

映像の感性情報表現を助けるビデオカメラ

御手洗紘子^{†1} 吉高淳夫^{†1}

ユーザの表現意図を適切に表したノンバーバル情報を伝達するためには、適切なカメラワークと編集が必要のため、質の高いビデオコンテンツを制作するのは困難である。撮影経験の少ないユーザに対する予備実験のもと、漸進的インタラクティブ撮影モデルを提案した。本論文では、映像が表す雰囲気などのような感性情報を撮影技法に関係づけるモデルを基にユーザの撮影知識や技術を補うシステムについて述べている。ノンバーバル情報として特定の雰囲気を選択すると、システムが撮影中の映像とカメラアングルやズーム速度などカメラ操作を解析しユーザを支援する。これまでに開発したシステムはカメラ部分が大きく、重量があったため、小型化を図った。評価実験より、システムは適切にユーザの意図を反映するような撮影支援を行うことが示された。提案システムにより撮影知識や技術の獲得を支援することができ、ユーザがショット撮影をより適切かつ効果的に行えるようになる。

Video Camera for Better Expression of Affective Information

HIROKO MITARAI^{†1} ATSUO YOSHITAKA^{†1}

In order to communicate nonverbal information which appropriately express user's expressional intentions, appropriate camerawork and editing are required; therefore authoring quality video content is difficult. Based on the preliminary experiment on the inexperienced users, we proposed the incremental shooting model. In this paper, based on the model which relates affective information such as atmosphere to shooting techniques, the system, which compensates for techniques and shooting knowledge of the user, is described. When the user selects specific atmosphere as nonverbal information, the system assists the user by analyzing current shooting image and camera manipulations such as camera angles and zoom speed. Since previously developed system contained a large camera component and was heavy, we downsized the system. System evaluation indicated that the system could assist the user to reflect his/her intention appropriately. By using our proposed system, the system assists the user to acquire shooting knowledge and techniques; it enables the user to shoot the video more appropriately and effectively.

1. はじめに

我々は会って会話をしたり、文通や電話で会話するなど様々な方法でコミュニケーションを図ってきた。映像は視覚情報を含み、豊富な言語・非言語コンテンツを伝えることのできるもっとも効果的なメディアの一つである。

1895年にルミエール兄弟によって最初の映像媒体である映画が生まれて1世紀以上が経過した[1]。映画は当初、事実の記録のみに用いられた。しかし1902年、フランス人映画監督のジョルジュ・メリエスによる世界初のSFフィクション映画「月世界旅行(Le Voyage dans la Lune)」が生まれてから、映画という媒体は我々の想像を目に見える形にする媒体としての可能性を広げ続けている。

近年、ビデオ制作は低価格かつ高品質のビデオカメラとノンリニア編集を可能にする編集ソフトの入ったコンピュータを持つ一般のユーザに普及している。それと同時に、YouTubeなどのプラットフォームを用いて誰でも映像情報を共有することができる。

効果的な映像の利用は人と人とのコミュニケーションを促進する一方で、その適切な制作を行うことは必ずしも容易ではないことから、一般のビデオカメラユーザが映像制作者として非言語情報を適切に伝えることは難しい。

映像制作には大きく分けて技術的問題・表現的問題とい

う2種類の問題がある。技術的問題の例としては、例えば走り回る元気な子供を捉えようとしてビデオカメラを動かすことなどが挙げられる。この場合、元気な雰囲気を表現することを意図していたとしても、撮影された映像は手振れと過剰なカメラワークが原因で見ると堪えない映像になる。

表現的問題の例としては、被写体を遠すぎる位置から撮影する場合がある。その場面において被写体の表情が重要であった場合、鑑賞者は被写体の表情が見られないだけでなく、被写体に物理的に近いという感覚を得られないため、そのショットの雰囲気は簡単に損なわれてしまう。その理由として、感情表現を行うためには、基本的にはクローズアップショットが必要であることが言える。映画監督や映像理論の研究者は適切な映像制作手法について研究を続けており、たとえば「映画の文法[2]」にその成果を見ることができる。

本論文では、撮影経験の少ないユーザの映像制作知識を補うシステムを提案する。ユーザが「力強さ」のような感性情報を入力すると、システムは同時にビデオカメラから撮影中の映像を取得し、分析を行う。また、加速度センサおよびズームセンサを用いてユーザの状態を検知する。システムはこれらの情報を用いて、選択された雰囲気を表現できるようにユーザをガイドする。このシステムを使用することで、鑑賞者にも理解可能な非言語感性情報を表現することができる。

^{†1} 北陸先端科学技術大学院大学 情報科学研究科
School of Information Science, Japan Advanced Institute of Science and Technology

2. 映画の文法

映像を撮影する際は、画面中の被写体の大きさ（ショットサイズ）、被写体の映っている角度（カメラアングル）、カメラ操作（カメラワーク）とそれらにより表現される感性情報を考慮する必要がある。鑑賞者は自分がカメラを通して実際にその場所を見ているような感覚を得ることから、カメラ操作は鑑賞者の印象を左右する[3]。図1は、ショットサイズ、カメラアングル、カメラワークの表現する感性情報について[2][4]で述べられていることを参考にまとめたものである。

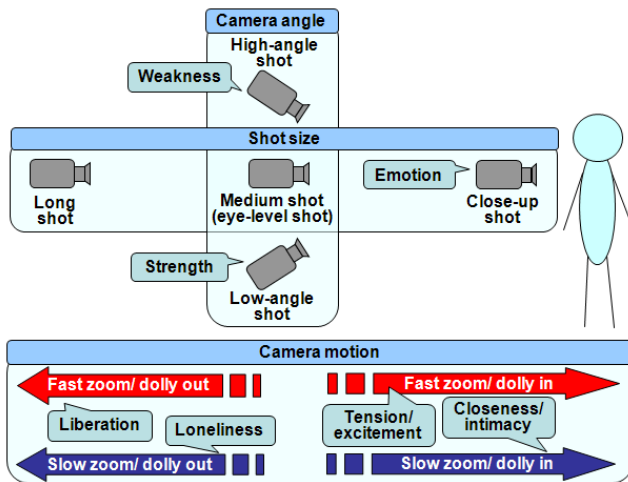


図1 カメラワークと感性情報の関係

ロングショットで被写体を撮影すると、被写体は小さく見え、どの部分がショットの重要な部分なのか分りにくくなる。ミディアムショットは、普段人と人が会話をするときの距離を想起させるため、特別な印象を与えることはないが、クローズアップで撮影する場合は、人物の顔により接している印象を与えるため、被写体の感情を強調することができる[5]。

3. 提案モデル

3.1 撮影経験の少ないユーザーへの予備実験

一般ユーザーの撮影における感性情報表現能力を測定するため、前項の図1を基に感情表現、力強さ、弱さ、共感・親密さ、孤独感、解放感の7つの感性語を選択し10名の様々なビデオ制作経験をもつ大学院生の被験者に提示し、事前知識なしにカメラの撮影方法のみによってそれら一つずつ表現するように指示を与える実験を行った。被写体は椅子に座っているのみで表情などは変わらない。実験の詳細は[6]に記述されている。

実験の結果、撮影経験があることが必ずしも適切な撮影につながることを、また被験者がうまく撮影できたと思っても実際に適切に撮影できているとは限らないことがわかった。このことから、撮影経験を増やすだけでは効果的でなく、また適切に撮影できているか正しく判断出来

ないことがあるため撮影時に支援を行う必要があることが分かった。

3.2 撮影支援のためのインタラクションモデル

図2はビデオカメラにおける初期のモデル、現在のモデル、本研究で提案する撮影支援モデルにおけるインタラクションの相違点を表している。



図2 インタラクションモデルの変遷

初期の撮影スタイルでは、ズームなどの機能を用いて、ユーザーがシステムを一時的に操作するのみだった。現在は光学・デジタル処理による色彩、焦点や露出の自動補正などの技術的支援に加え、水平線や三分割構図線などのガイドラインによりカメラ操作の決定に有用な指針を提示する機能を備えている。しかしながら、未だどのように撮影すべきかについて具体的な提案をしてはいない。そこで、表現意図に適切に沿うショット撮影を可能にするため、ユーザーの主体的な撮影をシステムが支える形でユーザーを密に支援する漸進的インタラクションモデルを考案し、以下のようなシステムデザインを行った(図3)。

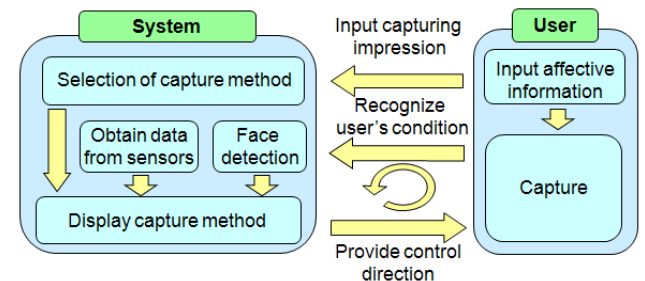


図3 漸進的インタラクションモデル

ユーザーが「感情表現」や「力強さ」など自分がどのような印象や雰囲気では撮影したいのかを入力すると、システムは選択された印象をもとにどのような撮影方法が最適か判断する。ユーザーがショットを撮影すると、システムは、得られる動画情報およびカメラ操作情報から映像を分析する。また、映像の分析結果とユーザーにより選択された印象からどのように撮影すればよいかユーザーに教示する。ユーザーは指示に従って撮影を行う。これをショット撮影が完了するまで繰り返す。

これにより、ユーザーは現在の撮影環境と撮影状況を考慮に入れた適切な撮影方法をその場その場で学ぶことができるため、ユーザーは適切なショットサイズ、カメラ角度とカメラワークを使用した適切な撮影経験と知識を得られるようになる。最終的にはプロの持つ撮影技術について知る機

会のない経験の少ないユーザのメディアリテラシーを向上させることに貢献すると考えられる。

4. システム実装

4.1 ハードウェア要件

適切な撮影法を教示する撮影支援システムの実現には、次の構成要素が必要になる。

- ユーザに撮影された映像を取得するビデオキャプチャデバイス
- ユーザの撮影状態を検知するセンサ
- 映像から顔検出を行う画像処理装置
- 表現したい雰囲気を入力するためのマンマシンインタフェース

4.2 ハードウェアデザイン

図 4 は EmoCap システムの外観である。

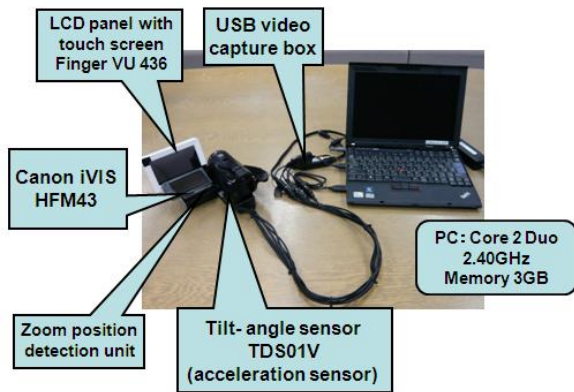


図 4 EmoCap システムの外観

ビデオカメラ Canon iVIS HFM43 の映像出力が RCA ビデオキャプチャユニットを通してノートパソコンに接続されている。カメラ下部にはカメラの傾きを取得するための加速度センサ (図 5) が USB インタフェースを介して接続されている。ズーム検出ユニット (図 6) はタッチセンサ付液晶パネル下部にあり、ビデオカメラに接続されている。



図 5 USB 接続の加速度センサ



図 6 ズーム検出ユニット

図中のノートパソコンの CPU は Core 2 Duo 2.40GHz でメモリは 3GB 搭載している。このパソコンは撮影映像の処理、ズーム速度およびカメラ角度検出による撮影状態の認識、これら情報の評価に基づくガイダンスの表示を行っている。システムを小型化するため、小型のタッチセンサ付

液晶パネルと一般的なビデオカメラを採用した。また、ズーム位置検出ユニットは小型のカメラからズーム操作を検出するため SOP 部品を用いた実装により小型化した。

図 7 はシステム構成である。ユーザがカメラ操作を行うと、ノート PC が同時に 4 種類のデータを取得する。

- ビデオカメラから取得された映像
- 液晶パネルのタッチスクリーンからの入力
- 加速度センサからのカメラ角度
- ズーム位置センサユニットからのズーム位置

ノート PC はこれらのデータを処理し、液晶パネルを通して対応するガイダンスを表示する。使用したビデオカメラのズームスイッチは可変抵抗を使用しているため、ズーム操作を検出するためにズームスイッチの操作位置における電圧を基準電圧と比較している。ズームスイッチの位置によって、ズーム位置センサユニットはズーム位置に対応するシリアルデータを返す。

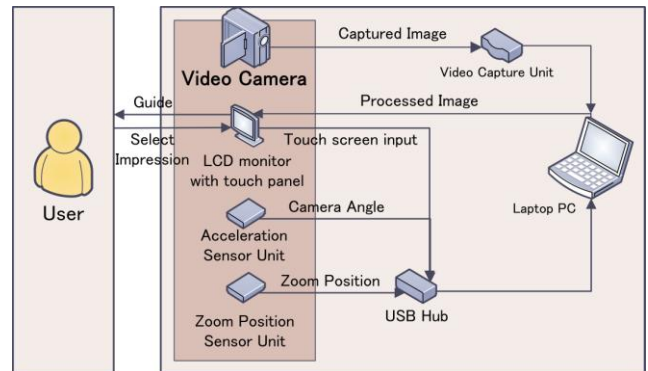


図 7 EmoCap システムのシステム構成

5. ソフトウェアデザイン

本システムは VisualC++およびインテル OpenCV ライブラリを使用して実装を行った。「通常撮影」「感性表示」「感性ナビ」の 3 つの機能があり、通常撮影が選択されると、システムは撮影中の映像をそのまま表示する。

5.1 感性表示

「感性表示」は、現在撮影している映像がどのような感性情報を表しているか表示する機能である。



図 8 感性表示画面

システムは映像とセンサデータを分析し、図 8 のように「感情表現」「弱さ」など対応する感性アイコンをビデオ画面に表示する。

5.1.1 感性ナビ

感性ナビは、ユーザが特定の雰囲気表現したいときにユーザを支援する機能である。この機能では、感性ナビボタンをクリックすると現れるウィンドウにて感情表現、力強さ、弱さ、緊張感・興奮、共感・親密さ、孤独感、解放感の 7 つの感性表現の中から一つを選択すると、システムがそれぞれに応じた撮影方法を提示する。適切な撮影のために、簡潔な指示がアイコンと文章で表示される。

5.2 感性表現支援アルゴリズム

前述の感情表現、力強さ、弱さ、緊張感・興奮、共感・親密さ、孤独感、解放感の 7 つの感性語に対し、参考文献 [2][4] を元にそれぞれ感性表現支援アルゴリズムを考案した。以下は各感性語に対してそれぞれどのように判定を行うかを表している。

5.2.1 感情表現

図 1 に示した通り、人物の感情表現を行う際には、顔のクローズアップを使用すると適切に表現することができる。顔のクローズアップの検出には、顔検出後、フレーム内における顔の高さ h_{face} をフレームの縦の長さで割ったものを計算し、閾値 Th_{emo} より高い場合に真（感情表現である）と判定を行う。

$$I_{emo} = \begin{cases} true, & \frac{h_{face}}{H_{fr}} > Th_{emo} \\ false, & otherwise \end{cases} \quad (1)$$

顔が十分大きく撮影されていれば ($I_{emo} = true$)、システムはサイズが正しいことを緑色の四角で表示する。そうでない場合 ($I_{emo} = false$) は赤色の四角に加えて外向きの矢印を表示することで顔のサイズをより大きくするよう指示を与える (図 9)。

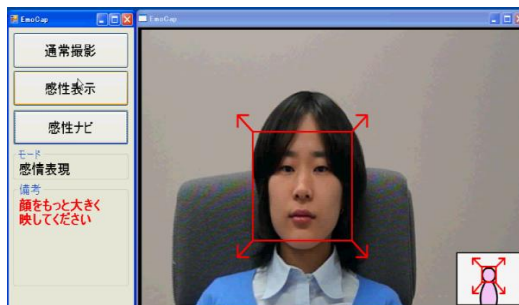


図 9 システム画面 (感性ナビ: 感情表現)

5.2.2 力強さ・弱さ

人物の力強さや弱さを表現するためには、顔を下から (力強さ)、あるいは上から (弱さ) 撮影すると適切に表現

できる。システムはこれらを検出するため、被写体の顔を検出し CAMSHIFT アルゴリズムによって追跡する。カメラの傾きを検出するため、ビデオカメラに搭載した加速度センサを使用し、「力強さ」の場合はカメラの傾き θ_{cam} が閾値 Th_{str} より高い場合に力強い ($I_{str} = true$)、「弱さ」の場合は θ_{cam} が閾値 Th_{wkn} より低い ($I_{wkn} = true$) 場合に弱さであると認識される。認識されない場合は、さらに角度を上か下に変更するよう赤色ビデオアイコンが表示される。

$$I_{str} = \begin{cases} true, & \theta_{cam} > Th_{str} \\ false, & otherwise \end{cases} \quad (2)$$

$$I_{wkn} = \begin{cases} true, & \theta_{cam} < Th_{wkn} \\ false, & otherwise \end{cases} \quad (3)$$

5.2.3 緊張感・興奮、共感・親密さ、孤独感、解放感

緊張感・興奮、共感・親密さ、孤独感または解放感が選択された場合、システムは顔を検出し追跡する。カメラのズームイン・ズームアウト速度を検知すると、それらに対応する閾値と比較する。緊張感・興奮は速いズームインによって表現されるので、システムは現在のズームイン速度 s_{zi} と閾値 Th_{ten} を比較する。速いズームが検出されたら ($I_{ten} = true$) 青色アイコンを表示する。そうでない場合 ($I_{ten} = false$) は、赤色アイコンを表示し速いズームを行うよう促す。

$$I_{ten} = \begin{cases} true, & s_{zi} > Th_{ten} \\ false, & otherwise \end{cases} \quad (4)$$

同様に、共感・親密さは遅いズームインによって表される。システムは現在のズームイン速度 p_{zi} と閾値 Th_{inn} を比較する。遅いズームが検出されたら ($I_{inn} = true$) 青色アイコンを表示する。そうでない場合 ($I_{inn} = false$) は、赤色アイコンを表示し遅いズームを行うよう促す。

$$I_{inn} = \begin{cases} true, & s_{zi} < Th_{inn} \\ false, & otherwise \end{cases} \quad (5)$$

孤独感、解放感も同様にズームアウト情報を処理する。

$$I_{lon} = \begin{cases} true, & s_{zo} < Th_{lon} \\ false, & otherwise \end{cases} \quad (6)$$

$$I_{lib} = \begin{cases} true, & s_{zo} > Th_{lib} \\ false, & otherwise \end{cases} \quad (7)$$

6. システム評価

10 人の大学院生を被験者として、撮影によって 7 つの感

性語を表現するシステム評価実験を行った。なお、評価実験は小型化前のシステムを用いて行った。このシステムはビデオカメラ HXR-MC2000J, タッチパネル付き液晶ディスプレイ HM-TL7T, 異なるズーム検出ユニットを使用している以外は前述のシステムと同一のものである。

評価実験はシステムによる支援のない状態, システムによる支援のある状態の2通りに対して行った。各タスクにおいて, 被験者は椅子に座っている被写体を撮影した。被写体の動作や表情は一定にした。

被験者は評価プロセスの間, 他の被験者に実験内容を教えないように指示された。システムは実験後に行う分析のため, カメラ角度とズーム速度を含むログファイルを作成している。分析は撮影された映像に対して行った。

ページ数制限の都合によりグラフの掲載は一部にとどめている。

6.1 感情表現

実験者はフレームサイズに対しての頭部のサイズを手動で測定した。このタスクでは, 被験者は「感情表現」を行うためにクローズアップショットを使用する必要がある。システムを使用する際, システムは顔を映すように指示を行い, 自動的に顔の高さを測定し正しい大きさになるよう支援を行う。

図 10 は「感情表現」タスクにおいて被験者がどのように撮影したか表している。縦軸はフレームサイズに対する頭部のサイズの比率, 横軸は被験者である。黒線はシステム未使用時, 赤線はシステム使用時を表している。縦に伸びる線は黒, 赤ともに被験者が撮影中にズームを使用したことを表している。短い横線はズームを使用しなかったことを表している。図より, 顔の最大サイズはシステムを使用しないときよりも大きくなる傾向にあることが分かった。また, システムが大きく撮影するように指示を与えたために撮影中にズームを行う傾向があった。

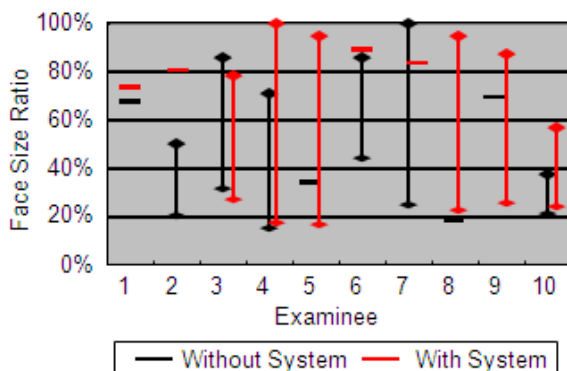


図 10 撮影による「感情表現」の表現における顔のサイズの比率

6.2 力強さ, 弱さ

「力強さ」を表現するためには, 被験者はカメラを上

に向けたまま被写体の顔を撮影する必要がある。図 11 は被験者が「力強さ」を表現するためにカメラを何度傾けたかを表している。

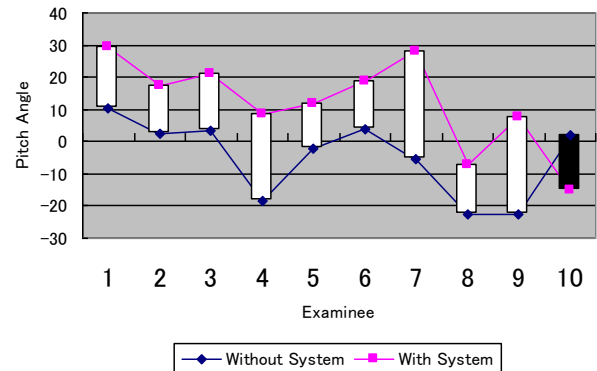


図 11 撮影による「力強さ」の表現におけるピッチ角度の違い

紫色の線はシステムを使用した際のカメラの平均角度, 紺色の線はシステムを使用しなかった際のカメラの平均角度である。図中において棒グラフが白く表示されている場合は, システムを使用した場合, システム未使用時に比べて値が上昇していることを表しており, 黒く表示されている場合は値が下降していることを表している。図中において, 被験者 10 はどの程度上に傾ければよいのか分からず, 指示通りに角度を増加させることができなかった。

「ひ弱さ」タスクにおいてはすべてのユーザが「下に傾ける」指示通り水平より下に向けて映像撮影を行っていた。

6.3 緊張感・興奮, 共感・親密さ, 孤独感, 解放感

「緊張感・興奮」, 「共感・親密さ」, 「孤独感」, 「解放感」はズームを使用することにより表現することが出来る。紫色の線はシステム使用時にどの程度ズームを使用したかを表しており, 紺色の線はシステム未使用時を表している。グラフの縦軸はショット自体の長さに対しどの程度該当のズームを使用したかを表している。

「緊張感・興奮」は速いズームインによって表現される。図 12 は「緊張感・興奮」タスクにおいて被験者がどのように速いズームを使用したかを表している。全ての被験者は速いズームの比率を増加させることができたが, 該当の操作が行われる秒数そのものが少ないために割合が大きくなりにくくなる結果となった。

「共感・親密さ」は遅いズームインによって表現される。このタスクでは殆どの被験者が遅いズームインの使用割合を増加させることができた。例外として, システム未使用時から理解した上で遅いズームインを使用していた被験者 1 名, ズームアウトと勘違いした被験者 1 名が見受けられた。

「孤独感」は遅いズームアウトによって表現される。このタスクでは殆どの被験者においてその割合を増加させる

ことができた。

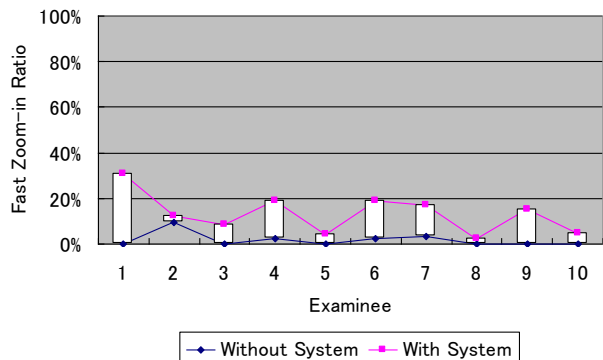


図 12 撮影による「緊張感・興奮」の表現におけるズームイン比率

例外として支援前から理解した上で遅いズームアウトを使用していた被験者や指示に気がつかなかった被験者 1 名が見受けられた。

「解放感」は遅いズームアウトによって表現される。このタスクにおいて殆どの被験者においてその割合を増加させることができたが、例外として指示文を読み間違えた被験者 1 名が見受けられた。

7. 議論および結論

本論文では、感性情報をより適切に表現することを支援する機能を持つビデオカメラシステムを提案した。提案システムは漸進的インタラクションモデルを基にしている。このモデルは、撮影経験の少ないユーザの実験分析結果を基にしており、撮影者の状態や撮影方法を認識して、適切なカメラ操作になるよう随時教示を行うものである。

一般のビデオカメラユーザを対象とした映像撮影に関する実験の結果から、ユーザは撮影経験が十分あればショットが適切に撮影されたことが認識できることが分かった。一方で、ユーザが適切に撮影できていると思っていたとしても、撮影経験が多いからといってより適切な撮影ができるわけではないことも分かった。漸進的インタラクションモデルは撮影中の意思決定プロセスにおいてユーザを支援する必要があるという分析結果を基にしている。

実装システムは、ビデオカメラ、USB 接続の加速度センサ、ズーム検出ユニット、タッチスクリーン付液晶パネル、ノート PC により構成されている。システムを使用する際、ユーザは感情表現、力強さ、ひ弱さ、緊張感・興奮、共感・親密さ、解放感、孤独感の 7 つの雰囲気の中から 1 つを選択する。これら雰囲気は映像のプロフェッショナルが映画を撮影するとき使用する映像制作技法[4]や映画の文法[2]に基づき選択した。

撮影に関する特別な技術や知識がなくても、ユーザはシステムの支援により適切かつ効果的にショット撮影を行うことができる。結果として、ユーザは基本的な撮影技術を

スムーズに身につけ、より効果的な撮影を行うことができるようになる。システムの効果を評価するため、システム評価実験を行った。評価実験より、ほとんどの被験者がシステムを使用することにより指定された印象表現のための適切な撮影を行うことができた。

提案システムにより期待される効果は、効果的な撮影方法を計画する時間を短縮させることができること、提案システムにより提案された撮影支援から撮影方法について学ぶことができることである。実際、被験者の一人は提案システムを使用することで実際の撮影操作を行う前に最良の撮影方法を考える時間を短縮することができたと述べている。

他の被験者は、システムにより提示された適切な撮影方法を理解することができたことインタビューで述べていた。また、システム評価実験から、システムの応答時間は十分に速く、ユーザに余分な負荷を与えないことがわかった。

一方、被験者のコメントから提案システムの問題も明らかになった。提案システムのインタフェースは映像撮影経験を反映した撮影支援をしていない。ユーザを支援する適切なタイミングに関してもまだ明らかになっておらず、「孤独感」などの感性語の意味に関してもユーザによって理解の仕方が違う可能性がある。撮影映像の質を向上させるためには、各ユーザに関してより効率的な支援手法が必要である。

言語はコミュニケーションのために使用される。映像はその使用方法が複雑なことから、熟練したユーザしか効果的にそれを使用することができない。一方で、この視覚的言語は様々な文化や言語のバックグラウンドを持つ人々でも理解することができる。われわれの視覚的表現能力が向上すれば、世界中の人々の映像によるコミュニケーションスキルを向上することが可能になると考えられる。

謝辞 JAIST ライフスタイルデザイン研究センター西本一志教授、宮田一乗教授に有益なコメントを頂いた。謹んで感謝の意を表する。

参考文献

- 1) D. Bordwell and K. Thompson: Film Art: An Introduction, McGraw-Hill (2003).
- 2) D. アリホン, 岩本憲児ほか訳: 映画の文法—実作品にみる撮影と編集の技法, 紀伊國屋書店 (1980).
- 3) J. オーモン, M. マリー, A. ベルガラ, M. ヴェルネ, 武田 潔訳: 映画理論講義—映像の理解と探究のために, 勁草書房 (2000).
- 4) B. Mamer: "Film production technique: creating the accomplished image", 2nd edition ed., Wadsworth Pub Co., Belmont (2000).
- 5) T. Levelle, "Digital Video Secrets: What the Pros Know and the Manuals Don't Tell You," Michael Wiese Productions (2008).
- 6) Hiroko Mitarai, Atsuo Yoshitaka, "Interaction Model for Emotive Video Production," International Journal of Information and Electronics Engineering, Vol. 2, No.5, pp.661-666 (2012).