初期スケッチを生成する

(2) 意味的損失の算出

- 目撃証言では曖昧で感覚的な表現(例えば「真面目そうな」)が用いられるため, GPT-4も用いてCLIPへの入力に適したプロンプトに変換する
- このプロンプトおよびスケッチ画像をそれぞれViT-B/32 CLIPモデルでエンコードし, これらの出力のコサイン類似度を求める
- これを意味的損失とする

$$L_{semantic} = dist(CLIP(I), CLIP(R(\{s_i\}_{i=1}^n)))$$

$$dist(x, y) = 1 - \frac{x \cdot y}{\|x\| \cdot \|y\|}$$

(3) 幾何学的損失の算出

- 生成されたスケッチが正面顔から大幅に乖離しないようにするため幾何学的損失も考慮した
- 幾何学的損失はResNet101 CLIPモデルに初期スケッチ画像と生成されたスケッチをそれぞれ入力し, このときの中間層のL2ノルムにより求める

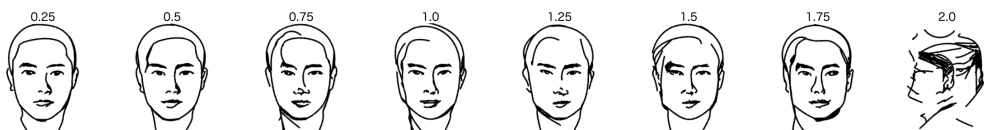
$$L_{geometric} = \sum_l \|CLIP_l(I) - CLIP_l(R(\{s_i\}_{i=1}^n))\|_2^2$$

(4) 似顔絵スケッチ生成

- 意味的損失と幾何学的損失を重み付けした和を損失関数とし, これを最小化することで, 似顔絵生成に用いるパラメータを最適化する
- このパラメータを元にラスタライザーによりスケッチを生成する

$$\min_{\{s_i\}_{i=1}^n} L_{geometric} + w_s \cdot L_{semantic}$$

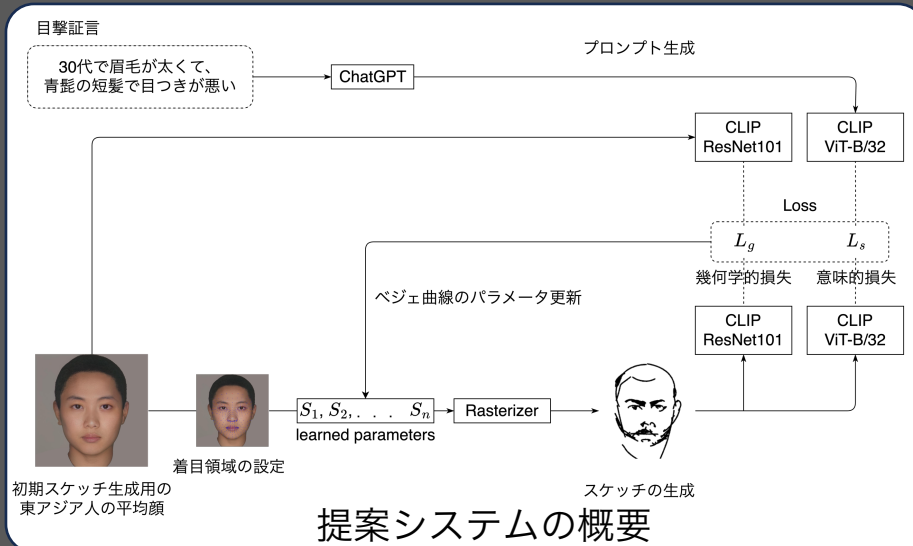
- 意味的損失の重み w_s は予備調査に基づき1.75とした



意味的損失 w_s と生成された第1回目生成スケッチ画像の関係

(5) スケッチの更新

- 生成されたスケッチを初期スケッチ画像の代わりに用いることで, スケッチを更新する
- 更新時の意味的損失の重み w_s は0.75とした



提案システムの概要

