一般社団法人 電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

HCGシンポジウム2014
HUMAN COMMUNICATION GROUP SYMPOSIUM 2014
HCG2014-C-8-3

集合的個人視点映像の自動編集に関する基礎検討

- 屋外グループ活動の効果的な記録・閲覧を目指して -

近藤　一晃† 　小幡佳奈子† 　中村　裕一†

† 京都大学 学術情報メディアセンター　〒606–8501 京都市左京区吉田本町
E-mail: †{kondo,obata,yuichi}@ccm.media.kyoto-u.ac.jp

あらまし　場を共有する複数人の主観映像の集まりである集合的個人視点映像は，単に並べて同時再生しても状況を把握するための良いメディアとはならない．本報告では発話者にフレーミングするような映像切り替え手法を提案し，その有効性および不足している要素について検討する．

キーワード　グループ活動記録, 集合的個人視点映像, 自動編集

Fundamental Study for Editing Collective First-person-view Videos

- toward effective record and review for outdoor group activities -

Kazuaki KONDO†, Kanako OBATA†, and Yuichi NAKAMURA†

† Academic Center for Computing and Media Studies, Kyoto University
Yoshidahonmachi, Sakyo-ku, Kyoto-shi, Kyoto, 606–8501 Japan
E-mail: †{kondo,obata,yuichi}@ccm.media.kyoto-u.ac.jp

Abstract　Collaborative First-person-view videos are captured by the camera attached on multiple persons in the same field. Simultaneous replay of CFPV videos are inappropriate way to review situations or iteractions. We proposed a basic video switching method to make it obvious what factors are important in editing CFPV videos.

Key words　Group activity records, Collaborative First-person-view Videos, and Automatic editing

1. はじめに

複数の人間が互いにインタラクションを行いながら活動・作業を進める場面は，コミュニケーションや社会学といった観点から重要な意味を持っている．我々は屋外におけるグループ活動・体験活動などを想定し，それらを効果的に記録する仕組みとして参加者全員にカメラを取りつける方法 (集合的個人視点映像) を提案している [1]．しかし自由に動くカメラにより撮影された個人視点映像は互いに一貫性を欠くため，単に並べて同時再生しても良いメディアとはならない．最も簡単な方法は各時刻において最適な映像に切り替えながら再生するもので，着席対話 [2], [3], スポーツシーン [5], 技能コンテンツ作成 [4] などを対象とした手法が提案されている．これら従来研究と集合的個人視点映像における条件の違いは，撮影者が被撮影対象でもあること，カメラが固定されておらず自由に動き回ることである．つまり写したい対象がいつも撮影されているとは限らず，また良いアングルで写っていることも保証されない．そこで本報告では Shen らが提案した Quality-of-View の枠組み [6] を集合的個人視点映像の切り替え編集に用いることで，従来手法の適用可能性および未対応な課題について検討を行う．

2. 集合的個人視点映像の自動編集手法

2.1 問題設定

集合的個人視点映像は個人視点映像の集まり $\mathbf{I_c}(t), c = 1, 2, ...N$ で表現される．ここで c はカメラ ID, N は参加者数を示す．これらを入力として，各時刻でどの映像を提示するのかを表す系列 $c_s(t) = \{1, 2, \cdots, N\}$ を状況をもっともよく表すように決定する．ただし「状況をもっともよく表す映像系列」の数学的定義は 1 つではない．本報告では人物間のインタラクションで重要かつ基本的な要素である対話に注目し，その様子を伝えたい状況とする．通常，対話には発話者・受け手が存在し，場合によっては話題となっている物体・シーン・人物など第三のファクタが存在するが，ここでは発話者の様子を伝えることのみを扱う．すなわち発話者がもっとも良く写っている映像を選択するように $c_s(t)$ を決定する．

2.2 映像の評価値導出

上記問題設定に基づいた映像選択を行うために各映像における発話者の写り方を点数化して用いる．まず発話区間 $e(t, T) =$

$\{0,1\}$ を人物 T 毎に抽出する．発話者が複数の場合もあるため発話者数で正規化した $\mathbf{E}(t) = \left[\frac{e(t,1)}{\Sigma_{T=1}^{N} e(t,T)} \cdots \frac{e(t,N)}{\Sigma_{T=1}^{N} e(t,T)} \right]^T$ をイベント系列とする．発話者がいない時刻では参加者全てが等しく提示対象となりうるとし，$E = \left[\frac{1}{N} \cdots \frac{1}{N} \right]^T$ とする．

次に人物の写り方評価値 $g(t,T,c) = (0,1)$ を導出する．これは時刻 t において参加者 T が映像 c にどれだけ良いアングルで写り込んでいるかを示している．評価値計算には対象人物が写っている矩形領域の面積・位置・縦横比・速度の 4 項目を点数化した $g_s, g_p, g_r, g_v = (0,1)$ を用いた．縦横比が大きかったり速度が早すぎる場合は，面積や位置に関わらず対象人物を提示する映像として不適とみなしたいため

$$g(t,T,c) = \frac{1}{2}\{g_s(t,T,c) + g_p(t,T,c)\} g_r(t,T,c) g_v(t,T,c) b(t,c)$$

のように設計した．ここで $b(t,c) = \{0,1\}$ は個人視点映像 $\mathbf{I_c}(t)$ 自体が提示に適しているかどうかを示す値であり，激しい揺れを含んでいる・全体的に暗すぎたり明るすぎたりする場合に 0 となる．各映像の点数 $S(t,c) = (0,1)$ は，イベント系列 $\mathbf{E}(t)$ および $g(t,T,c)$ を行列の形で表した $\mathbf{G}(t,c) = [\, g(t,1,c) \cdots g(t,N,c) \,]^T$ を用いて $s(t,c) = \mathbf{E}(t)^T \mathbf{G}(t,c)$ のように導出される．

2.3 制約付き評価値最大化による提示映像決定

各時刻を独立に考えれば，$c_s(t) = argmax\ s(t,c)$ のようにして最適な提示映像を選択することができるが，短時間のショット切り替えが連続すると視聴者の負担が増大するため，各ショットの長さ (同一映像が選択されている時間) に下限値 m_{th} を設ける．すなわち $c_s(t)$ が少なくとも m_{th} 秒連続して同じ値をとるという条件の下で $c_s(t) = argmin\ \Sigma_{t=0}^{t_{max}} |s(t, c_s(t)) - max\ s(t,c)|$ なる組み合わせ最適化問題を解く．しかし，この最適化を効率的に行うことは難しいこと，準最適解でも実用上問題ないことから本報告では **Algorithm 1** に示すような近似アルゴリズムを用いる．これは向こう m_{th} 間で積算した評価値が最も大きい映像を選択し，さらに以降も他のどの映像よりも評価値が高い限り同じ映像を選択し続けるものである．

Algorithm 1 Calculate $c_s(t)$ from $s(t,c)$

```
t_p ← 0
while t_p ≦ t_max do
  for c = 1 to c ≦ N do
    s_tmp(c) ← 0
    for t = t_p to t ≦ t_p + f_th do
      s_tmp(c) ← s_tmp(c) + s(t,c)
    end for
  end for
  for t = t_p to t ≦ t_p + f_th do
    c_s(t) ← argmax s_tmp(c)
  end for
  t_p ← t_p + f_th
  while s(t_p, c_s(t_p − 1)) = max s(t_p, c), c = 1, 2, ···, N do
    c_s(t_p) ← c_s(t_p − 1)
    t_p ← t_p + 1
  end while
end while
```

3. 実　　験

仮の体験活動として「大学キャンパス内の自転車駐輪問題を調査・解決するワークショップ」を実施・記録した．ワークショップでは 6 名の大学生・大学職員および進行・誘導役のファシリテータ 1 名が一団となって学内の駐輪状況を調査してまわった．実験にはファシリテータおよび 4 名の学生参加者により構成される集合的個人視点映像を入力として用いた．なお発話イベント・対象人物矩形・提示不適区間の抽出は手作業で行い，ショットの長さの下限値は $m_{th} = 3sec$ とした．提案手法を用いて編集することで 5 本の個人視点映像を同時再生するのに比べて誰がどのように話しているのかがわかりやすいことが確認されたが，その一方で以下のような問題点も見られた．

映像中のどの部分に注目すればよいか分かりづらい　従来のコンテンツ映像では視聴者に注目して欲しい対象を中央に大きく配置するなどの工夫がなされているが，個人視点映像ではそのような効果的なアングルは保証されない．発話者の写り方が小さい・映像の端に写っている・映像の切り替え前後で位置が移動する，といった場面では，注目すべき場所がわかりづらいことが確認された．これには吹き出し等を追加表示することが効果的と考えている．

発話者がどの個人視点映像にも写っていないときの提示映像　視聴者は「発話者がどの個人視点映像にも写っていない」ことを知らないため，対象ショットをどう見れば良いのか迷ってしまう傾向にある．前項と同様に吹き出しを追加することで，例えば対象人物は画面外にいることや，提示映像は対象人物の個人視点映像であることなどを伝えられるのではないかと考える．

この他にも，誰も発話していないときの提示映像の見方に困る，人物の写り方の評価に方向（正面・横・後ろなど），部分 (顔・バストショット・下半身など)，複数人物の配置といった指標が足りない，などの問題も明らかとなった．単に映像を切り替えるだけでなく，例えば文脈に合わせて映画の技法などをうまく援用することなどを含めて解決法を検討していきたい．

文　　献

[1] 近藤一晃, 高瀬恵三郎, 小泉敬寛, 中村裕一, 森　幹彦, 喜多一, “個人視点映像を用いた気づき体験の回想と整理支援　〜フィールド調査における問題発見を通じて 〜”, 電子情報通信学会:PRMU 研究会報告, PRMU2010-128, pp. 13-18, 2010.

[2] 尾形涼, 中村裕一, “制約充足と最適化による映像編集モデル”, 電子情報通信学会論文誌, Vol. J87-DII, NO. 12, pp. 2221-2230, 2004.

[3] 竹前嘉修, 大塚和弘, 武川直樹, “対面の複数人対話を撮影対象とした対話参加者の視線に基づく映像切替え方法とその効果”, 情報処理学会論文誌, Vol. 46, No. 7, pp. 1752-1767, 2005.

[4] 東海彰吾, 間瀬健二, 藤井俊彰, 川本哲也, “多視点の釘付け視聴による技能コンテンツ制作・提示”, 画像の認識・理解シンポジウム MIRU2008 予稿集, pp. 428-433, 2008.

[5] 永井有希, 丸谷宜史, 梶田将司, 間瀬健二, “プレーに着目したスポーツ多視点映像の評価尺度”, 情報処理学会:EC 研究会技報, 2011-EC-19, pp. 1-6, 2011.

[6] C. Shen, C. Zhang, and S. S. Fels., “A MultiCamera Surveillance System that Estimates Quality-of-View measurement”, In Proc. of IEEE International Conference on Image Processing(ICIP), pp. 193-196, 2007.