

# 顔検出に基づく構図モデルを用いた リアルタイム動画撮影支援の提案

鎌原淳三<sup>†1</sup>

**概要:** インターネットの動画配信はスマートフォンのカメラアプリケーションにより誰でも簡単に行えるようになったが、配信される動画では顔の大きさや位置が適切でなく見づらいものも多い。顔の位置と大きさについて経験的に人物の構図として知られているが、動画の撮影支援ができるほどモデル化されたものはない。本稿では実際のテレビ番組の動画から人物の顔の大きさと中心位置の関係を調べ、モデル化した。モデルに基づき動画撮影のリアルタイム支援の実装を提案する。

**キーワード:** 顔検出, 構図, 撮影支援

## Propose to Assist Recording Video in real time using Composition Model with Face Detection

JUNZO KAMAHARA<sup>†1</sup>

**Abstract:** Although anyone can distribute an Internet video which is recorded by the smart phone application, distributed video are not adequate to watch especially in the view of face size and position. The rule of face size and position in video frame is known as the video composition called “the rule of third.” However, there is no mathematical model for assisting video recording in the implementation of that rule. In this article, we present the mathematical model of video composition based on the study of face size and position with actual TV program scenes. Furthermore, we propose the implementation of assisting video recording for person in real time.

**Keywords:** Face Detection, Composition, Recording Assist

### 1. はじめに

カメラ内蔵の高性能スマートフォンの普及により、インターネットでの動画配信が非常に容易になっている。動画共有サイトやリアルタイムに動画配信するサイトも広く認知されており、配信に伴う広告収入により生計を立てるいわゆるプロの動画配信者も登場している。スマートフォンを利用する若年層の女子では SNS での動画投稿経験があるものは 62.1%に及ぶという調査もある[1]。しかし、多くの動画配信者は素人であり、動画撮影の知識やノウハウを持たずに動画配信を行っていると考えられる。

多くの人に見てもらえる動画にするには見やすさを考慮する必要がある。そのためにプロのカメラマンは撮影の際に構図というものを考える。構図とは画面の枠の中に被写体となる対象物をどのように配置するかというもので、経験的に決まっている。著名な手法に三分割法がある。三分割法とは、画面を三等分した位置に線を引いて分割したもので、縦横の三等分線上や交差した点に被写体が来るとバランスが良いとされている [2]。iPhone の標準カメラアプリケーションやその他のカメラにおいても、この三分割法を意味する線が撮影時に表示されている。しかし、撮影者

はこの手法に関する知識がないために、意識して活用できていない。

本研究では、撮影者に動画撮影時の構図を意識させる支援方法を考えるために、人物の構図のモデル化を行ってそれに基づくガイドを提示する。人物の構図を考える際には、人物の顔領域の位置と大きさが対象となる。動画撮影支援を行うには、人物の顔領域の位置と大きさが適切になるようにカメラが撮影者に何らかのガイドを提示する必要がある。つまり、ある一定のモデルに基づいて現在の構図を評価し、適切な構図からずれている場合には、モデルに適合した構図になるようなガイドを提示することで、撮影者により良い構図を選択できるようにする。

#### 1.1 関連研究

カメラに構図を判断する機能を持たせる技術は特許出願などがなされている。富士写真フイルムの特許出願[3]では予め定めた構図(3 分割構図を含む)と比較し画像を評価しガイドを表示するものであるが、具体的な評価の計算方法は示されていない。オリンパス株式会社の特許[4]は人物の輪郭を抽出し人物の構図が最適になるよう画面を加工またはガイドを表示するものである。この技術では最適な構図は予め設定されている。顔検出と構図を組み合わせたもの

<sup>†1</sup> 神戸大学大学院海事科学研究科  
Kobe University

としては、株式会社ニコンの特許出願[5]がある。この技術では、ウエストショット(W)サイズ、バストショット(B)サイズ、顔アップ(F)サイズとして3つの倍率が用意されており、顔領域の検出の結果、顔の大きさがあらかじめユーザが設定したサイズになるようカメラが自動的にズームする。自動ズームすることによって顔領域が画面外にはみ出す場合にはカメラの向きを変えるようガイドを表示するとしている。いずれの技術も各ショットサイズにおける適切な構図について固定の設定値を設けている。本研究では、顔領域の大きさと位置の関係について連続的に変化する計算モデルを提案している点が異なる。

また aesthetic photo composition (美的な写真の構図)に関する研究もなされてきているが[6,7], 3分割法と対角線法および視覚バランススコアなどからターゲットのスコアを評価ものであり、対象物の大きさとその位置の関係を明らかにしたものではなく、撮影時にリアルタイムに望ましい構図になるような指示をどのようにすべきか明確ではない。

## 2. 人物の構図に基づく構図判定モデル

### 2.1 構図の判定

ここで映像の構成要素であるショットについて述べる。人物撮影における基本的なショットにはフルショット、ニーショット、ウエストショット、バストショット、アップショット、クローズアップショットと言われるものがある(図1参照)。これらのショットを組み合わせることで映像の質を高めることができる。実際のテレビ番組では、被写体に対して複数台のカメラを用意し、それぞれにショットを割り振っている。これによりズーム操作が必要なく、カメラを切り替えるだけで視聴者はいろいろなショットの映像を見ることができる。しかし一般ユーザのような、スマートフォンやカメラ1台での動画撮影を考えると、自らの手でズームやワイドを操作する必要がある。つまり一般ユーザが複数の構図を実現するためには、ズーム・ワイド操作を行うことで、ショットの種類を自ら変えなければならない。

動画では同じショットが長く続くと、画面上に変化がないため単調でつまらなく感じることになる。そのため、ショットの種類を変化させることで変化をつける。図1(a)から(f)のそれぞれのショットでの人の顔の中心位置に注目すると、顔が大きくなるにつれて顔の中心位置が下がってきていることが分かる。そこで人の顔の大きさとその時の顔の中心位置に着目すれば、構図の判定ができるのではないかと考えた。

ここで本研究における制約について述べる。本研究ではスマートフォンの顔検出技術を用いることが前提である。そのため人の顔が全部写っていることが条件となる。顔が画面から切れてしまっていると顔認識できない。ここから本研究ではクローズアップショットのような構図はないも

のとする。また顔検出ができないほど顔が小さい場合も適用できない。



図1 ショットの種類

Figure 1 The type of video shot.

また被写体の人数についての制約だが、被写体の数が3人以上になると、様々な配置をとることが考えられるため今回は顔検出で検出した被写体が1人(1Face)か2人(2Faces)の場合を対象とした。

### 2.2 顔領域の検出

顔領域の検出は目的としてスマートフォンのアプリに適用することから、スマートフォンの顔検出機能を用いる。以下では実験に用いたスマートフォン(ASUS Zenfone 5/Android 4.4.2)に対応して便宜的に画面サイズを横1280ピクセル、縦720ピクセルとして示す。

#### (1) 顔領域

顔領域は、Android APIのFaceDetectorを用いて矩形で検出した。当APIでは顔領域は正方領域として検出される。プレビュー表示の画像から①顔検出数の他、顔領域の②left, ③right, ④topを記録し、これらの値から顔領域の面積及び顔領域中心座標を求めた(図2参照)。

#### (2) ワイプ領域の対応

予備実験で放送されたTV番組を対象として検出された中心座標のヒストグラムをプロットしたところ、画面の隅の領域に多く顔領域が検出されたが、内容を精査したところ別カメラで人物の顔のみを小窓で重ね合わせたワイプと呼ばれる部分であった。

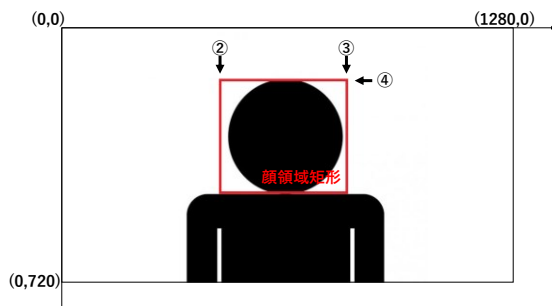


図 2 顔領域矩形

Figure 2 Rectangle of detected Face

そこで前処理でワイプと考えられる領域は後の解析では除いているが詳細は紙面の関係で割愛する。

### 2.3 構図判定モデル

実験で 2016 年に日本で放送されたトーク・情報番組から 7 番組(200 分以上)を用いて顔領域を検出し集計した。その結果、顔領域面積  $S$  と顔領域中心の  $y$  座標( $center.y$ )の間にリニアな関係があることが示唆された[8]。このデータを用いて  $center.y$  と  $S$  との関係を以下のように表す。

$$center.y = aS + \left(\frac{h}{3} - a * minS\right) \quad - (1)$$

$h$  は画面の高さ、 $minS$  は顔検出できる最小の顔領域面積の大きさである。ここで、係数  $a$  を

$$a = \frac{\lambda h - \mu h}{(\alpha h)^2 - (\beta h)^2} \quad - (2)$$

とする。これは顔領域中心の  $y$  座標が画面上部から  $\lambda h$  の時に顔領域の高さ(長さ)が  $\alpha h$  となり、顔領域中心の  $y$  座標が  $\mu h$  の時に顔領域の高さが  $\beta h$  と変化するように定めたものである。これらのパラメータは顔領域の数によって異なる。

#### (a) 1Face の場合

$$\lambda=0.5, \mu=0.33, \alpha=0.6, \beta=0.1$$

#### (b) 2Faces の場合

$$\lambda=0.5, \mu=0.44, \alpha=0.4, \beta=0.1$$

1Face では、顔の高さ(長さ)が画面の 1 割の時に顔領域中心の高さが画面の三分割線の上の線に合致(フルショットサイズ)し、顔の高さが画面の 6 割の時に顔領域中心が画面の真ん中の高さ(アップショットサイズ)となるよう、顔領域面積に対する顔領域中心の  $y$  座標の増加の傾きを定めている(図 3(a)参照)。

2Faces では、顔の高さ(長さ)が画面の 1 割の時に顔領域中心の高さは画面上部から 44%の高さとなり、顔の高さが画面の 4 割の時に顔領域中心が画面の真ん中の高さとなるよう定めた(図 3(b)参照)。このパラメータは実験により経験的に定めた。人物が二人並んだ 2Faces の場合には頭の上の空間が 1Face よりも大きくなる傾向があり、顔の大きさ

もバストショットサイズ程度で横に並ぶ傾向がある。

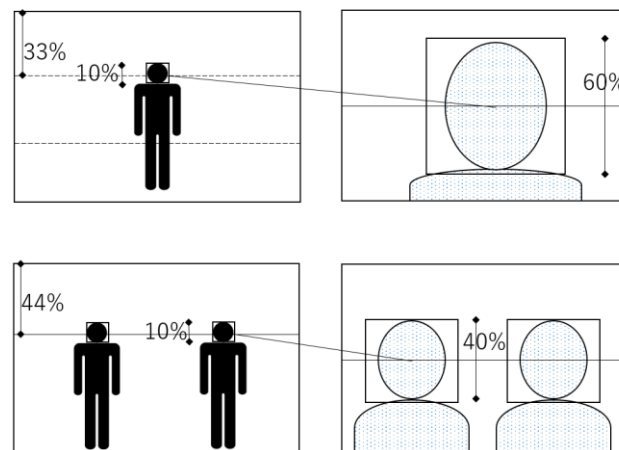


図 3 設定パラメータの意味

Figure 3 Meaning of Parameter Setting

### 3. 構図判定モデルの評価

評価では実験とは異なる 2 番組(2016 年に日本で放送された TV 番組で番組名も異なる)を用いて評価を行った。

評価では、番組で検出された顔領域が 1Face および 2Faces のみを対象とし、事前にワイプ領域の除去を行ったものを検証データとした。評価は、ある顔領域面積の時に顔領域中心の  $y$  座標( $center.y$ )がモデル直線の近傍(実験の際の  $\pm 2\sigma$ )にどの程度分布しているかを対象とした。これは撮影者に人物の構図がモデルに完全に合致することを求めるものではなく、あくまでもガイドを提示するためのものだからである。

結果は次のとおりである。

1Face における検証データの全データ数(フレーム数)は 50,916 個あった。このうちモデルの  $\pm 2\sigma$  範囲内に含まれたデータ数は 49,154 個で 96.5%となった。

2Faces の左の顔については、検証データの全データ数は 22,013 個で、モデルの  $\pm 2\sigma$  範囲内に入っていたデータ数は 21,400 個で 97.2%、2Faces の右の顔については、検証データの全データ数は同じく 22,013 個で、モデルの  $\pm 2\sigma$  範囲内に入っていたデータ数は 21,177 個で 96.2%であった。

実験でのモデル直線からの距離のヒストグラムは見た目では正規分布のように見える(図 4 参照)が、正規性は確認できなかった。このことから、顔領域の面積と位置は複数の確率変数から構成されており、いずれかの変数は期待値が存在しない分布である可能性がある。

これにより、今回の構図判定モデルはあくまでも近似的なモデルであると考えられる。ただ、1Face、2Faces それぞれで一定の範囲内に 95%以上のデータが含まれることから、リアルタイムに簡易に構図を判定するためのモデルとしては有効であると考ええる。

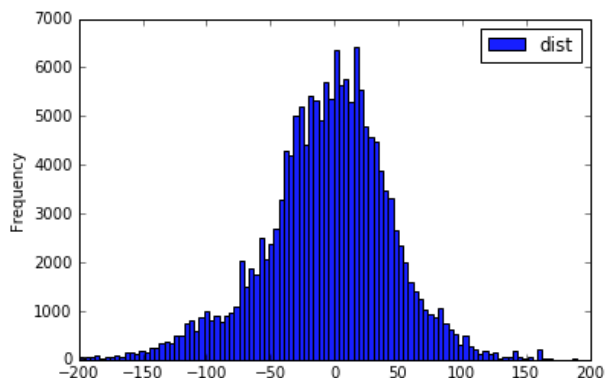


図4 モデル直線からの距離のヒストグラム

Figure 4 Histogram of Distance from Model

## 4. リアルタイム構図判定に基づく撮影支援の実装

今回提案する構図判定モデルに基づいて撮影支援を行うカメラアプリケーションの実装を行った。実装はソニーモバイルコミュニケーションズ製の Xperia Z5 Compact (Android 7.1.1)で行った。なお、実装は端末を横向き (landscape)で保持している前提で行っている。これは視聴される動画がほとんど横長を前提としているためである。

撮影支援のガイド提示を2種類例示する。

### 1)モデル適合範囲表示

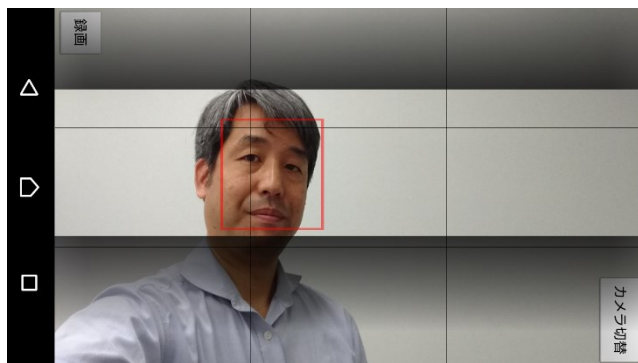


図5 ガイド表示1のスクリーンショット

Figure 5 Screenshot of Recording Guide 1

### 2) カメラ移動方向表示



図6 ガイド表示2のスクリーンショット

Figure 6 Screenshot of Recording Guide 2

モデル適合範囲表示では、顔領域面積から計算できるモデル直線から $\pm 2\sigma$ 範囲については通常の表示とし、そこから外れた部分について半透明のマスク表示を行う(図5参照)。

カメラ移動方向表示では、検出している顔領域がモデルの範囲から外れている場合に、顔領域をどちらに動かせば良いか矢印記号で示す(図6参照)。

なお、顔検出が行えた場合のみ、赤枠で顔領域が表示されるとともに、この2種類の撮影支援が表示される。

支援の有無によって撮影映像における構図がどの程度モデルに適合するかの比較は今後行う予定である。

## 5. おわりに

本研究では放送された番組の人物構図について、統計的なデータから構図判定のためのモデルを提案し、モデルに基づいた撮影支援の実装例を示した。モデルを構築する際に用いたのとは別のTV番組に対してもモデルが適用できるか評価を行い、95%以上がモデルの範囲内にあることが確認できたが、正規性は確認できずリアルタイムに処理するための近似的なモデルであると言える。

今後、実装の有効性の評価や、3人以上の人物の構図に関する分析、モデルを構成する要因の詳細分析等を行っていく予定である。

## 参考文献

- [1] “LIVE 配信・動画投稿に関する調査結果”.  
<https://lab.testee.co/grp-movie-result>, (参照 2018-06-14).
- [2] 益子広司・内田一夫 (2013) 映像カメラマンのための構図完全マスター, 玄光社出版, pp.22
- [3] 特開 2001-167253 (2001), 富士写真フイルム株式会社, 特許公開報(A)
- [4] 特許 3971240 (2007), オリンパス株式会社, 特許公報(B2).
- [5] 特開 2010-41365 (2010) 株式会社ニコン, 特許公開報(A)
- [6] Ligang Liu et. al., “Optimizing Photo Composition,” EUROGRAPHICS 2010, Volume 29 (2010), Number 2, pp.469-478, Blackwell Publishing, 2010.
- [7] Roberto Gallea, et. al. “Automatic Aesthetic Photo Composition,” : ICIAP 2013, Part II, LNCS 8157, pp. 21–30, Springer-Verlag, 2013
- [8] 塩山昌志,鎌原淳三, “顔認識を用いた動画撮影支援のための構図判定モデルの提案,” 電子情報通信学会総合大会学生ポスターセッション ISS-P-96, 2017.