

輪郭適応パッチを用いたワーピング予測ビデオ符号化方式

宮本義弘 横山 裕 太田 睦

日本電気(株) 情報メディア研究所

〒216 川崎市宮前区宮崎 4-1-1

E-mail: miyamoto@DSP.CL.nec.co.jp

低ビットレートでのビデオ符号化において、オブジェクト階層符号化 (Object Scalability) を実現する方式を提案する。ISO/IEC MPEG では次世代低ビットレートビデオ符号化方式 (MPEG-4) の標準化を進めているが、従来の方式に比較して符号化効率を改善するだけでなくマルチメディアデータアクセスに適した符号化技術の開発を目標としている。例えばオブジェクト階層符号化の技術により、画像中の注目する被写体領域だけを符号化伝送したり、圧縮データのままの状態でも領域単位での画像の合成や編集などが可能になる。提案方式はワーピング予測を基本方式とし、ワーピング処理の単位であるパッチを注目する被写体領域の輪郭に適合させることで、オブジェクト階層符号化を実現する。領域形状はパッチにより近似表現するので、形状符号量は僅かですむ。ワーピング予測は予測効率に優れ、予測による歪が視覚的に目立ち難いので低ビットレート符号化に適している。またパッチ形状に合わせた可変形状 DCT を用いることで予測誤差を効率的に変換符号化する。計算機シミュレーション結果により方式の効果を示す。

Warping Video Coding using Contour Adaptive Patch Structure

Yoshihiro Miyamoto, Yutaka Yokoyama and Mutsumi Ohta

Information Technology Research Laboratories, NEC Corp.

Miyazaki 4-1-1, Miyamae-ku, Kawasaki 216 JAPAN

This paper presents a video coding method for realizing object scalability at low bitrate. The proposed method employs contour fitting patch structure. An object region is defined with selected patches, and it is independently encoded to other objects. Owing to the scheme, object scalable functionalities, such as selective object decoding, object manipulation etc., are obtained. Each patch is encoded using warping prediction and variable shaped DCT. The warping prediction is a motion compensation method suitable for low bitrate coding, because it does not cause block distortions. The variable shaped DCT transforms the prediction error which have arbitrarily shapes corresponding to the patch. Experimental results indicate the performance of the proposed method.

1. はじめに

ISO/IEC MPEGではこれまでにビデオ符号化の国際標準化方式として、MPEG-1、MPEG-2の規格化を進めてきた [1],[2]。現在は、次世代低ビットレート符号化方式 MPEG-4の検討を開始したところである。MPEG-4では低ビットレートでの応用に適するように符号化効率を改善する事はもちろん、従来の規格ではカバーしていない新機能の実現も目標としている。すなわち符号化方式の柔軟性を高め、マルチメディアデータアクセスにおける様々な要求に応える符号化技術を開発することが期待されている。これを実現する手段の一例としてオブジェクト階層符号化の技術が注目されている。オブジェクト階層符号化とは画像中の注目する被写体(オブジェクト)を互いに独立して復号できるように階層符号化する方法である。こうすることで特定のオブジェクトの圧縮データの内容を修正したり、圧縮データのままで複数のオブジェクトから新しい画像を合成できる。

ここでMPEG-1、MPEG-2などの従来の符号化方式でオブジェクト階層符号化を実現する方法を考えてみる。従来の方式は何れもブロックベースの動き補償フレーム間予測方式を採用している。このためオブジェクト階層符号化を実現するには、ブロック単位で画像を領域分割する。例えばMPEG-1ではGeneral Slice構造を利用し、注目するオブジェクトに係るマクロブロックだけを符号化し、他はスキップする。

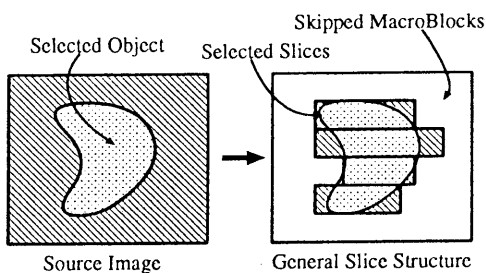


Fig.1 Object Scalable Coding using MPEG-1 Slice Structure

しかしこの方法では任意形状のオブジェクトをブロック単位で切り出すため、同時に多くの不要な画像データが切り出される。またオブ

ジェクトの動きに伴い、この不要な画像データの内容と大きさはフレーム間で変動する。これはフレーム間予測符号化の効率を阻害する大きな原因となる。

そもそもブロックベースの動き補償フレーム間予測方式ではブロック境界で不連続な予測誤差、いわゆるブロック歪を発生する。そこで我々はこれを解決する方法としてワーピング予測を検討してきた [3]-[5]。本稿では更に、ワーピングの処理単位であるパッチを、注目するオブジェクトの輪郭に適合させることで、オブジェクト階層符号化を実現する方式を検討した。輪郭適合させたパッチはオブジェクト間にまたがる部分が少なくなる。このため注目するオブジェクトに係わるパッチだけを切り出せば、効率よく符号化できる。また予測誤差の変換符号化もオブジェクト内で処理が完結するように、可変形状 DCT [6],[7] により行った。

2. 方式の説明

2.1 階層符号化処理の構成

図1にオブジェクト階層符号化を実現する場合のブロック構成を示す。

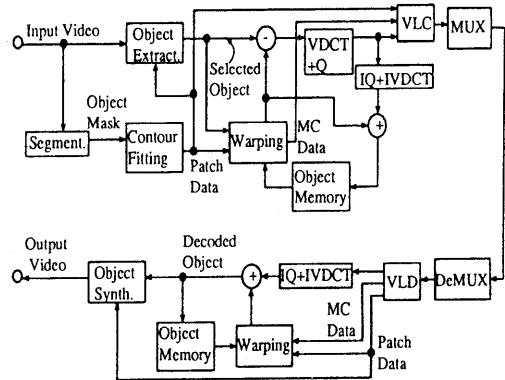


Fig.2 Block Diagram

処理の流れを以下に説明する。

- 入力画像を領域分割し、注目するオブジェクトの輪郭線を抽出する。

- パッチの辺がオブジェクトの輪郭線に適合するように可変形状のパッチ配置を定める。
- パッチの中で注目するオブジェクトを蔽うものだけを有効パッチとみなし、入力画像から当該パッチの画像データだけを抽出する。
- 抽出した画像データをその他の画像部分とは独立して符号化する。まずワーピングにより動き補償フレーム間予測し、予測誤差を可変形状 DCT で符号化する。

オブジェクトの形状は領域単位で表現する方法 [8],[9] が広く検討されているが、今回はワーピング予測との組み合わせを考慮して可変形状パッチによる方法を採用した。個々のパッチは異なるオブジェクト間にはできるだけ跨らないように、輪郭線を考慮して配置する。こうすることでパッチ毎に何れのオブジェクトに属するかを分類できる。なおパッチが小さいほどオブジェクト形状を高精度に表現できるが、付加符号量は反比例して増加する。このためパッチのサイズは符号化レートとのトレードオフで決定する。

オブジェクト階層化の必要が無く、入力画像を全体で符号化する場合には全てのパッチを有効とする。この場合にも被写体構造を反映した可変形状パッチでワーピング予測を行なう方が予測効率が良くなる [5]。

2.2 輪郭適応パッチの生成

パッチのサイズや配置は、注目するオブジェクトの内容に従い最適に定めるのが望ましい。しかし低ビットレートの符号化システムでは、パッチ構造の再現に要する付加符号量を抑制する必要がある。そこで基本パッチ構造を予め定めた範囲内で変形し、オブジェクトの輪郭形状に適合させることとした。Nieweglowski ら [10] は正方格子のパッチ構造から出発し、パッチの頂点をそれぞれ独立に輪郭適合させる方法を検討している。本稿では更に、複数のオブジェクト間に跨るパッチの発生を抑制するために、三角格子のパッチ構造から出発し、パッチの辺が輪郭線に沿うように再配置した [5]。以下にパッチの修正手順を説明する。

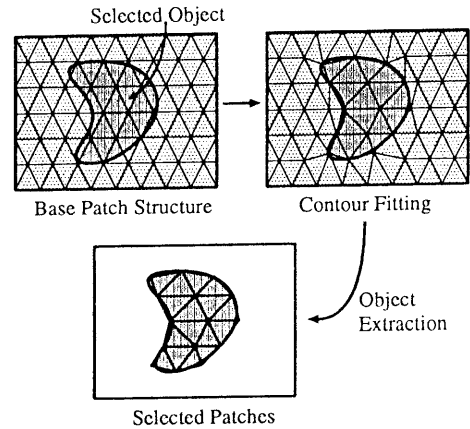


Fig.3 Contour Fitting Patch Structure

まず輪郭線からの距離 $d(x, y)$ に反比例したポテンシャル場を作る。入力画像上の画素位置 (x, y) におけるポテンシャルの値 $P(x, y)$ を次のように定めた。

$$P(x, y) = -\frac{A}{d(x, y)^2 + 1} \quad (1)$$

+1 は分母を 0 にしないための項、 A は適当な定数である。輪郭に近い程強く、且つ輪郭から離れた位置では影響が少なくなるよう、距離の 2 乗に反比例する形式とした。次に基本パッチ構造である三角格子をポテンシャル場に当て、格子のエネルギーが最小となるように格子点 (パッチの頂点) の位置を変位させた。このとき格子の辺が輪郭線に沿うように、辺単位でエネルギーを評価した。

なおフレーム間予測符号化を行う場合には、参照フレームでのパッチ構造を現フレームでの初期パッチ構造とした。こうすることで輪郭適合の収束を早め、フレーム間でのパッチの対応を安定化できる。

2.3 ワーピング予測

ワーピングとは画像の幾何学的変型処理の一種で、マッピングやモーフィングによる画像の合成や特殊効果に応用されている [11]。またワーピングを動き補償フレーム間予測に適用する方法も幾つか報告されている [10],[12]。本稿では入

力画像を三角パッチに分割したので、アフィン変換によりワーピング予測を実現した。

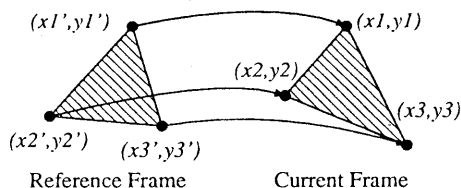


Fig.4 Warping Prediction

図4にアフィン変換によるワーピング予測の概念を示す。アフィン変換は6つのパラメータ $a \sim f$ により (x, y) から (x', y') への座標変換を表現する。

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix} \quad (2)$$

現在の入力画像上の三角パッチの3頂点の位置を $(x_1, y_1), (x_2, y_2), (x_3, y_3)$ とし、これに対応する参照画像上での3頂点の位置を $(x'_1, y'_1), (x'_2, y'_2), (x'_3, y'_3)$ とすると、

$$\begin{pmatrix} x'_1 & x'_2 & x'_3 \\ y'_1 & y'_2 & y'_3 \end{pmatrix} = \begin{pmatrix} a & b & e \\ c & d & f \end{pmatrix} \cdot \begin{pmatrix} x_1 & x_2 & x_3 \\ y_1 & y_2 & y_3 \\ 1 & 1 & 1 \end{pmatrix} \quad (3)$$

が成り立つ。この式からパラメータ $a \sim f$ を得る。三角パッチ内の任意の画素の対応位置は、パラメータ $a \sim f$ を用いた計算により求める。従来のブロック単位の動き補償予測は、式(2)で e, f だけが有効な場合に相当し、画像平面上の平行移動しか補償できない。ワーピング予測では式(2)の $a \sim d$ により回転や拡大・縮小の動きも補償できる。このため予測効率が改善できる。また画素毎の動きベクトルの値は、パッチの境界部分でも連続的に補間されるのでブロック歪が発生しない。

一方、ワーピング予測ではパッチ内部に動きの異なるオブジェクトが混在するとオブジェクト境界部分で不自然な伸縮歪が発生する[3]。本稿ではパッチがオブジェクト間を跨がないように輪郭適合させており、こうした歪を抑制する効果もある。

2.4 予測誤差の変換符号化

ワーピング予測誤差は可変形状 DCT により変換符号化する。以下にその理由を説明する。オブジェクト階層符号化では、動き補償フレーム間予測誤差もオブジェクトの有効領域内で発生する。このため通常のブロックベースの DCT で変換符号化したのでは必ずしも十分な変換効率が得られない。図5にその例を示す。

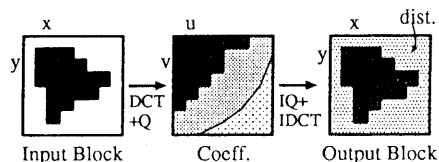


Fig.5 Block Based DCT

オブジェクト境界に跨るブロックでは、予測誤差の有効な画素と無効な画素とが混在する。しかし DCT 変換係数上では無効成分の判別がつかないため、符号化する係数の数が元の有効画素数より多くなることがある。また無効画素に起因する変換係数の量子化歪は、逆 DCT によりブロック全体に拡散する。無効画素部分ではこの歪をオブジェクト境界に従い切り捨てれば良いが、有効画素部分に重畳した歪は画質劣化となる。

次に今回の検討で用いた2種類の可変形状 DCT について説明する。

2.4.1 固定ブロックベース可変有効領域の DCT(SA-DCT)

この方法は通常の DCT と同様に、固定サイズのブロック単位で可変形状 DCT を行う。すなわちワーピング予測と処理の単位が異なることになる。

まずオブジェクトの予測誤差画像を固定サイズのブロックに分割する。オブジェクトの内部のブロックは通常の DCT で変換する。オブジェクトの境界部分に跨るブロックでは、有効画素の予測誤差だけを可変形状 DCT で変換する。固定ブロック内の有効画素だけを変換符号化する方法としては、KL 変換や Schmit の直交変換が

知られている [13],[14]。これらは変換効率は高いが、領域形状に従い変換基底を算出するため演算量が膨大になるという問題がある。そこで本稿では松田ら [6],[7] が提案した簡易な可変形状 DCT を用いた。すなわちブロック内の行と列毎に、各々の有効画素数に合せた 1 次元 DCT 変換 (SA-DCT) を行う。

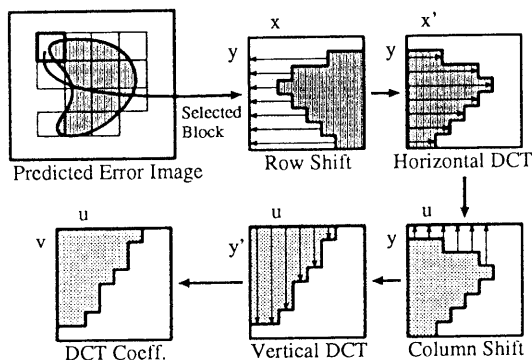


Fig.6 Shape Adaptive Block-DCT

図 6 を参照し、ブロック内の有効画素だけを変換符号化する手順を説明する。

- ブロック内の有効画素データを水平方向にシフトし左辺に詰める。
- 行毎に可変長の 1 次元 DCT を行う。
- 1 次元 DCT して得た変換係数を垂直方向にシフトし上辺に詰める。
- 列毎に可変長の 1 次元 DCT を行い、最終的な変換係数を得る。

変換における行列の順序は入れ替えても良い。なお、前述の 1 次元 DCT は次のように定義する。例えばある行 (列) の有効画素の値が $(p_0, p_1, p_2, \dots, p_{N-1})$ であれば、この行 (列) に対する 1 次元 DCT は、

$$C_u = \frac{2a_u}{N} \sum_{k=0}^{N-1} p_k \cos\left(\frac{(2k+1)u\pi}{2N}\right), \quad (4)$$

$$a_u = \begin{cases} \frac{1}{\sqrt{2}}, & \text{for } u = 0, \\ 1, & \text{for } u > 0. \end{cases} \quad (5)$$

である。この 1 次元 DCT に対する逆変換は以下のように定義される。

$$p'_i = \sum_{u=0}^{N-1} a_u C_u \cos\left(\frac{(2i+1)u\pi}{2N}\right). \quad (6)$$

SA-DCT では変換演算に無効画素が含まれないため、変換係数の数は有効画素数以上には増えない。ただし行・列毎に有効画素数に応じて変換基底を切替えるため、通常の DCT よりも変換処理は若干複雑になる。

2.4.2 可変形状パッチベースの DCT (Pch-DCT)

前節で説明した SA-DCT に対し、ワーピング予測処理と同じパッチ単位で可変形状 DCT を行う方法 (Pch-DCT) を説明する。Pch-DCT を用いれば、図 1 におけるワーピング予測から可変長符号化までをパッチ単位で一貫的に処理できる。

まず選択したパッチを長方形の擬似ブロックに当てはめる。擬似ブロックは前節の SA-DCT と同じ手順で変換符号化する。このとき、パッチを包含する擬似ブロックの中で、パッチ内の画素のみ有効画素とし、パッチの外部は無効領域とする。またパッチを順次選択し、別々に変換符号化する際に、隣接するパッチ間で境界線上の画素が二重に符号化されないように、いずれかのパッチにのみ帰属させる。

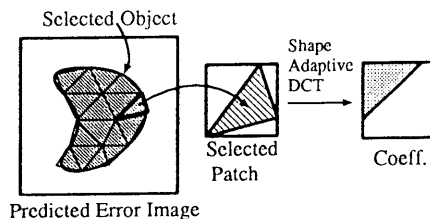


Fig.7 Patch Based DCT

Pch-DCT ではパッチの大きさに伴ない擬似ブロックのサイズも変動するため、変換処理は SA-DCT よりも複雑になる。なお、ワーピング予測ではパッチ単位で予測モードを制御できる。Intra / Inter 判定などにより予測誤差分布がパッチ単位で変化する場合には、変換符号化もパッチ単位とした方が良い。

3. シミュレーション

本方式の性能をシミュレーションにより評価した。なお現在の画像解析技術では、背景が複雑な画像から注目するオブジェクトを正確に抽出することは困難である。抽出したオブジェクト形状がフレーム間で不安定だと、符号化効率を阻害する。そこで今回は、予め画像から抽出したオブジェクトのマスクパターンを利用した。[15]

基本パッチ構造は固定サイズ(22x24画素)の三角格子とし、フレーム毎にマスクパターンに輪郭適合して可変形状パッチを得た。パッチの各頂点のフレーム間対応位置は、ブロックマッチングで粗く探索した後、微小補正を繰り返して検出した。パッチ単位でのワーピング予測を行い、予測誤差を可変形状DCTで変換符号化した。シミュレーション条件を表1に示す。

Table 1 Specification of Simulation

Sequences	Foreman, Mother&Daughter
Frame Size	176x144 pixels
Frame Rate	7.5Hz
ME Range	(+8.0)x(+8.0):Half Pel
Shape-Ad.DCT	8x8 SA-DCT or Pch-DCT
VLC	MPEG-1 VLC Table

写真1はオブジェクト(人物)の輪郭に適合したパッチ構造を示す。写真2は写真1からオブジェクト領域に関わる有効なパッチだけを切出した状態を示す。パッチの辺が輪郭適合することで、有効な領域を効率良く切出していることがわかる。写真3は、写真2で切出した領域を独立してワーピング予測し、予測誤差を可変形状DCT(SA-DCT)で符号化および復号したものである。写真5,6,7は別のシーケンスでオブジェクト抽出し、符号化した結果である。写真3,7の圧縮データはオブジェクト単位で独立して符号化しており、単独で復号可能であることはもちろん、任意の背景の圧縮データとも容易に合成できる。

図8に抽出したオブジェクトだけをフレーム間予測符号化した時の特性(10フレーム平均)を示す。いずれも可変形状パッチの符号量は300bit/frame程度である。可変形状DCTは2種類の方法で比較した。SA-DCTでは無駄な

係数が発生せず効率が良い。一方、Pch-DCTはSA-DCTに比較して符号化効率がやや劣る。原因は、Pch-DCTは一度に変換する領域が大きく高域成分を含み易いこと、パッチの変形が激しく変換時のシフト処理の影響が大きいこと、可変長符号化に8x8ブロックベース用のMPEG-1 VLCテーブルを用いたため、必ずしも効率良く符号変換できていない等の可能性が考えられる。

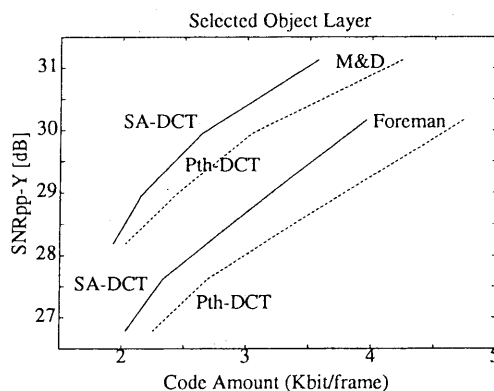


Fig.8 Simulation Result (1 object)

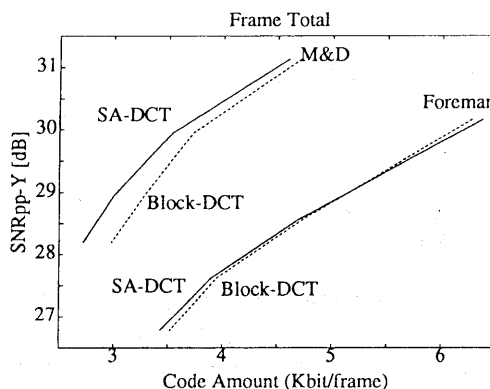


Fig.9 Simulation Result (Frame Total)

図9に抽出したオブジェクトと背景とを独立して符号化し、全体を統合した結果(10フレーム平均)を示す。比較のためMPEG-1方式でブロック単位でオブジェクト階層符号化した場合も示した。いずれも24~40kbpsの低ビット

レートで符号化できた。本方式の符号化効率はブロックベース方式に較べて若干良い程度ではあるが、注目するオブジェクト領域をより正確に分離できる点で優れている。写真4,8に本方式でオブジェクト階層符号化し、画像全体を復号合成した結果を示す。オブジェクトの境界部分も自然に再現できている。

4. まとめ

輪郭適合パッチベースの手法によりオブジェクト階層符号化を実現する方法を提案し、その有効性を示した。本方式は画像全体を一括して符号化する場合にも、そのままの符号化シタックスで適用できる。

なお本稿では低ビットレート伝送を考慮し、パッチ構造による大間かな形状表現を検討した。一方、高いビットレートが許容され、高度な画像の合成を目的とする応用分野では、より高精度に領域形状を表現したい場合もある。オブジェクト形状をより正確に表現するには、パッチサイズを階層的に小さくしたり、マスクボタンを追加データとして符号化すれば良い[16]。マスクボタンを符号化した場合には、ワーピング予測や可変形状DCTの有効領域もマスクボタンに従いOn/Off制御する[3]。なお、オブジェクト階層符号化における形状表現の精度を、段階的に改善する方法(クオリティ階層符号化)は今後の検討課題である。

文献

- [1] ISO/IEC 11172-2: "Information technology - Coding of moving pictures and associated audio for digital storage media at up to about 1.5Mbit/s - Video Part."
- [2] ISO/IEC 13818-2: "Information technology - Generic coding of moving pictures and associated audio."
- [3] 宮本: "被写体輪郭に適應した動き補償方式", PCSJ93, 2-11 (1993)
- [4] 宮本: "動き内挿予測を用いた領域階層符号化方式", 1994 秋季信学全大, D-154 (1994)
- [5] 横山, 宮本, 太田: "輪郭適合パッチを用いた適應 Warping 予測", 1995 春季信学全大, D-319 (1995)
- [6] 松田, 伊東他: "DCTを用いた画像の可変ブロック形状変換符号化の検討", 1992 秋季信学全大, D-146 (1992)
- [7] T.Sikora and B.Makai: "Shape-Adaptive DCT for Generic Coding of Video", IEEE Trans. on CSVT, Vol.5, No.1 (1995)
- [8] M.Hotter: "Object Oriented Analysis-Synthesis Coding Based on Moving Two-dimensional Objects", Signal Processing, Image Comm., Vol.2, No.4 (1990)
- [9] Y.Yokoyama, Y.Miyamoto and M.Ohta: "Very Low Bitrate Video Coding with Object Based Motion Compensation and Orthogonal Transform", VCIP, SPIE, Vol.2094 (1993)
- [10] J.Nieweglowski, T.G.Campbell and P.Haavist: "A Novel Video Coding Scheme Based on Temporal Prediction using Digital Image Warping", IEEE Trans. on CE, Vol.39, No.3 (1993)
- [11] G.Wolberg: "Digital Image Warping", Los Alamitos, CA, USA, IEEE Computer Society Press (1990)
- [12] 中屋他: "三角形パッチによる動き補償の基礎検討", 信学技報, IE90-106 (1991)
- [13] 山本, 松田, 伊東他: "画像の可変ブロック形状KL変換符号化", 1993 春季信学全大, D-244 (1993)
- [14] M.Gilge, T.Engelhardt and R.Mehlan: "Coding of Arbitrarily Shaped Image Segments Based on a Generalized Orthogonal Transform", Signal Processing, Image Comm., Vol.1, No.2 (1989)
- [15] P.Kauff: "HHI Test Sequences", ISO/IEC JTC/SC2/WG11 MPEG95/50 (1995)
- [16] 柴藤, ブン, 角野: "階層表現と多重テンプレートを用いた画像符号化", 信学技報, IE94-159 (1995)

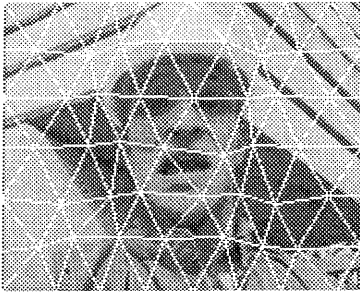


Photo 1 Patch Structure (Foreman)

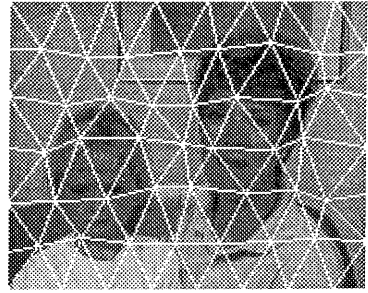


Photo 5 Patch Structure (M&D)

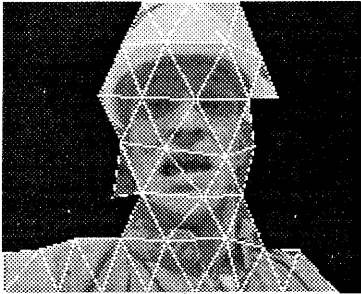


Photo 2 Selected Patches (Foreman)

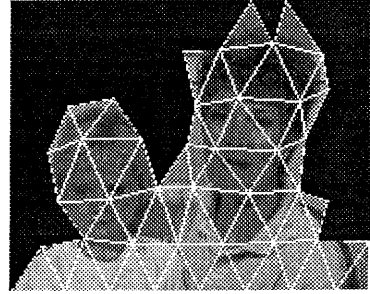


Photo 6 Selected Patches (M&D)



Photo 3 Decoded Object(2.0Kbpf)



Photo 7 Decoded Object(1.9Kbpf)



Photo 4 Decoded Frame (3.4Kbpf)



Photo 8 Decoded Frame (3.2Kbpf)