
Mobile Crowd Photographing における類似度に基づく キャッシュ置換

鄧 千宜 上山 憲昭

立命館大学 情報理工学部

2023.03.07

Mobile Crowd Photographing

- Mobile Crowd Photographing (MCP)
 - スマホで撮影された写真を様々なサービスに活用
- 大学のキャンパス内でゴミ箱の中の写真を収集し、リサイクルボックスの設置場所を決定
- スマホの写真を用いて屋内のナビゲーション
- 災害発生時に、スマホで撮影された被災状況に関する写真を収集し、避難者の誘導や救助活動に活用

MCPの課題

- ネットワーク上に数千万枚以上の写真が存在
 - 画像は冗長性が高い
 - ユーザの要求画像と完全に一致する画像を配信する必要性は低い

- キャッシュサーバから配信される機会が多い
 - CDNやエッジキャッシュなど
 - キャッシュ容量が有限
 - キャッシュ置換法(LRUやFIFOなど)が必要

研究目的

- ユーザに低遅延で配信
- ユーザのニーズをできるだけ満たしたい

- キャッシュ内に存在する画像の中で、要求画像との類似度が最大の画像を配信

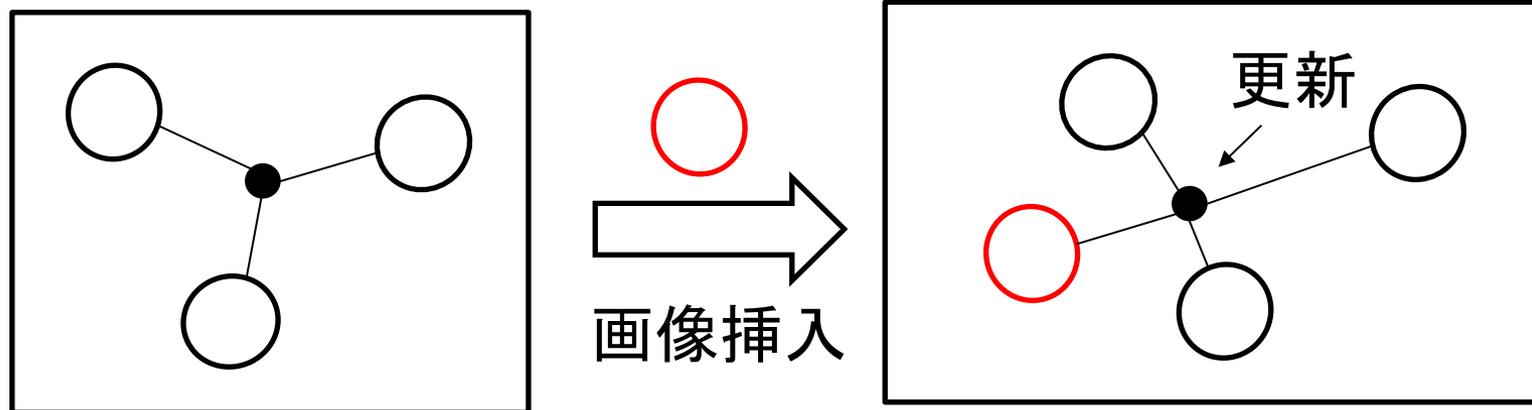
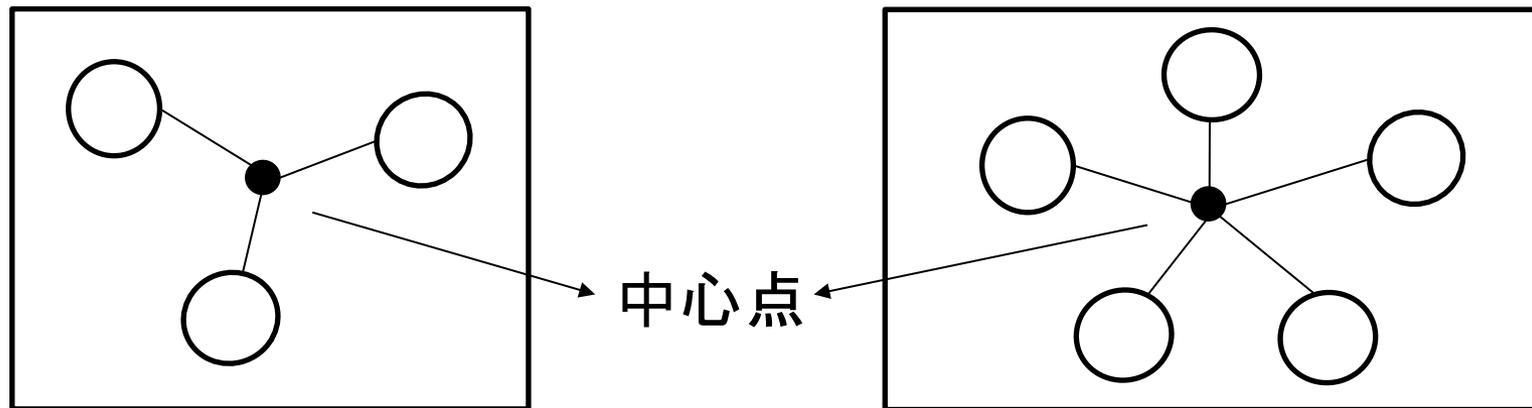
- **提案キャッシュ置換法:** 他のキャッシュ画像との類似度が高く、冗長性が高い画像を優先的に削除
 - 類似度計算に要する時間を抑制
 - 配信画像と要求画像との差異を向上
 - 画像の冗長性に加え、画像の人気度も考慮

提案方式の処理の流れ

- 画像の類似性を高速に判定するため、**SIFT** (scale-invariant feature transform)アルゴリズムで画像の特徴点を抽出
 - 画像データを全体的に利用する必要性なし
- 特徴量を, Bloom Filterで一つベクトルに変換
 - 特徴点数が異なる多様な画像を統一的に処理可能
- E2LSH (exact euclidean locality sensitive hashing)アルゴリズムで類似度に基づき画像を複数のグループに分類
 - 画像の類似度判定に要する時間を短縮
- キャッシュ内の画像を, **冗長性と人気度の両方を考慮した総コスト**によりキャッシュ置換

E2LSHによる画像のグループ化

- 特徴量が近い画像は同じグループに入れる可能性が高い



提案キャッシュ置換法

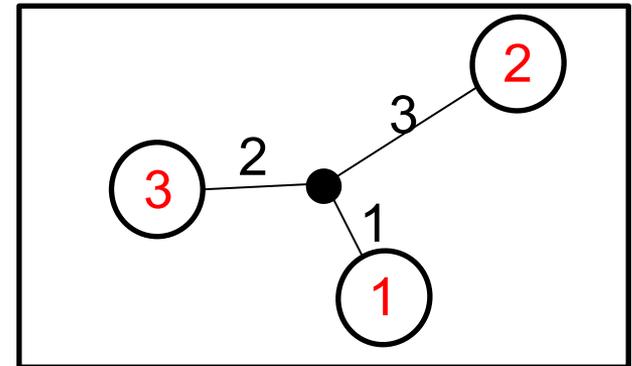
- キャッシュ内の画像の中から，要求画像と特徴量の距離が最小の画像をユーザに配信

- 距離に関するコスト(C_d)
 - 類似性が高い画像を削除
 - m 個の画像中， k 番目に冗長な

$$\text{画像データ: } C_d = \frac{m-k}{m-1}$$

- 人気度に関するコスト(C_p)
 - m 個の画像中， j 番目に人気の

$$\text{画像データ: } C_p = \frac{j-1}{m-1}$$



	1	2	3
C_p	0	0.5	1
C_d	1	0	0.5
ω	0.5	0.5	0.5
C_{total}	0.5	0.25	0.75

- キャッシュ内の**画像の総コスト**によってキャッシュ置換
 - $C_{total} = C_p * \omega + C_d * (1 - \omega)$ ※ ω は人気度の重み
 - C_{total} が最大の画像を削除 $0 \leq \omega \leq 1$

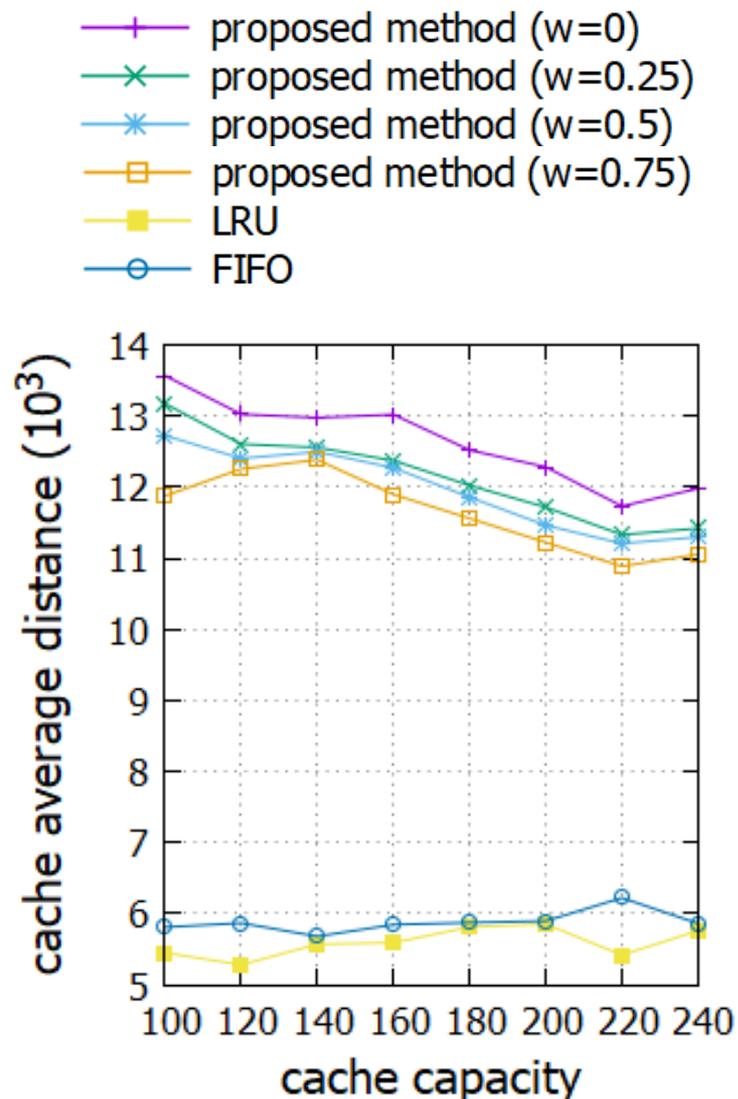
性能評価条件

- ユーザから1万回の要求
- Zipf分布(パラメータ:0.8)でランダムに要求画像を選択
- 三つの画像データセットを各々, 用いた結果の平均で評価
- LRUとFIFOと性能を比較

- 評価尺度
 - キャッシュ平均距離 (Cache average distance)
 - 平均キャッシュヒット距離 (Cache hit distance)

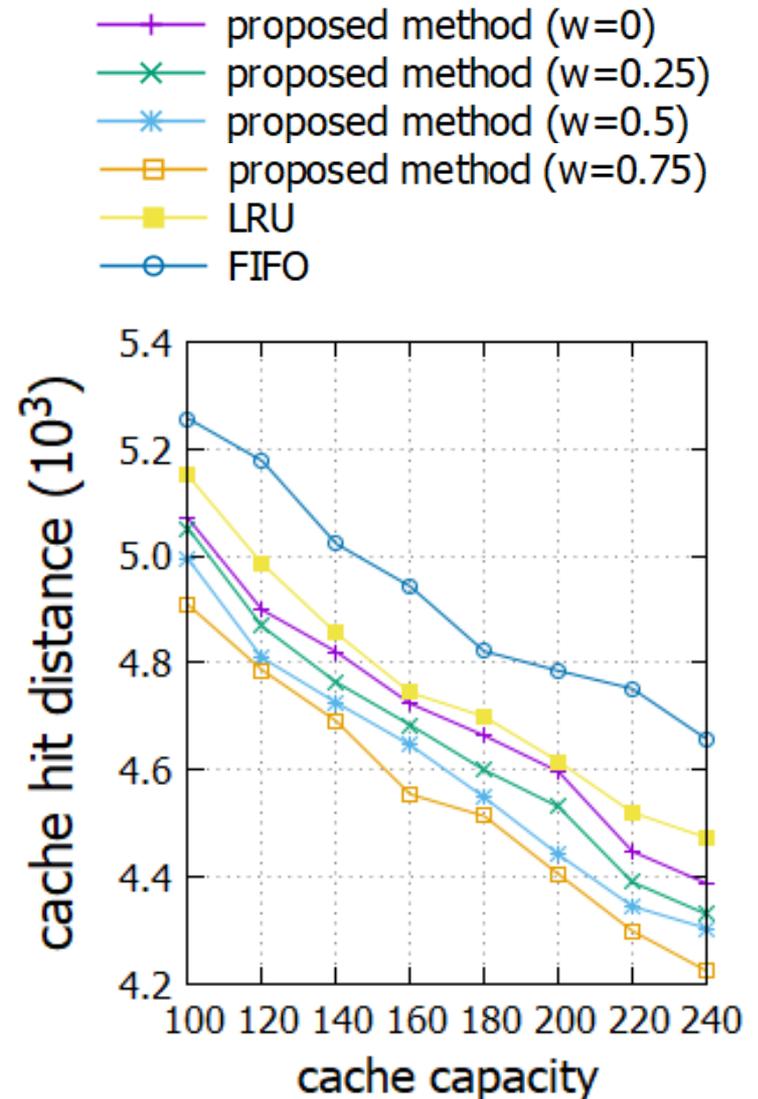
評価結果: キャッシュ平均距離

- キャッシュ平均距離
(cache average distance)
 - キャッシュ内の全て画像から中心点までの距離の平均値
- 類似度が高い画像を優先削除
 - キャッシュ内の多様性が高い
 - キャッシュ平均距離が高い
- ω の増加 \Rightarrow 距離の重み減少
 - 冗長な画像が残る可能性増加
 - キャッシュ平均距離減少



評価結果:平均キャッシュヒット距離

- 平均キャッシュヒット距離 (cache hit distance)
 - ユーザーに送信する画像と要求画像の距離の平均値
- キャッシュ内の多様性の向上
 - 平均キャッシュヒット距離減少
- ω の増加 \Rightarrow ヒットの機会増加
 - 平均キャッシュヒット距離減少



まとめ

- MCPでは膨大な数の画像がネットワーク上に存在
⇒ ユーザの要求画像に完全に一致しなくても、類似度の高い画像を配信すれば十分
- MCPのキャッシュ内の画像の冗長性を低減することで、配信画像と要求画像の差異を低減するキャッシュ置換法を提案
- 画像を複数のグループに分類することで、画像の類似度の判別の計算処理量を抑制

- 今後の取り組み
 - 機械学習など、他の方法でグループ化する方法を検討
 - 画像の特徴点の抽出と判別の効果を評価

ご清聴ありがとうございました