Vol.2012-AVM-79 No.25
2012/12/7

映像符号化情報を用いたシーン判定手法の一検討

三反崎暁経[†1]　小野尚紀[†1]　清水淳[†1]

あらまし　本報告では，映像符号化技術を応用し，符号化情報のみを用い，機器追加，復号化（デコード），画像処理等をすることなく低負荷に撮像シーンを判定する手法を提案する．本検討では，符号化方式としてフレーム間予測が可能な H.264/AVC を用いた．シーン判定に利用する符号化情報として，発生符号量や量子化値から算出される複雑度指標値を用いた．本手法を用いることで処理負荷の増加なしで，シーン（暗さ，動きの有無）を高精度に判定することが可能となる．

A Study on　Method of Estimating Scene by Video-coding Information

Tokinobu MITASAKI[†1]　Naoki ONO[†1]
Atsuhi SHIMIZU[†1]

Abstract　In this paper, we propose the method of estimating scene like the darkness and the motion by applied technology of the video-coding.　As video-coding method, we use H.264/AVC which is the inter-frame codec.　As the video-coding information, we use the complexity composed of data amount and the quantization value.　It is able to estimate the scene with high accuracy by the only coding information without adding any sensors, the image-processing, and decoding.

1.　はじめに

近年，ブロードバンドサービスの普及に伴い，カメラ映像を用いたサービスが数多く展開されている[1]，[2]．これらのサービスでは，カメラで撮影し，符号化した映像を受信者側で復号化することによって，監視，遠隔地の高齢者見守り，ＴＶ（テレビ）電話等を実現している．監視，見守りという観点では，カメラ映像に赤外線センサや照度センサ等を併用することで発生イベントの有無を自動的に検知し，高機能化したものも存在する[3]．この場合はセンサが必要となり，価格面，システムとしての故障率，カメラとセンサの連動面（汎用性）で課題が残る．これに対し，カメラのみで，センサを用いずにイベントの発生を検知する手法が提案されている[4]～[7]．これらの手法では，カメラ映像を取り込み，取り込んだ映像を符号化した映像データをイベント検出装置に送信する．イベント検出装置は，映像データの情報量，量子化レベル，動きベクトル等の情報からイベントが発生したことを検出する．しかし，いずれの手法も明るさに関して言及または検討していない．ＩＳＯ（国際標準化機構）感度が明るさにより自動的に変化するカメラを用いる場合は，明るさが一定値以下になると感度が高まり，イベントが発生していないにもかかわらず，ノイズが増加することで，映像データの各ピクチャにおける情報量，量子化レベルおよび動きベクトルは変化するため，誤判定をしてしまう．従来手法では，映像データの各ピクチャにおける情報量，量子化レベルおよび動きベクトルにより判定を行うため，明るさの変化に対応できない．つまり，明るさ情報が判定に関与しないため，例えば夕暮れのように明るさ環境が時刻とともに大きく変化する環境では正常に判定ができず，実用面から考えると現実的ではない．現在の多くのカメラは，ある一定の暗さになるとＩＳＯ感度が自動的に高くなるよう制御され，カメラ感度が高まることでカメラノイズが大きくなり，映像情報量や量子化レベルといった絶対値は変化するため，何もイベントが起きていないにも関わらず，警報が誤通知されることが考えられる．

図 1 に TV 電話を例として，明るさを考慮することのメリットを説明する．TV 電話を主に用いる環境（屋内）では，照明のオン／オフ等が日常的に行われており，例えば，動きの有無をイベントとして検知したい場合に，明るさを考慮しないシステムでは，明るさが変わるとピクチャの情報量が変化するため，動きがあったと判定してしまうことが考えられる．一方，明るさを考慮すると，明るさと動きの有無は別々に判定されるため，明るさが変わっただけでは動きがあったと判定されず，イベントの誤判定を抑止できる．例えば，自身と通話先（対向）とは通常ＴＶ電話接続はなく，対向で人や物の動きがあった場合や，明るさが変化した場合等にのみ対向側に通知することで，NW 帯域の有効利用，通話不成立の低減，およびコミュニケーションのきっかけ作り（推進）に役立てることができる．

本検討は上述のような明るさが変化する状況であっても，映像からイベント（明るさの変化や，カメラ前での人や物の動き）の発生を高精度に検出する手法を提案する．

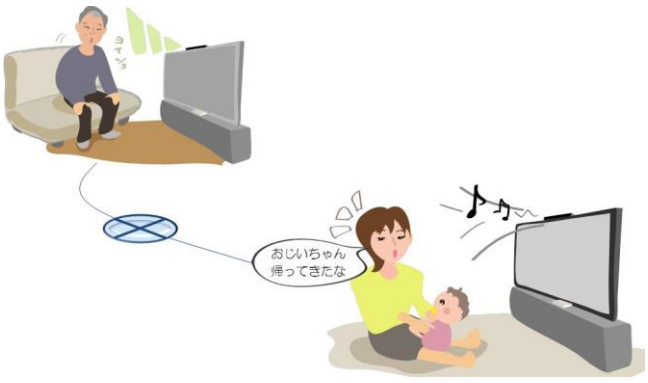

図 1　TV 電話での適用例

Figure 1　Example of application of videophone

†1　日本電信電話株式会社　NTT メディアインテリジェンス研究所
NTT Media Intelligence Laboratories, NTT Corporation

2. 符号化情報を用いたシーン判定手法

2.1 シーン判定手法の概要

図 2 に提案手法のフロー概要を示す．まず，カメラで撮像し，その映像をフレーム間予測方式（例えば MPEG2 や H.264/AVC，HEVC 等）で符号化する．次に映像解析部において，符号化ストリーム内の発生符号量と量子化値から以下の式で複雑度指標値 Xt(N)をフレーム毎に算出する．

・Xt(N) = Qp(N)×R(N)

ここで，t はピクチャタイプ（I/P/B），N はピクチャ番号，Qp はスライスヘッダ中に含まれる slice_qp_delta, R はピクチャの発生符号量をそれぞれ示す．次に本指標値を用い，明暗判定部において撮影環境の明暗判定を実施し，明るいと判定された場合は，動き判定部において動き判定を実施する．一方，暗いと判定された場合は，カメラ映像がほぼ真っ黒な状態であり，動き有無の判定が困難であることから，動き有無の判定は実施しない．

符号化以外の判定に要する処理としては，複雑度指標値を算出する乗算, および判定時の閾値処理のみであるため，非常に低負荷にシーン判定を実現できる．

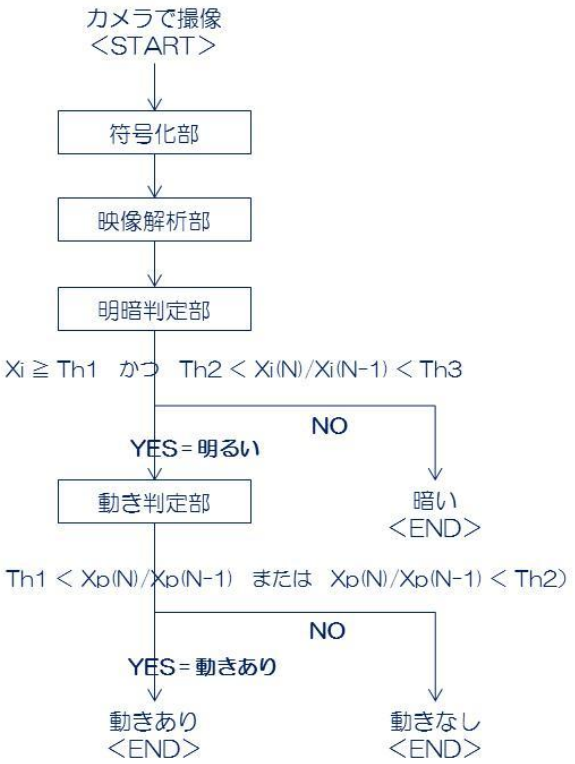

図 2　シーン判定手法フローの概要

Figure 2　Flow diagram of estimating scene

2.2 明暗判定

明暗判定部における明暗判定手法について説明する．明暗判定には，I ピクチャの複雑度指標値 Xi を使用する．Xi が以下の条件 A，B をともに満たす場合は「明るい」と判定し，満たさない場合は「暗い」と判定する．

・Xi(N) ≧ Th1　　（A）

・Th2 ＜ Xi(N) / Xi(N-1) ＜ Th3　　（B）

条件 A が示す内容を以下に説明する．例えば，撮像シーンが明るい場合，背景，被写体には，なんらかのテクスチャが存在することが多い．テクスチャが存在する場合は，Xi が大きくなるため，閾値 Th1 以上となる．一方，撮像シーンが暗い場合，背景，被写体は真っ黒な映像となり，テクスチャがほとんど存在しないため，Xi は非常に小さくなる．例えば，明るい環境下における Xi = 1000，暗い環境下における Xi = 100 とすると，Th1 = 500 とすることで，「暗い」という判定結果が得られる．Th1 を非常に小さく設定することで，照明，太陽光ともにほぼない状態のみを「暗い」と判定することができる．

次に条件 B が示す内容を以下に説明する. Xi(N) / Xi(N-1) は N 番目の I ピクチャとその一つ前の I ピクチャの複雑度指標値の比率であり，例えば，本比率が大きく変化する場合とは，照明の ON/OFF のような明るさが大きく変化する場合を意味する. 一方, 本比率が大きく変化しない場合は，一定の明るさが継続されていることを意味する．例えば，Xi(N) = 1000，Xi(N-1) = 800，Th2 = 0.9，Th3 = 1.1 とした場合，条件 B を満たさず，「暗い」という判定結果が得られる．

条件 A と B をともに満たす場合とは, 一定以上の明るさがあり，明るさ変化があまりない状況と判定する．なお，本判定の時間的粒度は，I ピクチャの時間方向間隔（符号化設定）によるものであり，例えば 1 秒に 1 回 I ピクチャを挿入する設定であれば，判定の最小間隔は 1 秒となる．

2.3 動き有無判定

動き判定部における撮像シーンの動き有無判定手法について説明する．動き判定部では，P ピクチャの複雑度指標値 Xp を使用する. Xp が以下の条件 C を満たす場合は撮像シーン，つまりカメラ前の人，物，カメラ自身において「動きあり」と判定し，満たさない場合は「動きなし」と判定する．

・Th4 ＜ Xp(N) / Xp(N-1) ，または

・Xp(N) / Xp(N-1) ＜ Th5　　（C）

条件 C が示す内容を以下に説明する．例えば，カメラ前で人の動きが発生すると，動きがない場合に比較して P ピクチャの複雑度指標値が大きくなり，1 枚過去の P ピクチャの Xp と判定対象の Xp との比率が大きくなる. これにより「動きあり」と判定する．一方, 動きが全くない場合は，Xp は低い値のまま, ほとんど変化しないため,「動きなし」と判定する. 例えば, Xp(N) = 500, Xp(N-1) = 300, Th2 = 0.7, Th3 = 1.4 とした場合，条件 C を満たさず,「動きなし」という判定結果が得られる．Th4 を小さく，Th5 を大きく設定することで，小さな動きでも「動きあり」と判定でき，

逆に Th4 を大きく，Th5 を小さくすることで，大きな動きのみを「動きあり」と判定できる．なお，本判定の時間的粒度は，P ピクチャの時間方向間隔（符号化設定）によるものであり，例えば 30fps の映像で，IPPP の（B を使用しない）GOP 構造である場合，判定の最小間隔は 1/30 秒となる．

3. 事前実験（PC ベース）

提案手法の有効性を確認するため，まず CPU 負荷を考慮せず，高スペックな PC ベースで明暗および動き有無の判定について事前実験を行った．本実験では，カメラ映像を入力とし，複雑指標値 Xi および Xp を出力とする．

3.1 事前実験系および条件

表 1 に事前実験条件を，図 3 に事前実験系をそれぞれ示す．汎用的な web カメラで映像を撮像し，それを H.264 方式で符号化する．web カメラと PC 端末を USB 接続し，符号化ストリームを同端末に入力する．同端末内では符号化ストリームから複雑度指標値 Xi，Xp(N)を算出し，本値を基に閾値処理により「明暗」，「動き有無」を判定する．なお，事前実験では，照度は日中の居室で照明を消灯（15lx）または点灯（700lx）とし，カメラ～被写体間距離は 1.5m で固定とした．

表 1　事前実験条件
Table 1　Pre-experimental conditions

符号化パラメタ	値
符号化方式	H.264 / AVC
解像度	1280 x 720
フレームレート [fps]	30
ビットレート [kbps]	2000
OS 環境	Windows 7
カメラ位置	固定
カメラ設置場所	居室（場所は 1 か所のみ）
CPU	3.5GHz 程度
照度 [lx]	15 または 700
距離 [m]	1.5（固定）

図 3　事前実験系
Figure 3　Pre-experimental setup

3.2 事前実験結果（明るさ判定）

図 4 に平日一日間の Xi の推移を示す．図中の時刻 A は日の出付近の時刻であり，日の出とともに Xi の値は徐々に増加する．これは，太陽光が窓から入り，それに伴いカメラ前の映像が見え始めたこと，およびカメラの ISO 感度が高まったことに起因すると考えられる．次に時刻 B は居室において照明が点灯された時刻付近であり，ここから時刻 C の照明の消灯までは，多少の変動はあるものの，ある一定値以上を保持し続ける．時刻 C は居室での消灯（夜）時刻を示しており，照明光も太陽光もないことから，カメラ前が真っ暗になり，Xi の値が一気に落ち込んだものと考えられる．以上より，Xi の値を用いることで，カメラ前の明るさ環境が判定できると考えられる．

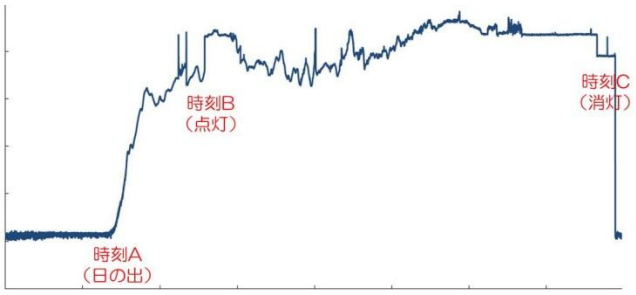

図 4　1 日間の Xi 推移
Figure 4　Trend of Xi for a day

3.3 事前実験結果（動き有無判定）

次に動き有無判定の実験を行った．カメラ前での動作，背景，被写体のテクスチャおよび明るさは以下の表 2 のパタンで実施した．

図 5 に 1 回だけフレームインした時の Xp 変動推移を示す．撮像環境が薄暗い場合では，カメラの ISO 自動感度調整機能によりカメラノイズが増加し，動きがあった際の Xp 変動を確認することが困難であった．なお，手を振る，カメラ画角のフレームイン，アウトの繰り返しについても同様に実験したが，やはり暗い環境下では Xp の変動を得ることは困難であった．一方，明るい環境下では，背景，被写体（動作人物の洋服の模様）の平坦，複雑に関わらず，

表 2　動き判定の実験パタン
Table 2

動作	明暗	背景	被写体
手を振る	・明るい ・薄暗い	・平坦 ・複雑	・平坦 ・複雑
フレームイン，アウトの繰り返し	・明るい ・薄暗い	・平坦 ・複雑	・平坦 ・複雑
1 回だけフレームイン	・明るい ・薄暗い	・平坦 ・複雑	・平坦 ・複雑

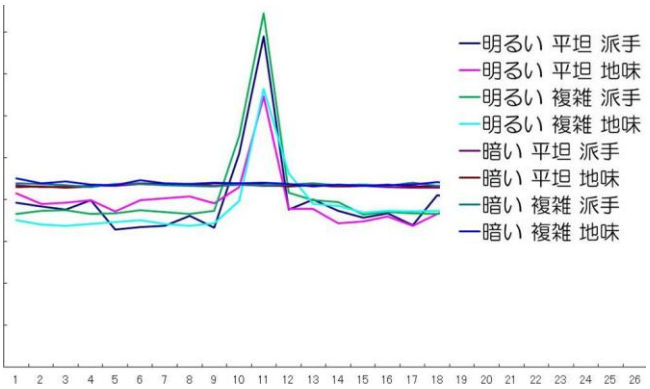

図 5　Xp 推移（1 回だけフレームイン）

Figure 5　Trend of Xp (once frame-in)

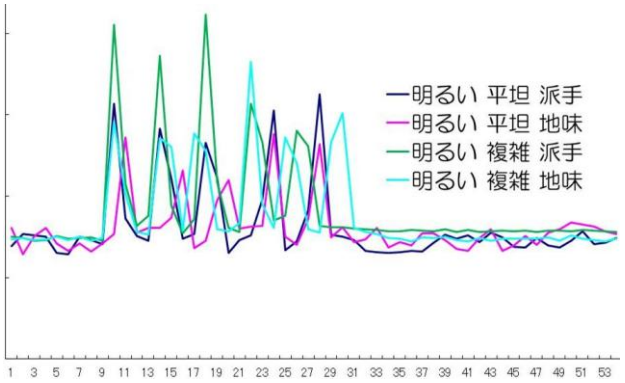

図 6　Xp 推移（フレームイン，アウトの繰り返し）

Figure 6　Trend of Xp (multiple frame-in/out)

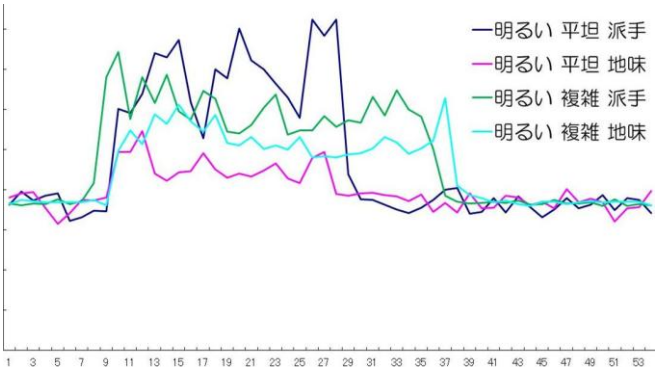

図 7　Xp 推移（手を振る）

Figure 7　Trend of Xp (shake hands)

フレームイン（動き）の検知により，Xp が変動することが確認できた．

図 6 にカメラ前を複数回フレームイン，アウトを交互に繰り返した場合の Xp の推移を示す．なお，薄暗い場合は上述の通り，Xp の変動が確認できなかったことから，グラフは省略した．フレームインの場合と同様に，フレームアウトについても，動きの検知により，Xp が変動することが確認できた．

図 7 にカメラ前で一定時間内，手を振る動作をした際の Xp 推移を示す. 図 6 と同様に暗い場合のグラフは省略した．手を振る動作の場合，フレームイン，アウトに比較し，映像の動領域が小さいため, Xp の変動も全体的に小さくなったが，変動の傾向は確認できた．

以上より，明るい環境下では，フレームイン，アウト等の大きい動きをはじめ，手を振る等の小さい動きもある程度確認できることを示した．

4. 実験

事前実験で，本手法を用いた明暗および動き有無の判定の実現性をハイスペック CPU の PC 端末ベースで確認できた. 次に本手法を低スペック CPU の Android 端末に実装し，低負荷に実現できることを確認する．本実験ではカメラ映像を入力とし，明暗（消灯），動きの有無の 3 パタンを出力結果とする．

4.1 実験系および条件

表 3 に実験条件を，図 8 に実験系をそれぞれ示す．web カメラで映像を撮像し，それを H.264 方式で符号化する．web カメラと Android 端末を USB 接続し，符号化ストリームを同端末に入力する．同端末内では符号化ストリームから複雑度指標値 Xi，Xp(N)を算出し，本値を基に閾値処理により「明暗」,「動き有無」を判定する．なお，本実験では，照度とカメラ～被写体間距離をパラメタとし，事前実験と同様に，カメラ前にフレームイン，アウトおよび手を振る動作で動き有無とした．

表 3

Table 3　Experimental conditions

符号化パラメタ	値
符号化方式	H.264 / AVC
解像度	1280 x 720
フレームレート [fps]	30
ビットレート [kbps]	2000
OS 環境	Android 2.X
カメラ位置	固定
CPU	1GHz 以下
照度 [lx]	0.01～800
距離 [m]	0.5～4

図 8　実験系

Figure 8　Experimental setup

4.2 実験結果（明暗判定）

照度を可変パラメタとして割り当て，今回は「消灯 = 暗い」状態を判定することとし，閾値設定を行った．図 9 に各照度におけるカメラ画像を示す．図 9 に示す通り，照度が非常に低い場合，カメラ画像はほとんど真っ黒な状態となる．これにより Xi の値は非常に小さくなり，2 章における条件 A を満たさないため，暗いと判定された．図 10 は日中に部屋の照明を消灯し，照度が 11 [lx]の場合のカメラ画像であるが，この場合は「消灯」とは判定されず，高精度な判定結果が得られた．また，図 11 に示すような，背景のテクスチャが細かい場合においても，正常に判定ができた．一方，図 12 は背景が平坦な壁で，照明をつけた（700 [lx]）環境下で実験したところ，背景のテクスチャがほとんどなく，かつ明るさも十分であり，映像にノイズもほとんどないことから，条件 A，B を両方見たし，明るいにも関わらず「消灯」と誤判定された．ただし，本環境条件は一般的なシーンでは非常に稀であり，同様の環境で追加実験を行ったが，一度も誤判定は起きなかった．そのため，実用上では同様の誤判定が起きる可能性響は非常に低いと考えられる．

図 9　照度 0.05 [lx]
Figure 9　Illuminance : 0.05 [lx]

図 10　照度 11 [lx]
Figure 10　Illuminance : 11 [lx]

図 11　照度 800 [lx]
Figure 11　Illuminance : 800 [lx]

図 12　照度 700 [lx]
Figure 12　Illuminance : 700 [lx]

4.3 実験結果（動き有無判定）

照度と距離をパラメタとして割り当て，カメラ前で手を振る，フレームインおよびフレームアウト等を行い，「動きあり」状態を判定することとし，それに応じた閾値設定を行った．図 13 に横方向にカメラ～被写体間距離，縦方向に明るさ（背景の複雑さ）をそれぞれ変えた場合の判定結果を示す．ここでは，TV 電話への適用を仮定し，明るい環境での結果を整理することとする．明るい環境では，背景のテクスチャの有無に関わらず，3m 程度の距離までは 90%以上の確率で正しい判定結果を得ることができた．カメラ

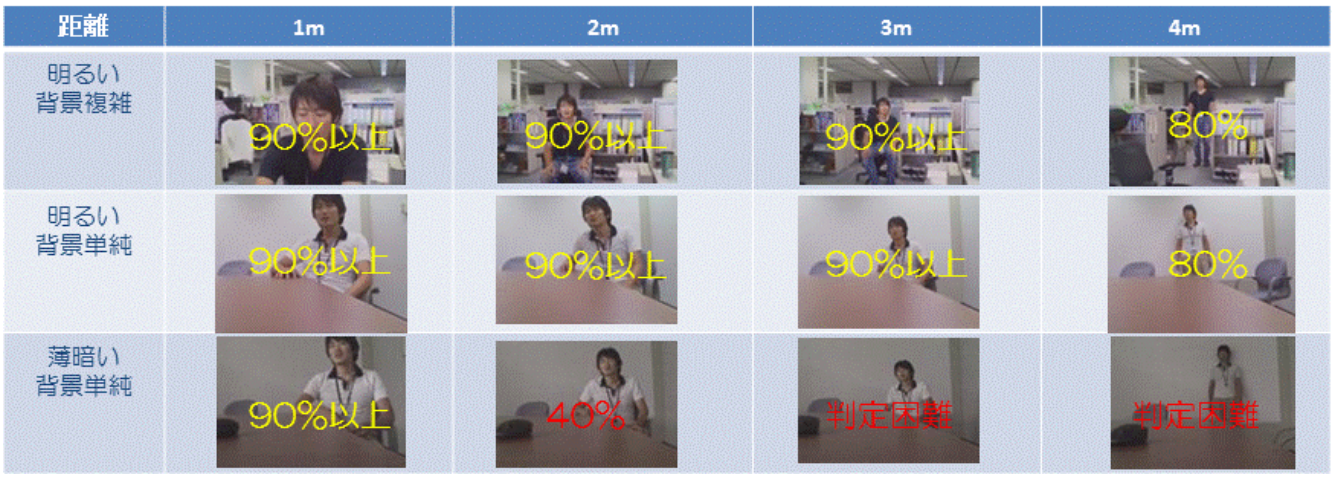

図 13　動き有無判定結果
Figure 13　Result of estimating existence or non-existence of motion

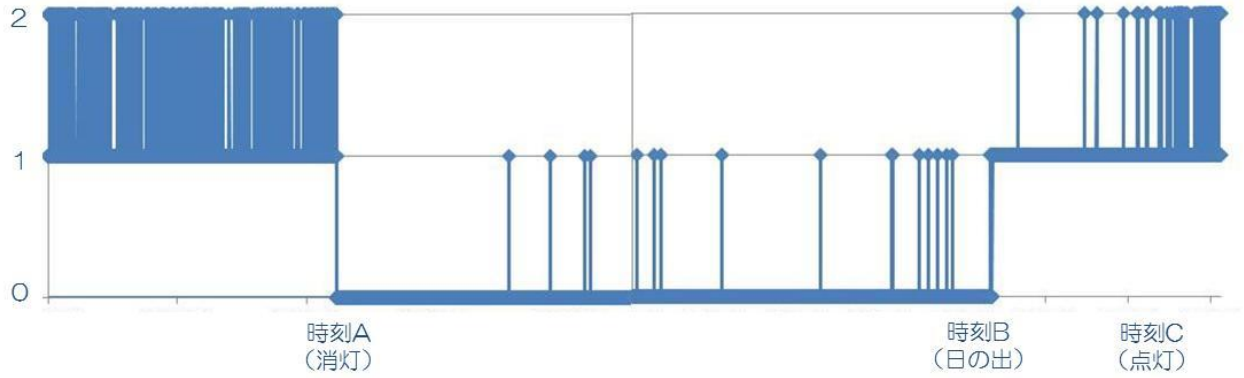

図 14　終夜試験時の判定結果推移
Figure 14　Trend of estimating result in all-night test

～被写体距離の 3m は一般的な家庭における TV 電話利用環境のほとんどを包含できる距離であると考えられる．次に日中に部屋の照明を消灯し，薄暗い環境における動き有無判定結果について説明する．薄暗い環境になると，距離が 1m では明るい場合と同様の高い精度で正しい判定結果が得られるが，距離が長くなると，判定精度が著しく低下した．理由としては，カメラの ISO 感度自動調整機能が自動的に動作し，多くの光量を取り込もうとするため，ノイズが多くなり，結果として複雑度指標値も高くなってしまう．これにより，動き．が発生したことによる複雑度指標値の変動が検知しづらくなったためと考えられる．

4.4 実験結果（終夜試験）

居室にカメラを設置し，終夜試験を実施した際の明暗，および動き有無の判定結果推移を図 14 に示す. 縦軸は 0 が消灯, 1 が動きなし, 2 が動きありをそれぞれ意味する. また，横軸における時刻 A は照明消灯時刻，B は日の出付近，C は照明点灯時刻をそれぞれ示す．消灯後は一度も点灯つまり明るい状態にはなっていないため，A～B 間に縦軸が 1 になっている点は誤判定の点（回数）を示している．今回は 1 秒に 1 回の頻度で明暗判定を行っているため，消灯から日の出までの間，25642 回の明暗判定が実施され，そのうち 15 回が誤判定という結果となった. 判定精度としては 99.9%以上を実現した．また，動き判定を実施した回数は 19336 回であり，動きあり，なしと判定された回数がそれぞれ 732 回，18604 回となった．人の在室状況を考慮すると 90%以上の確率で適切に判定を実現できたと考えられる．

表 4　判定結果
Table 4　Result of estimating scene

判定	判定精度（判定回数）
暗い（消灯）	99.9%以上（25627 回 / 25642 回）
動きなし	18604 回
動きあり	732 回

4.5 CPU 負荷

シーン判定に対する CPU 負荷を測定した．用いた Android 端末の CPU は 1GHz 以下である．実行結果としては，本判定を実施しても，CPU 負荷は 0%増加，つまり，処理負荷は最大でも小数第一位を四捨五入すれば, 0%となる程度であり，処理負荷はほとんどないという結果が得られた．

5. おわりに

本報告では，映像符号化情報のみを用いたシーン判定手法について提案した．判定可能なシーンとしては，撮像環境の明暗，およびカメラ自身または被写体（カメラ前の人や物の動きの有無）を取り上げた. 本手法を用いることで，明暗およびカメラ前の動き有無を 90%以上の精度で判定することができた．

今後，映像のみでなく，音声と組み合わせることで，映像でカバー困難な状態（薄暗い環境下での精度向上や，カメラ画角外の動き判定等）をカバーすることや，時刻情報と組み合わせることで，さらに判定精度を高めることを検討する．また，今回は符号化方式としてフレーム間予測方式を用いたが，JPEG 等の画面内予測方式を用いたシーン判定手法について，合わせて検討していきたい．

参考文献

1) http://www.ntt-east.co.jp/business/solution/camera_visual/
2) http://www.webcam-service.jp/
3) http://www.ntt-at.co.jp/page.jsp?id=1793&content_id=764
4) 新倉 他：MPEG 符号化データからのシーンチェンジ検出方法の検討，電気学会通信研究会 Vol.CMN-97, No.37-42, pp.7-12 (1997)
5) 加藤 他：MPEG 符号化データからの移動物体検出に関する一検討, 2001 年電子情報通信学会・システムソサイエティ大会, pp.89
6) 玄蕃 他：MPEG 符号化されたサッカー映像に対するシーンの自動分類のための情報抽出，映像メディア学会誌 Vol 55, No.3, pp.417-421 (2001)
7) 正木 他：MPEG 符号化データを用いた人物領域と動きの抽出手法，2010 年電子情報通信学会総合大会，pp.121