パノラマ映像生成機能を持つテレビ会議用ビデオカメラの開発に関する研究

米本良*，上杉徹，川村尚生，菅原一孔（鳥取大学）

Video Camera for Teleconferecing with a Panoramic Image Generation Function

Ryo Yonemoto*, Toru Uesugi, Takao Kawamura, Kazunori Sugahara (Tottori University)

Abstract

In most of the current teleconferencing systems, a conference room is taken in one fixed video camera. For that reason, a face image of speakers becomes small. To take both full view images of a conference room and face images of speakers, it is required to develop a multi camera system with human operations. In this paper, we develop a new video camera for a teleconferencing without human operations in above mentioned situations. The proposed video camera outputs the images which composed with panoramic images generated with images from three NTSC video cameras and the facial images of speakers. In this paper, the Fast Robust Correlation technique proposed by Fitch et al. is adopted and its hardware realization is mentioned.

キーワード：パノラマ映像，高速フーリエ変換
(Panoramic image，Fast Fourier Transform)

1. はじめに

最近ではネットワークの通信速度の高速化や高性能の映像撮影装置の低価格化に伴い，テレビ会議が開催される場面が増えてきている．ところが一般のテレビ会議では，1台のビデオカメラで会場全体を撮影するため，撮影された映像から話者の表情を読み取ることが困難である．また，逆に話者の表情を読み取ろうとして話者が大きく映るように撮影すると，会場の狭い範囲しか撮影できなくなるため会場の雰囲気を知ることが困難になるなどの問題がある．これらの問題を解決するため，複数台のビデオカメラで会場全体の映像と話者の映像を別々に撮影する場合もあるが，複数のビデオカメラを用意するにはコストがかかる上にその切替操作には人手がかかるなどの問題が出てくる．そこで，我々は会議会場全体の広範囲の映像を表示するためのパノラマ映像と，話者を映している映像をカメラ内部で合成したものを映像信号として出力するテレビ会議用ビデオカメラを開発している．このような装置を構成する際には，話者追従の手法と共に複数の映像からのパノラマ映像の合成手法が重要な機能として上げられる．本稿では，パノラマ映像を合成する手法として，Alistair J. Fitch らが提案した Fast Robust Correlation 手法 [1] を取り上げ FPGA を用いてそのハードウエア実現 [2] [3] を試みた．

2. システムの構成

システムは，3 つの NTSC ビデオカメラを用いた映像撮影装置，NTSC ビデオデコーダ LSI，メモリ，デジタルビデオエンコーダを実装したビデオ信号入出力ボード，FPGA ボード，モニタから構成されている．構成している各装置を図 1 に示す．そして，FPGA の仕様を表 1 に示す．

メモリは 3 バンク実装している．そして，各バンクは 1 アドレスに 24 ビットの保存が可能であり，RGB 各 8 ビットの 1 画素分のカラーデータを 1 つのアドレスに保存することができる．そして，リアルタイムで映像処理を行うために 3 バンクを用いてバンク切り替えを行っている．これにより，１つのバンクにアクセスが集中することで起こる

表 1　FPGA の仕様
Table 1.　FPGA specifications

開発元	ALTERA Co.Ltd.
型番	EP1C20F400C7
ロジックエレメント数	20,060
最大 I/O ポート数	301
パッケージ	400-Pin FineLine BGA
サイズ	21 × 21 [mm]
動作周波数	48[MHz]

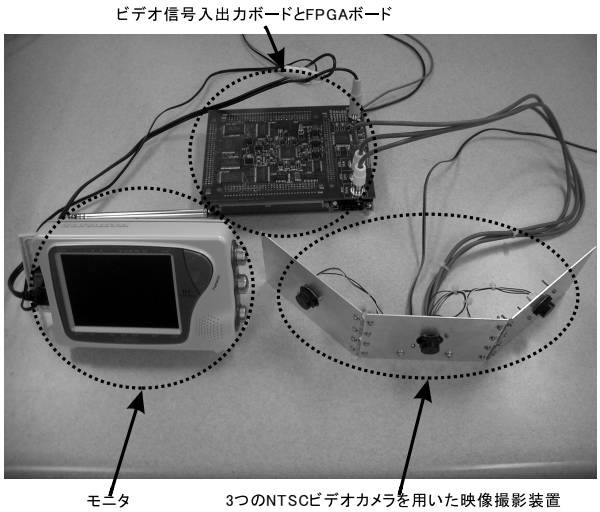

図 1　装置
Fig. 1.　System configuration

処理の待ち時間を効率よく防ぐことができる．各部での 1 フレーム分の処理が終了するのに同期してバンク切り替えを行った．バンク切り替えを行っている様子を図 2 に示す．

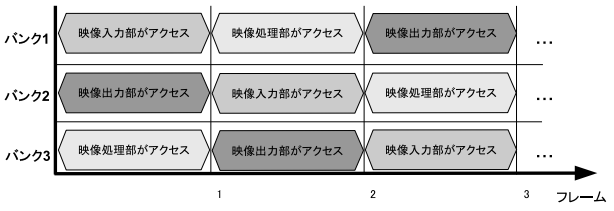

図 2　バンク切り替え
Fig. 2.　Bank change

図 2 に示した処理の流れは以下のとおりである．バンク 1 には映像入力部がアクセスし，NTSC ビデオカメラから送られてきた NTSC アナログ映像信号をデコーダによってデジタルの RGB 各 8bit のデジタル映像信号に変換し，ビデオ信号入出力ボード上のメモリに蓄積する．同時に，バンク 2 には映像出力部がアクセスし，FPGA 上の各回路で処理を行ったデータをエンコーダで NTSC アナログ映像信号に変換しモニタに表示する．バンク 3 には映像処理部がアクセスし，FPGA 上の各回路においてメモリに蓄積したデータを基にパノラマ映像生成の処理を行う．そして，1 フレーム分の処理が全てのバンクで終了したら，次に，バンク 1 には映像処理部，バンク 2 には映像入力部，バンク 3 には映像出力部がアクセスし処理するというように，バンク切り替えを行いながら処理を行う．

3.　パノラマ映像生成

パノラマ映像生成では，会議会場全体の広範囲の映像を表示するために，視野角の違う 3 つの NTSC ビデオカメラで撮影した入力映像を基にパノラマ状の映像を生成する．しかし，撮影した入力映像の映像サイズ (640×480[pixel]) のままでは，パノラマ映像を生成する際，図 3 に示すように想定している出力映像の映像サイズ (640×480[pixel]) より大きくなり，パノラマ映像を生成することはできない．そこで，それを防ぐために入力映像を 3/8 サイズの 240×180[pixel] に縮小し一部分を重ねて合成し，640×180[pixel] で表示することにした．補正処理を行う前のパノラマ映像の生成の様子を図 4 に示す．

映像サイズの縮小方法として，まず，撮影した入力映像をメモリに書き込む時に，1 画素ずつ飛ばして書き込むことで，入力映像の映像サイズを 1/2 に縮小する．さらに，入力映像を 3/4 サイズにするために，パノラマ映像を生成する際に縮小前の映像から画素を縦横方向にそれぞれ 4 画素に 1 画素間引くという手法を用いた．またパノラマ映像の生成方法は，入力映像データをメモリの空き領域にコピーすることによって行われる．まず，映像サイズを縮小しながらメモリの決められたアドレスに 3 つの映像データの横 1 行分をコピーする．続いて 2 行目，3 行目とコピーするが，縦方向も 4 回に 1 回行を飛ばして縮小しながら全映像をコピーしてパノラマ映像を生成する．

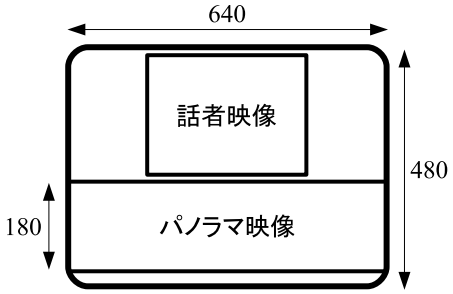

図 3　出力映像
Fig. 3.　Output image

4.　パノラマ映像生成の自動補正処理

〈4•1〉 自動補正処理　自動補正処理とは，上記で述べたようにパノラマ映像を生成する際に，映像同士の最適な結合位置を自動で検出し，最適な位置に映像を補正する処理である．この補正処理を図 4 に示す入力映像①と入力映像②，入力映像②と入力映像③の各々の境界領域で行う．

〈4•1•1〉 Fast Robust Correlation 手法[(1)]　今，図 5 に示すように，画像 g を画像 f に対して x 軸方向に m，y 軸方向に n ずらすと仮定する．このとき，$\boldsymbol{x}$,$\boldsymbol{m}$ を $\boldsymbol{x} = (x, y)$,$\boldsymbol{m} = (m, n)$ と定義すると，2 つの画像 f, g の画素差は以下のように表すこととする．

$$R = f(x) \cdot \alpha_f(x) - g(x-m) \cdot \alpha_g(x-m) \cdots\cdots (1)$$

また，2 つの画像 f, g の全体としての画像の差は次のようになる．

$$R(m) = \sum_x h\left(f\left(x\right) - g\left(x-m\right)\right) * \alpha_f(x) \cdot \alpha_g(x-m) \cdots\cdots\cdots\cdots\cdots (2)$$

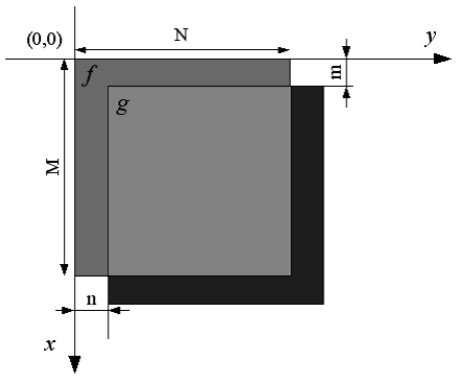

図 5　画像比較
Fig. 5.　Image comparison

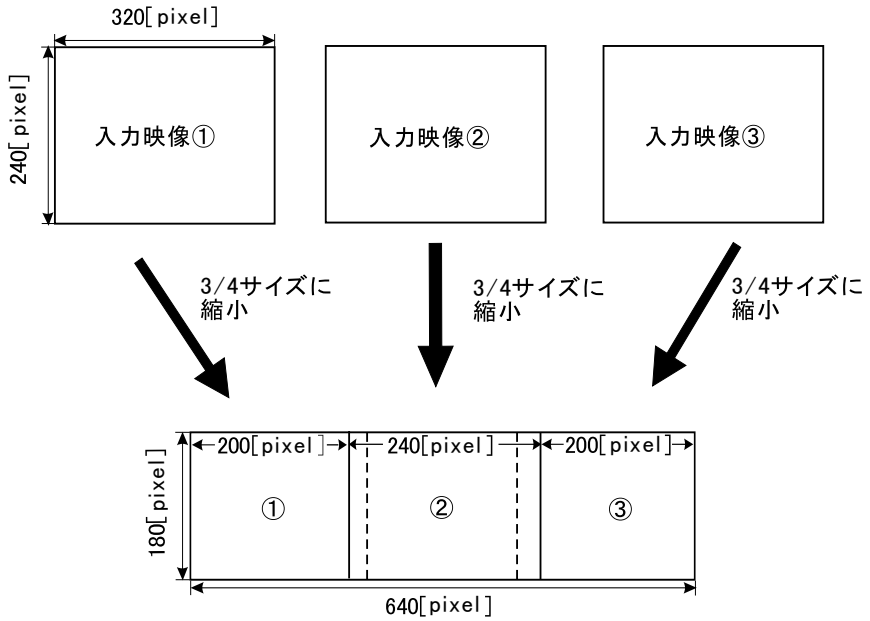

図 4　補正前のパノラマ映像を生成する様子
Fig. 4.　Panoramic image generation

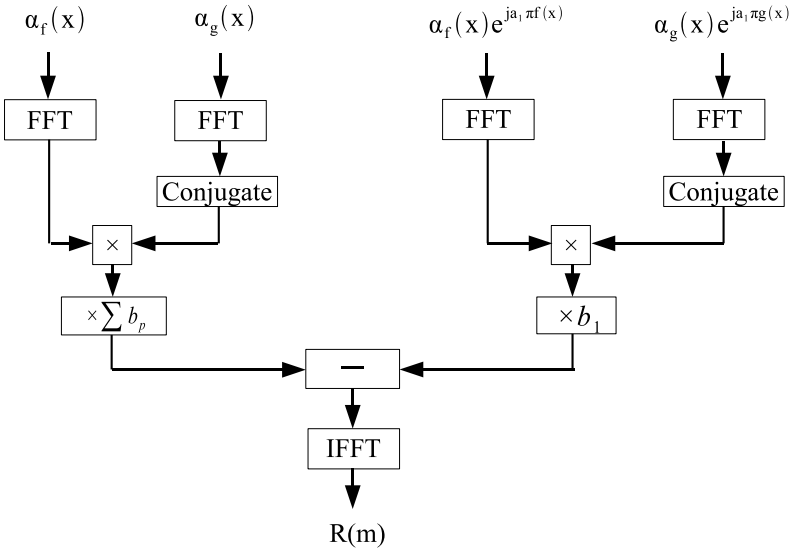

図 6　画像比較の流れ図
Fig. 6.　Flow chart of image comparison

ここで，式 (1) と式 (2) に示している $\alpha_f(x)$ と $\alpha_g(x)$ はアルファマスクと呼ばれ，任意の形状の領域に柔軟に誤差評価の際の重みを持たせることができる．本研究では画像全面に対して比較するので，$\alpha_f(x)$，$\alpha_g(x)$ 共に，全面が 1 の値を持つものとする．また，式 (2) 中の，各画素の誤差 r に対する評価関数 $h(r)$ は，偶関数である cos 関数を用いて式 (3) で与える．

$$h(r) \approx \sum_{p=1}^{P} b_p \left(1 - \cos\left(a_p \pi r\right)\right) \cdots\cdots\cdots\cdots\cdots\cdots \quad (3)$$

ここで，式 (3) は b_p と a_p というパラメータを含んでいるが，$P = 1$ の場合には $b_1 = 1$, $a_1 = 1.0$ となる．式 (3) を式 (2) に代入して整理すると

$$R(m) = \sum_{x}[(\alpha_f(x) \cdot \alpha_g(x-m)) \sum_{p=1}^{P} b_p(1 - \cos(a_p\pi * (f(x) - g(x-m))))] \cdots\cdots\cdots \quad (4)$$

となる．これを整理すると

$$R(m) = \sum_x [(\alpha_f(x) \star \alpha_g(x)) \sum_{p=1}^{P} b_p - \sum_{p=1}^{P} b_p((\alpha_f(x)e^{ja_p\pi f(x)}) \star (\alpha_g(x)e^{ja_p\pi g(x)}))] \cdots\cdots\cdots\cdots\cdots\cdots (5)$$

の関係を得る．ただし，式 (5) 中の $\star$ は畳み込み演算を示している．

〈4•1•2〉 周波数領域での画像比較　式 (5) を空間領域で直接求めずに，空間周波数領域で間接的に求めることにする．また，その計算の流れを図 6 に示す．まず，f, g には RGB 成分の R 値を与え，上の 4 つの項に対してそれぞれ二次元 FFT [4] を行う．結果を F, G とし FG^* を求める．ただし，G^* は G の複素共役を表す．その後，$\times \sum b_p$ や $\times b_1$ という演算があるが，上記で述べたように，$P = 1$ として $b_1 = 1$ なので，そのまま差を取る．そして最後に，その差の値に対して二次元 IFFT(inverse FFT) を行い，$R(\boldsymbol{m})$ を得る．この得られた $R(\boldsymbol{m})$ について最小値を求め，最小値となる (m, n) により，画像 f, g の合成画像となるよう画像 g を移動させることで自動補正処理を行う．また，画像サイズ $N \times M$[pixel] の画像を FFT を用いずに比較した場合の計算量は $O(M^2N^2)$ であるが，FFT を用いて周波数領域で処理を行うと計算量は $O(MN \log_2 N)$ となる．

〈4•2〉 FFT のハードウエア化

〈4•2•1〉 バタフライ演算　図 6 に示すとおり，本手法は FFT の処理が中心的な役割を果たす．FFT をハードウエア化するために，バタフライ演算をハードウエア化する必要がある．バタフライ演算とは，図 7 に示す計算処理である．

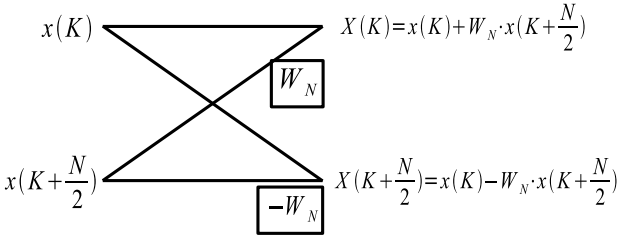

図 7　バタフライ演算
Fig. 7.　Butterfly operation

バタフライ演算をハードウエア化するにあたり，記述方法により FPGA 上に形成されるハードウエア量が大きく異なるので，その記述には注意が必要である．

本稿で扱うデータはすべて浮動小数点の IEEE の 754 規格で，仮数部 23 ビット，指数部が 8 ビット，符号ビット 1 ビットの合計 32 ビットで表現されている．FPGA 上でこのような浮動小数点の計算を行うために，ALTERA 社の QuartusII がもつ Mega Wizard の機能により，乗算器，加減算器を実現した．しかし，メモリの 1 アドレスに保存できるのは最大 24 ビットである．そこで，データをメモリに書き込む際には上位 24 ビットを書き込み，残りの 8 ビットは切り捨てることにする．また，データをメモリから読み込む際には読み込んだ 24 ビットのデータの後に 8 ビット分の 0 を付加することで 32 ビットにする．

バタフライ演算の流れは，以下のとおりである．図 8 にバタフライ演算の様子を示す．

（1） 複素数 W_N の値を予めレジスタ W_{Nr} と W_{Ni} に保存しておく．
（2） メモリから 2 つのデータとして，それぞれの実部虚部合せて 4 つのデータを読み込み FPGA 内のレジスタ x_0, x_1, x_2, x_3 に保存する．
（3） これらのデータから，$(x_2 + jx_3)(W_{Nr} + jW_{Ni})$ を求め，一時レジスタ t_0, t_1 に保存する．
（4） $(x_0 + jx_1) + (t_0 + jt_1)$ を求め，その結果をレジスタ X_0, X_1 に保存する．
（5） レジスタ X_0, X_1 の内容を，定められたメモリのアドレスに格納する．
（6） $(x_0 + jx_1) - (t_0 + jt_1)$ を求め，その結果をレジスタ X_0, X_1 に保存する．
（7） レジスタ X_0, X_1 の内容を，定められたメモリのアドレスに格納する．

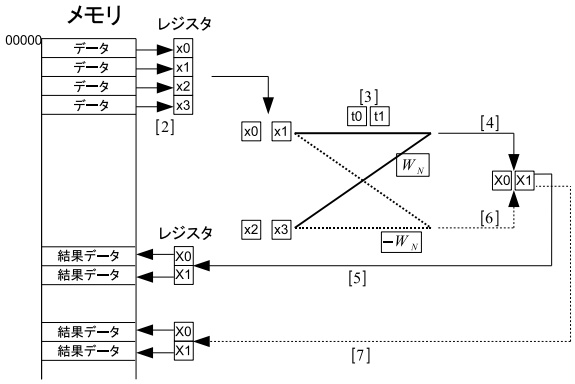

図 8　バタフライ演算の流れ ([ ] 内は本文中の処理の番号)
Fig. 8.　Flow of butterfly operation

このように 1 つのバタフライ演算につき，実数の乗算 4 回と実数の加減算が 6 回必要である．1 回の乗算には 5 クロック，1 回の加減算には 8 クロック必要である場合には 68 クロックで 1 つのバタフライ演算を実行できる．

M 点の 1 次元 FFT を実装するためには $\frac{M}{2} log_2 M$ 回のバタフライ演算が必要である．さらに $M \times N$ 点の 2 次元 FFT を実現するためには $\frac{MN}{2}(log_2 M + log_2 N)$ 回のバタフライ演算が必要となる．本稿では後述するように 128×128 の大きさの画像の合成を行なうので，114688 回のバタフライ演算が必要となる．

図 10　パノラマ映像生成の結果
Fig. 10.　Generated panoramic image

(a) 入力映像①

(b) 入力映像②

(c) 入力映像③

図 9　入力映像
Fig. 9.　Input images

5. 実　　験

〈5•1〉 実験方法　　3 つの NTSC カメラより撮影した入力映像①と入力映像②，入力映像②と入力映像③の映像を各々境界領域で比較し合成する．上述のとおり 1 つの入力映像の大きさは 320 × 240[pixel] であるが，FFT を用いる場合，映像サイズを 2 のべき乗することを考慮して，本システムでは映像サイズとして，128 × 128[pixel] とする．ただし，x 方向には境界側から 128 点の領域を合成に利用するものと考え，y 方向には中心から上下に 64 点，計 128 点の領域を合成に利用するものと考えた．

〈5•2〉 実験結果　　実際に 3 つの NTSC ビデオカメラの入力映像図を図 9 に示す．そして，図 9 の入力映像について，パノラマ映像生成の結果を図 10 に示す．使用した各回路規模を表 2 に示す．また，使用可能なロジックエレメント数の約 65 %で実現できた．そして，パノラマ映像生成でかかるクロック数は約 1.1×10^8 である．その結果，パノラマ映像生成には約 2.3 秒の処理時間が必要となる．

表 2　回路規模
Table 2.　Circuit scale

処理回路名	ロジックエレメント数
映像入力制御部	219
パノラマ映像生成部	12,409
映像出力制御部	329
メモリ等	252
合計	13,209

6. おわりに

本研究では，テレビ会議用ビデオカメラの開発において，重要な機能の 1 つであるパノラマ映像生成の自動補正処理を行うために Alistair J. Fitch らが提案した Fast Robust Correlation 手法のハードウエア実現を試みた．

参考文献

（1） Alistair J. Fitch, Alexander Kadyrov, William J. Christmas and Josef Kittler : Fast Robust Correlation,IEEE (2005)

（2） 長谷川 裕恭 : VHDL によるハードウェア設計入門，CQ 出版社 (2002)

（3） 森岡 澄夫 : HDL による高性能ディジタル回路設計，CQ 出版社 (2002)

（4） 大浦 拓哉 : FFT (高速フーリエ・コサイン・サイン変換) の概略と設計法 : http://www.kurims.kyoto-u.ac.jp/~ooura/fftman/index.html