

ユーザの興味度による複数映像表示方式

山足 公也 谷 正之 谷越 浩一郎

E-mail : yamaashi@hrl.hitachi.co.jp

日立製作所日立研究所

319-12 日立市 大みか町7丁目1-1

遠隔会議システムやプラント制御システムでは、多数の映像を表示する必要がある。多数の映像を表示するには、ネットワークなどの多量の計算機リソースを必要とし、その必要量がシステムの計算機リソースの限界を超えてしまう場合がある。本稿では、限られた計算機リソースの中で効率的に複数の映像を表示する手法について述べる。複数の映像を見るとき、ユーザは、興味あるビデオを詳しく見ており、その他の映像に対しては、その全体的な動きを見ている。そこで、ウィンドウレイアウトからユーザの興味度を推定し、興味のある映像は詳しく、興味の薄い映像は荒く表示する。これにより、映像表示に関する計算機リソースの必要量を効率的に削減でき、限られた計算機リソースを有効に利用することができる。

A Technique of Displaying Multiple Videos According to Users' Interests

Kimiya Yamaashi Masayuki Tani Koichiro Tanikoshi

E-mail : yamaashi@hrl.hitachi.co.jp

Hitachi Research Laboratory, Hitachi Ltd.

7-1-1 Omika-cho, Hitachi-shi, Ibaraki-ken, 319-12, Japan

Many applications, such as tele-conference systems, need to display multiple videos. In those applications, displaying multiple videos overwhelms limited computing resources due to the vast data. This paper describes a technique that allows multiple videos to display in the limited computing resources. When users look at multiple videos, they only look at videos of interest in detail and get global context of others. The technique displays videos of interest in more detail by degrading others. The technique distorts videos in the space and time domain according to users' interests, estimated based on the screen conditions.

1. はじめに

遠隔会議システムやプラント制御システムでは、たくさんのビデオを表示する必要がある。例えば、原子力プラントや鉄鋼プラントでは、20から30台の工業用カメラを配置し、オペレータは、プラント全体の状態を把握するために同時に複数のカメラ映像で現場の状況を監視している。また、遠隔会議システムでは、ユーザは、映像を用いて複数の参加者の顔の動きや状況を見ている[1]。

映像の情報量が膨大であるため、映像を表示するには、多量の計算機リソースが必要であり、その必要量がシステムの計算機リソースの容量を越えてしまう場合がある。例えば、ハーフサイズの映像(320*240)を秒30こまで表示すると、その情報量は、 $320 * 240 * 30 =$ 約2Mバイトとなる。この映像を10Mビット/秒のイーサネット で伝送しようとする と、1つの映像を伝送するだけでイーサネットの伝送限界を越えてしまい、伝送することができない。

映像の情報量を少なくするため、JPEG (Joint Photographic Experts Group) など多数の映像圧縮技術が検討されている。JPEG方式の場合、情報量を1/20-1/50の映像圧縮を行うことができる。[2][3][4]しかし、複数の映像を表示しようとする と、画像圧縮技術を用いてもすべての映像を表示することはできない。このような場合、より映像の情報量を削減するため、映像の品質を劣化させる必要がある。例えば、JPEG方式の場合には、量子化のビット数を少なくするなどの方式が用いられる。

複数の映像の情報量を削減して表示する場合、すべての映像を一様に劣化させる必要はなく、ある映像は品質良く表示し、別の映像は劣化させて表示することもできる。この場合、どの映像を品質良く表示し、どの映像の品質を落とすのか、といった映像の優先度制御を行う必要がある。

従来の画像圧縮技術では、1つの映像の表示方法に関して主に議論されており、これまで、複数の映像の優先度制御方法については議論されていない。本報告では、この複数の映像を表示する場合の優先度制御方法について検討する。特に、映像のレイアウト情報から映像のユーザの興味度が推定できることに注目し、推定したユーザの興味の高い映像は高品質に表示し、ユーザの興味の薄い映像は画像を劣化して表示する手法について述べる。本手法を用いることにより、ネットワーク等の限られた計算機リソースを効率良く使用することができる。

2. 映像の優先度制御

映像の優先度制御方法は、一度に1つの映像だけを表示し他の映像を止める方式(シングルタスク方式)と、複数の映像を同時に表示する方式(マルチタスク方式)との2つの方式に大きく分類できる。

シングルタスク方式を用いると、ユーザが指定した映像だけが表示され、他の映像は静止画の状態となる。このため、一度に1つの映像しか見れない。その代わりに、すべての計算機リソースを1つの映像のために使用することができ、最も高品質な映像を表示することができる。マルチタスク方式を用いる場合には、すべての映像を常に動画で見ることができる。その代わりに、複数の映像で計算機リソースを共有して使用するため、シングルタスク方式に比べ、映像の表示品質は悪くなる。

映像の優先度制御方法を検討するに際し、遠隔会議システムやプラント制御システムなどの複数の映像を表示するシステムでの映像の利用方法について考えてみる。遠隔会議システムやプラント制御システムでは、ユーザは、一度にすべての映像を見ているわけではない。ユーザは、主に興味

のある映像を見ながら、ときにそれ以外の映像も見て、全体の状況を把握している。例えば、遠隔会議システムでは、ユーザは、主に、直接話をしている人の顔を見ているが、ときどき、他の人の動きなども見て、会議全体の雰囲気や状況に注意を払っている。そのため、映像を表示する際には、直接の話し相手だけでなく、他の会議の参加者の映像も表示する必要がある。プラント制御システムでも、オペレータは、主に、点検などの作業を行っている現場の映像を見ているが、他の現場で異常がないかどうかを確認するため、ときどき、作業現場以外の映像も見て、プラント全体の状況を把握している。

遠隔会議システムやプラント制御システムのような複数の映像を表示するシステムでは、ユーザの興味がある映像だけを表示すればよいのではなく、それ以外の映像も同時に表示する必要がある。そのため、複数の映像を表示するシステムでは、マルチプロセス方式を用いる方が有効である。

マルチプロセス方式における個々の映像の優先度について考える。遠隔会議システムやプラント制御システムの例で述べたように、ユーザは興味のある映像を主に見ており、それ以外の映像はときどきしか見ない。さらに、その映像を詳しく見ているのではなく、全体の雰囲気や状況を把握することが主な目的である。このユーザニーズを満たしながら映像の情報量を削減するため、ユーザの興味のある映像は詳しく表示し、ユーザの興味のない映像は荒く（但し、全体の状況が理解できる）表示する手法を提案する。この手法を用いることにより、必要な映像を必要な量だけ表示することができ、限られた計算機リソースを効率良く使用することができる。

3. ウィンドウレイアウトによる興味度推定

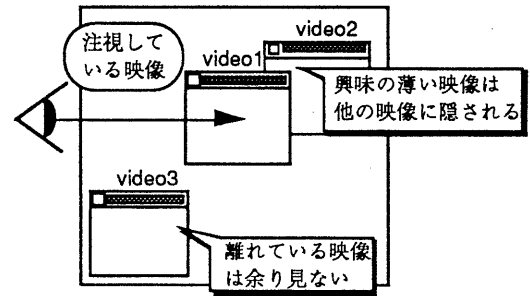


図1 ウィンドウレイアウト

3.1 ウィンドウレイアウト

画面上に複数の映像を表示する場合、図1に示すような、それぞれの映像を1つのウィンドウとして表示するビデオウィンドウを用いることが多い。ユーザは、このウィンドウを任意の位置に移動したり、ビデオウィンドウを重ねて表示したりして、興味にしがたって自由に映像をレイアウトする。

このウィンドウのレイアウトはユーザの興味度を反映している。例えば、ユーザは複数の映像を同時に扱う場合、それらのビデオウィンドウをできるだけ近くに配置する。これは、人間の視力が中央の物ほど良く、端に行くほど低下するため[5]、注視するビデオウィンドウが離れてしまうと、どちらかのウィンドウが良く見えなくなってしまうからである。したがって、映像に対するユーザの興味度は注視しているウィンドウに近いほど大きい、と考えられる。また、ユーザは興味のあるビデオウィンドウはできるだけ前面になるように配置する。その結果、興味のないビデオウィンドウ（例えば、video2）はユーザの興味の高いウィンドウにどんどん隠されてしまう。したがって、ユーザの興味度は、ビデオウィンドウの隠されている領域が大きいほど低くなると考えられる。

以上の検討から、ユーザの興味度をウィンドウ

のレイアウト状態から推定する、次のようなルールを仮定した。

- 1) ユーザの興味度は、注視しているウィンドウからの距離が大きいほど減少する。
- 2) ユーザの興味度は、他のウィンドウに隠されている領域が大きいほど減少する。

3.2 興味度関数

ユーザの興味度を計算するため、上記のルールを定式化する。ユーザが注視しているウィンドウ (y : 例えばvideo1) に対する、あるビデオウィンドウ (x) のユーザの興味度(DOI: Degree Of Interest)を以下の関数で表現する。

$$DOI(x;y)=R(x) \cdot D(x;y)$$

ここで、 $R(x)$ は、ビデオウィンドウ (x) の表示率、 $D(x;y)$ は注視しているウィンドウ (y) からビデオウィンドウ (x) への距離による興味度を示す。

$D(x;y)$ は、ユーザが注視しているウィンドウからの距離に関する興味度を表現しているおり、 $D(x;y)$ は、ビデオウィンドウ (x) がユーザの注視しているウィンドウ (y) から遠くなるに連れて小さくなる。ここでは、 $D(x;y)$ を以下のような指数関数で表現する。

$$D(x;y) = \exp(-d)$$

(d : ビデオウィンドウ x と

注視しているウィンドウ y との距離)

上記の式より、 $D(x;y)$ は、注視しているウィンドウ自身の時 ($d=0$) に最大値 1 となり、ユーザが注視しているウィンドウから離れるにしたがって値は小さくなる。

表示率 $R(x)$ は、ビデオウィンドウ全体の大きさ

に対する表示されている領域の比率である。この値は、ビデオウィンドウの表示されている領域が大きくなるほどその値は大きくなる。ビデオウィンドウのすべてが表示されている場合には、最大値 1 となり、すべての領域が他のウィンドウで隠されている場合には、最小値 0 となる。 $R(x)$ が最小値 0 の場合には、そのビデオウィンドウが表示されていないことを示しており、ユーザの興味度も 0 になる。

ユーザは、1つのウィンドウだけでなく、複数のウィンドウに興味を持っている場合がある。例えば、興味のある2つのウィンドウを見ようとすると、その間の映像も必然的に目に入ってくる。そのため、2つの興味のあるウィンドウの間の映像の興味度は、それぞれの興味あるウィンドウで計算した場合に比べて大きくなる。これを実現するため、それぞれの興味あるウィンドウに対する興味度をポテンシャルとして考え、映像のDOIをそれぞれの興味あるウィンドウに対する興味度DOIの和で表現する。

$$DOI(x) = \sum_{y=0}^n DOI(x;y)$$

4. 興味度によるリソース配分方式

計算機リソースをユーザの興味度にしたがって割り当てるため、推定した各映像の興味度 $DOI(i)$ に比例して各計算機リソースを割り当てる。映像を表示するための計算機リソースとして、ネットワーク、表示能力、画像圧縮/伸張能力などがある。この中で、ネットワークシステムの場合には、複数のデジタル映像の情報量の総計がネットワークの伝送容量を越えないように、ユーザの興味度に比例して、ネットワーク容量を表示する各映像に割り当てる。映像 i に対する許容伝送容量 $Capacity(i)$ は、全体の伝送容量 $TotalCapacity$ をユー

ザの興味度DOI(i)に対して比例して計算される。

$$\text{Capacity}(i) = \text{TotalCapacity} * \text{DOI}(i) / \sum \text{DOI}(i)$$

各映像の圧縮を行う場合には、この伝送容量以下になるように映像を圧縮して伝送する。

5. プロトタイプシステム

5.1 プロトタイプシステム

興味度による複数の映像表示方式の有効性を確認するため、プロトタイプシステムを作成した。本プロトタイプシステムでは、ハーフサイズの大きさ(320*240ドット)で映像を4つ表示する。本システムの処理能力は、各映像の情報量(画像の総画素数・フレームレート)に比例とし、ハーフサイズの映像を20こま(約10Mbps)で表示することができる。すべての映像の情報がハーフサイズの大きさである場合には、それぞれの映像の平均フレームレートは5こまとなる。

5.2 時間領域による情報削減

限られた処理能力の中で複数の映像を効率的に表示するため、まず、ユーザの興味度に応じて映像を時間領域で情報量を削減する方法を検討した。本システムでは、ユーザは、映像ウィンドウをマウスで選択することにより、興味のあるウィンドウを指定できる。その出力例を図2に示す。この図では、ビデオウィンドウのバー上に「FO」で示されているvideo0がユーザが興味ある映像である。

図2に示すように、注視している映像video0から離れているほど、各映像の興味度DOIは減少し、それに伴ってフレームレートも小さくなっている。(video1のフレームレートが3.5であるのに対し、

video2のフレームレートは1.2になっている。) video0から同じような距離でも他の映像(video1)に隠されているvideo3は、表示率が小さいのでフレームレートが小さくなっている。(video1のフレームレートが3.5であるのに対して、video3のフレームレートは1.9しかない。) ユーザの興味ある映像(video0)は、他の映像のフレームレートが小さくなった分、フレームレートを平均の5こまから13.4こまに上げることができている。

実際に時間領域だけで映像の情報量を削減してみると、興味のある映像は見やすいのだが、他の映像のフレームレートが小さすぎ、実際に動いているのかどうか分からない、という問題があることが分かった。

5.3 時間と空間領域による情報削減

ユーザの興味度により、時間だけでなく、映像の空間的な情報、つまり、画素数を変化させる。ここで、画素数を変化させるとは、映像をサンプリングする際に、間引いてサンプリングすることを意味し、表示する際には、所定の大きさ(320*240ドット)に拡大して表示する。例えば、縦方向の画素数を半分にする場合には、映像をサンプリングする際に1ライン毎に間引き、画素数を半分にする(320*120ドット)。表示する際には、縦方向に2倍に拡大して表示することにより、表示サイズを所定の大きさ(320*240ドット)にする。

映像の情報量が一定の中でフレームレートと画素数双方を変化させる場合には、フレームレートと画素数とはトレードオフの関係になる。本プロトタイプシステムでは、フレームレートと画素数のパラメータを以下の方法で決定する。ここで、画素数は、縦横両方向とも整数倍で変化する。

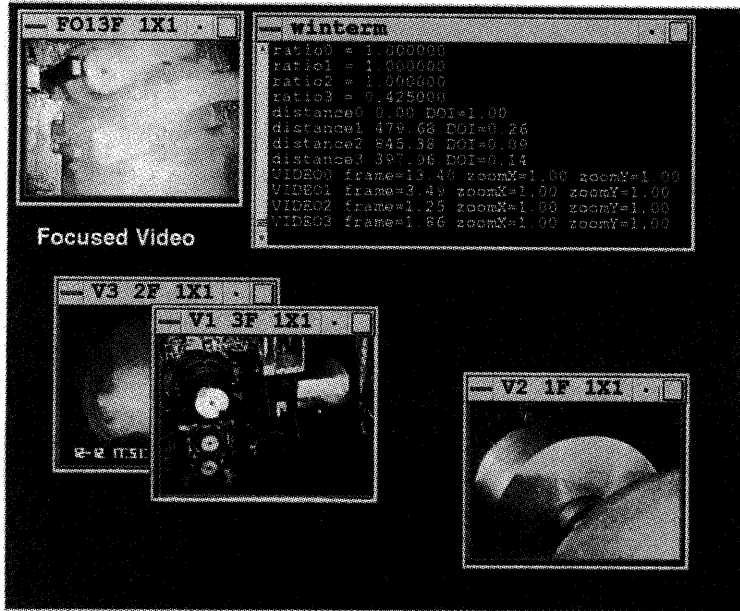


图2 出力例1



图3 出力例2

ステップ1：興味度DOIより時間領域だけで情報を削減する。

ステップ2：フレームレートが平均フレームレート5こま以上になるまで、横方向、縦方向の画素数を順次落としていく。

本方式で、前述の図2の例を空間的にも変化させた出力例を図3に示す。図3中、映像ウィンドウのバーには、フレームレートと縦横へのズーム率を示している。この出力例で示すように、ユーザの興味のある映像は画素数ならびにフレームレートとも高品質で見ることができる。また、図2の例ではフレームレートが極端に小さかったvideo2などの映像も画素数を下げることにより、相対的にフレームレートを上げることができ、見やすくなっている。

6. 今後の課題

6.1 興味のある映像の指定方法

本プロトタイプシステムでは、ユーザが興味のある映像を指定する方法として、マルチウィンドウなどで入力を行うアクティブウィンドウを指定する方法と同じ、ウィンドウをマウスで選択する方法を用いた。しかし、興味のある映像の指定方法として、上記のようなアクティブウィンドウ指定方式の他に、音声認識による指定方法やアイトラッカーなどの画像認識による注視点の検出など多数の方法があり、利用するシステムによって指定方法を工夫する必要がある。

6.2 音声の合成出力

複数の映像を表示する場合、音声をどのように合成して出力するかが問題となる。複数の映像を表示する際に、それぞれの映像の音声を単純に合

成したのでは、どの映像からどの音声が出力されているのか理解できない、ということである。

音声に対するユーザニーズを考えてみると、遠隔会議システムの場合、ユーザは話し相手の音声を主に聞きたいはずである。また、不意の呼びかけなどの注視していない映像の音声も重要である。これは、複数映像を表示する場合と同じユーザニーズであり、本手法を音声にも応用する。

本手法を音声に応用する方法として、以下の式のように、音声の大きさを興味度に合わせて変化させることができる。

$$\text{Sound}(i) = \text{Filter}(\text{DOI}(i), \text{Sound}(i))$$

ここで、Filterは、DOI(i)の大きさに合わせて音声の減衰率を制御する。Filterは、興味度DOI(i)の小さい映像の音声では、減衰率を大きくし、興味度の大きい映像の場合には、減衰率を小さくする。これにより、興味のある映像の音声は大きく聞こえ、興味のない映像の音声は小さくなる。このようにしておけば、遠隔会議システムにおいて、ユーザは、話し相手の音声は良く聞くことができ、別の相手から不意の呼びかけがあった場合でも、その声は通常より大きいので、ユーザはその音声を聞くことができる。この音声システムを次のプロトタイプシステムに反映して、その有効性を今後確認する。

6.3 ネットワークへの対応

実際の遠隔会議システムやプラント制御システムに応用する場合には、ネットワークが問題となることが多い。本プロトタイプシステムを図4に示すようなネットワークを用いたシステムに拡張する。この際、図に示すようなメディアサーバを実現する。メディアサーバは、映像をサンプリン

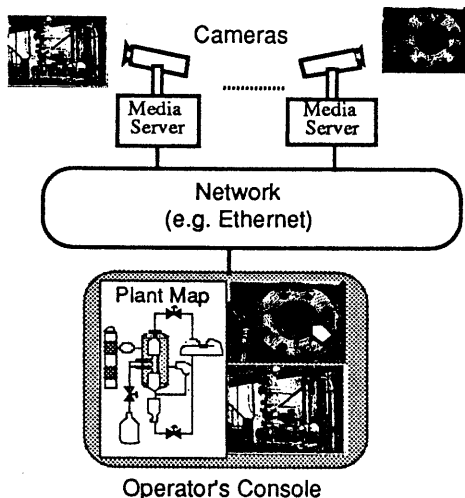


図4 ネットワークシステムへの適用

グし、それをユーザのしている計算機（クライアント）にネットワークを通して伝送する。

このシステムにおいて、クライアントは、映像の表示状態から各映像の興味度を推定し、それぞれの映像に割り当てるネットワークの許容量を計算する。クライアントは、伝送する映像の情報量が計算した割当許容量に収まるように、映像の各種パラメータ決定し、それをメディアサーバに設定する。メディアサーバは、設定されたパラメータに従って、映像をサンプリングし、映像圧縮し、クライアントに送信する。

メディアサーバは、複数のクライアントから共有されており、それぞれのクライアント毎に制御パラメータを管理する。さらに、もし、同じ情報を複数のクライアントから要求された場合には、UDP等の同報通信を利用して一度に送信する、といったメディア管理も行う。

7. まとめ

限られた計算機リソースで複数の映像を表示するため、ユーザの興味度により映像を表示する手法を提案した。本方式は、まず、ユーザの興味度を映像の表示状態から推定し、その推定した興味度に応じて限られた計算機リソースを配分し、効率的に映像を表示する。

本方式を用いたプロトタイプシステムを開発し、その有効性を確認した。

1) 本方式を用いることにより、ユーザの興味のある映像を高品質に表示できる。

2) 他の映像も画質を落としてはいるものの、常に更新表示されるため、ユーザは全体的な状況を把握することができる。

今後、本方式をネットワークを用いた映像伝送システムに応用していき、その有効性と問題点を検討していく。

8. 参考文献

- [1] A. Borning and M. Travers, Two Approaches to Casual Interaction over Computer and Video Networks, CHI'91 Conference Proceeding (New Orleans, April 2 -May 2 1991), ACM Press, pp. 13-19
- [2] Gregory K. Wallace, The JPEG Still Picture Compression Standard, Communicatins Of The ACM, (April 1991, Vol.34 No.4), ACM Press, p30-p44
- [3]原島 博、画像情報圧縮、1991、オーム社
- [4]安田 浩編、マルチメディア符号化の国際標準、1991、丸善
- [5]野呂 影勇、図説エルゴノミクス、1990、日本規格協会