

ビデオ画像中のシーンの 3 次元形状に適合する三角網の自動生成

Automatically Generating a Triangular Mesh Adapted for 3-D shape from Video Images

中辻敦忠† 菅谷保之‡ 金谷健一‡

Atsutada Nakatsuji†, Yasuyuki Sugaya‡, Kenichi Kanatani‡

† NEC エンジニアリングインターネットターミナル事業部

‡ 岡山大学大学院自然科学研究科

†Internet Terminals Division, NEC Engineering, Ltd.

‡Department of Computer Science, Okayama University.

E-mail: nakatsuji@suri.it.okayama-u.ac.jp

Abstract

特徴点に基づくビデオ画像からの 3 次元復元では通常, それらの特徴点を頂点とする三角網が定義され, シーンが多面体として表示される. しかし, シーン自身が多面体の場合は物理的な辺に矛盾する辺が現れる. 本論文ではそのような誤りを検出するフィルタを用いて初期の三角網を物理的な辺に適合するように自動的に変換する新しい手法を提案する.

1. まえがき

ビデオ画像からシーンの 3 次元形状を復元するとき, 復元した形状をどう表現するかが大きな問題である. 代表的な方法は, 未校正カメラで撮影した画像上に特徴点を抽出し[5], それら特徴点間の対応を決定し[4][10], その特徴点の 3 次元位置を計算し[1][3], それを頂点とする三角網を定義して物体を多面体表示し, テクスチャマッピングを施して表示する方法である. これに必要な三角網は, 指定した画像上で特徴点を頂点とするドロネー三角網を用いればバランスのとれたメッシュが自動的に生成できる.

しかし, シーン自体が多面体の場合は, ドロネー網によって定義される多面体の辺と物理的な辺が合致せず, 矛盾した 3 次元形状が復元されることがある. そこで我々は 2 画像から矛盾を検出するテンプレートを導入して三角網を物理的な形状に適合するように最適化する手法を提案した[8]. 室内や建物などの人工的な環境では物体のほとんどが多面体であるから, この方法は実際の応用に非常に有益である.

本論文では, 前報[8]の 2 画像に対する処理をビデオ画像に拡張する. そして, 実画像を用いて従来手法[7][9]と比較し, 提案手法が優れていることを実証するとともに, 従来手法の問題点を明らかにして, 提案

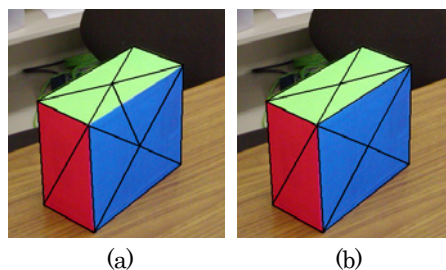


図1: (a) 物理的な辺に矛盾する三角網. (b) 物理的な辺に適合する三角網.

手法の特性と対比する.

2. 最適化の原理

図 1(a) のように物体頂点が特徴点として選ばれたとする. これから図 1(a) に示すドロネー網が得られるが, この多面体表現は物体形状を正しく表していない. 一方, 図 1(b) の三角網は物体形状を正しく表している. 本論文でいう三角網の“最適化”とは図 1(a) のような物理的な辺と矛盾する三角網を図 1(b) のような物理的な辺に適合する三角網に自動的に変換することである. Morris ら[7]と Perrier ら[9]は, 3 次元データを用いずに (3 次元復元は行わずに), 2 次元画像のみから計算する方法を提案した. その原理は, シーンの平面部分を表す三角形パッチは他の画像の対応する三角形パッチにアフィン変換で写像され, 平面部分でなければ写像後に輝度差が現れるという事実である. これを利用して, Morris ら[7]は全パッチを対応するパッチに写像したときの残差平方和を最小にするように辺の入れ換えを繰り返すグリーディ法を用い, Perrier ら[9]は残差平方和の代わりに正規化相関を用いた. 各三角形パッチの輝度値から定数を引いてパッチ内の平均輝度値を 0 に正規化し, さらに定数で割ってパッチ内の輝度値の分散を 1 に正規化しておけば, その残差平

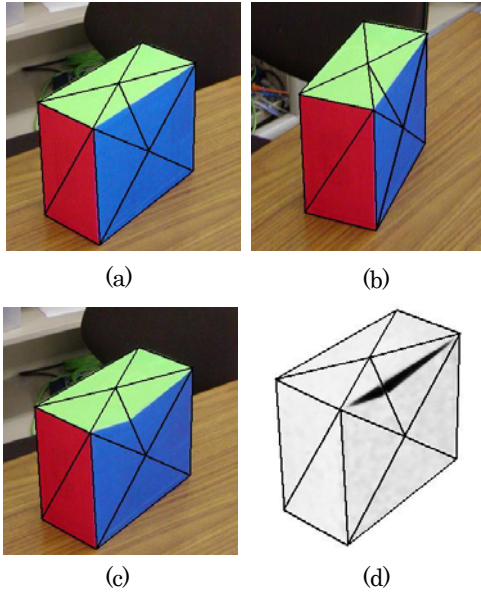


図2: (a), (b) 入力画像と初期三角網. (c) 図(b)を図(a)にテクスチャマッピングしたもの. (d) 図(a)と図(c)との差画像.

方和の最小化と正規化相関の最大化は等価であるから, Morrisら[7]の方法と Perrierら[9]の方法は本質的には輝度値を正規化するかしないかの違いでしかない.

わが国でも飯田ら[2]が Morrisら[7]と同じような考え方を述べていた. 栗原ら[6]は Morrisら[7]の方法で実画像実験を行い, 2画像間の照明変化を打ち消す必要性を指摘した. これは実質的に Perrierら[9]に他ならない.

3. 不適合の検出テンプレート

図2(a), (b)は多面体物体の画像上に初期三角網(図2(a)の頂点に対するドロネー網)を表示したものである. 図2(c)は図2(b)の各三角形パッチを図2(a)の対応する三角形パッチ上にテクスチャマッピングしたものであり, 図2(d)は図2(a)と図2(c)の差画像である. このように, 輝度差は誤った辺を横切る細長い三角形領域(不適合領域)に顕著に現れることがわかる. そこでこれを検出するために次の不適合検出テンプレートを定義する.

図3(a)はこれを濃淡表示したものである. 大きさは $l \times l$ であり, 実験では $\alpha = 0.1$ とした. $T(x, y)$ の等高線は R, T を端点として OS 上に頂点を持つ折れ線となっている. そして, OS に関して対称であり, RT に関して反対称(対称な値の符号を換えたもの)である.

図3(b)はその対角線 OS 上の切り口の値のグラフであり, 平均 $l/\sqrt{2}$, 標準偏差 $\alpha l/\sqrt{2}$ の正規分布の片側の符号を換えたものになっている.

反対称なテンプレートを用いるのは, 差の現れる不適合領域がどちら側に現れるかあらかじめわからないためである. 反対称にすれば, それ領域がどちらにあ

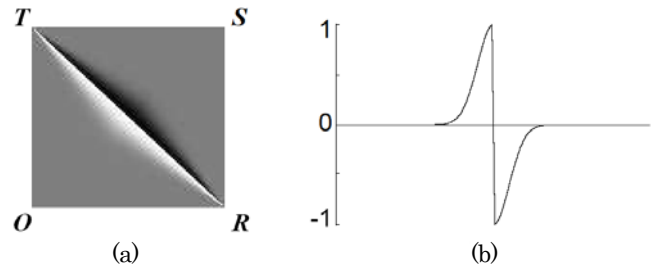


図3: (a) 誤った辺を検出するテンプレートの濃淡表示(白が正, 黒が負). (b) 対角線 OS に沿った切り口の値.

$$T(x, y) = \begin{cases} e^{-\frac{(x+y-l)^2}{2\alpha^2(x-y-l)^2}} & x+y < l, x \geq y \\ T(y, x) & x+y < l, x < y \\ -T(l-y, l-x) & x+y > l \end{cases} \quad (1)$$

っても反対側の輝度差はほとんど0であるから, 相関の絶対値を計算することによって不適合が検出できる. さらに, テクスチャマッピングのずれのためにパッチ全体に分布するランダムな誤差パターンや照明変化による様な輝度変化はこのような反対称なテンプレートによって正負が打ち消される.

ただし, 特徴点の検出の誤差によって三角形パッチの $\triangle ORS$, $\triangle OTS$ 上への写像にわずかなずれが生じることがあるので, 図3のテンプレートの対角線 RT 上の画素, および対角線 OS の上下2画素とフレームの周囲の2画素の値は0とする.

4. 最適化の手順

Perrierら[9]は2画像の場合しか考慮していないが, Morrisら[7]は多画像の場合に各三角形パッチの全フレームに渡る平均と各パッチとの残差平方和を最小化している. 本論文では, M 枚のフレーム辺 AB の不適合の程度を測る評価値 $w(AB)$ を次のようにして最適化を行う.

1. 辺 AB の片側に一つの三角形しかない場合は境界辺であるから, $w(AB) = -1$ を返す.
2. 辺 AB の両側の三角形を $\triangle ABP$, $\triangle ABQ$ とするとき, 四辺形 $APBQ$ が凸でないフレームが一つでも存在すれば, $w(AB) = 0$ とする.
3. 値を0に初期化した $l \times l$ 正方形領域 $ORST$ を用意し, 以下を $\kappa = 1, \dots, M$ に対して計算する.
 - (a) 第 κ フレームの四辺形 $APBQ$ を正方形領域 $ORST$ に射影変換し, 輝度値を足す.
 - (b) 第 κ フレームの $\triangle ABP$, $\triangle ABQ$ をそれぞれ正方形領域 $ORST$ の $\triangle OSR$, $\triangle OST$ にアフィン変換し, 輝度値をそこに書かれている値から引く.
4. 得られた正方形領域 $ORST$ の値と検出テンプレートとの相関の絶対値を $w(AB)$ とする.

ステップ3はどのフレームも同等に扱う対称な操作であり、その意味を図4に示す。ステップ3(a)は考えている辺 AB を囲む四辺形の形がフレームごとに異なるので、同じ形（正方形）にそろえる操作である。

ステップ3(b)のように辺 AB 両側の三角形をそれぞれアフィン変換すると、辺 AB が正しければ写像した結果はステップ3(a)で得られる射影変換とほぼ同じになるが、誤っていれば輝度差が辺 AB を横切る不適合領域に現れる。これはシーンの見え方によって辺 AB を横切る対角線のどちらかに現れるか一定しないが、どちら側かで符号が異なる。これを全フレームに対して合計すると、図4の右のように、辺 AB を横切る対角線を境界として両側に符号の異なる不適合領域が現れる（一方のみに現れる場合もある）。これと図3(a)のテンプレートとの相関をとることによって不適合の程度を測定する。

ステップ2は、四辺形 $APBQ$ がどのフレームかで凹であれば、入れ換えるべき対角辺が内部にないので、入れ換えは行わないという意味である。以下、 w の値の大きい辺から入れ換えていく。そして、入れ換えた辺の w を計算し、それがより大きくなれば入れ換えを元に戻す。これをすべての辺に行いもはや三角網が変化しなくなるまでこれを反復する。この過程で何らのしきい値を導入する必要はない。

5. 実画像実験

5.1 手法の比較

Morrisら[7]の方法、Perrierら[9]の方法、および提案手法を2画像例（特徴点は手動で指定）に適用して性能比較を行った。図5~7の(a)は入力2画像と初期三角網、(b)は提案手法で最適化した三角網を第1画像上に重ねたものである。図説中に辺数と計算時間(秒)を示す。ただし、CPUにはPentium4 3.2GHz、主メモリ2GB、OSにはLinuxを用いた。

1は3手法による正解率(%)を示す。正解率とは、境界辺は除くすべての辺に対する正しい（すなわち端点在同一平面上にある）辺の割合である。異なる特徴点在同一平面上にあるかどうかは目視で判定した。3手法を比較すると、全般的に提案手法がMorrisら[7]やPerrierら[9]の方法より優れていることがわかる。

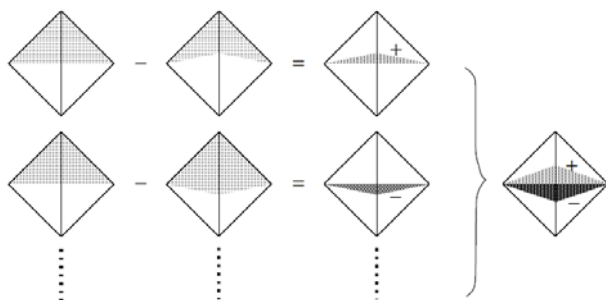


図4: ビデオ画像からの不適合の検出の原理。

図8に対してパッチ全体を比較するMorrisら[7]の方法が提案手法を上回ったのは、誤りを生じた三角形パッチが小さ過ぎて不適合領域が十分に抽出できなかったためと思われる。

一方、Perrierら[9]の方法は照明変化を打ち消すために輝度値を正規化したものであるが、2画像間で日照変化の激しい図7以外はMorrisら[7]の方法に比べて正解率が悪化している。輝度値を正規化するのは対応するパッチ間の類似性を増加させる目的であるが、副作用として対応しないパッチ間の類似性も増加してしまう。それに対して、提案手法は反対称テンプレートを用いているので、照明変化に影響されない。

図7のように複数のパッチが物理的な辺を横切ると、対角辺の入れ換えが一回では不十分で、連続した入れ換えが必要である。このような場合の不適合領域を観察すると、図2(d)のような明瞭な三角形ではなく、やや狭まった領域に縮んでいる。しかし、図3(a)の検出テンプレートは依然として有効であり、誤った辺が次々に修正されて、最終的にすべての辺が正しい位置に置かれた。

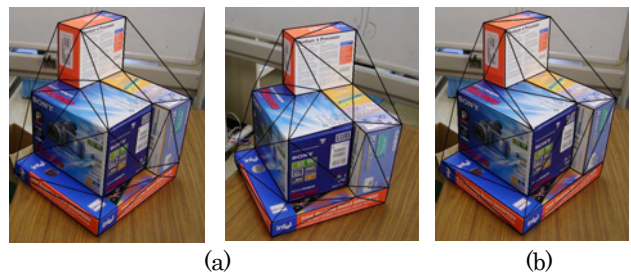


図5: (a) 初期三角網（辺数58）. (b) 提案手法による最適化（3.43秒）.



図6: (a) 初期三角網（辺数31）. (b) 提案手法による最適化（3.15秒）.

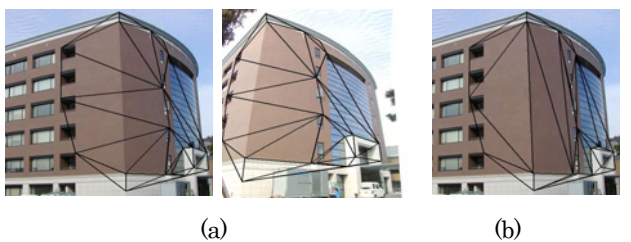


図7: (a) 初期三角網（辺数47）. (b) 提案手法による最適化（4.03秒）.

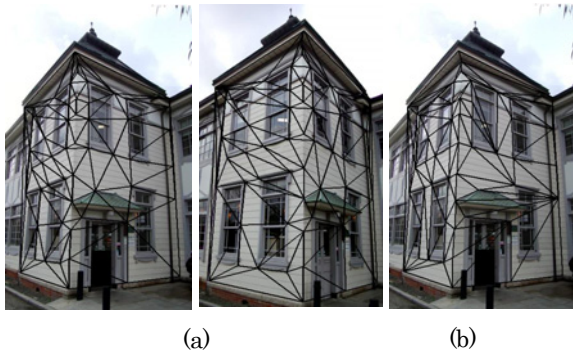


図8: (a) 初期三角網 (辺数 157) . (b) 提案手法による最適化 (11.85 秒) .

表 1: 最適化した辺の正解率(%) .

	図 5	図 6	図 7	図 8
Morris ら	100	73.9	89.2	100
Perrier ら	95.7	69.6	91.9	98.7
提案手法	100	100	100	98.7

5.2 ビデオ画像との比較

2 画像のみ用いる場合と連続ビデオ画像を用いる場合の比較を行った. 図 9 の (a), (b) はそれぞれ 30 フレームの画像列である.

Perrier ら [9] は 2 画像の場合しか考慮していないが, ここでは Morris ら [7] と同様にして多画像に拡張した. Morris ら [7] も Perrier ら [9] もパッチ同士を比較しているが, ここでは提案手法の技法を入れてテンプレート領域に射影変換して形を合わせてから比較した. 正解率 (%) と実行時間 (秒) を表 2 に示す. また, 比較のために初期フレームと最終フレームの 2 画像のみを用いた最適化も行った.

表 2 から, 提案手法は全部のフレームを用いても前後の 2 フレームのみでも正解率に変化がないことがわかる. また, Morris ら [7] や Perrier ら [9] の方法では, 提案手法に比べてかなり正解率が低下し, しかも全フレームを用いると正解率が悪化することが多い.

6. まとめ

本論文では, 多面体シーンの三角網による 3 次元表現から物理的な辺と矛盾する辺を除去する新しい手法を提案した. 実画像例を用いて Morris ら [7] や Perrier ら [9] の方法と比較し, 次の結論を得た.

1. 提案手法は Morris ら [7] や Perrier ら [9] の方法より高性能である.
2. 最適化はビデオ系列の初期フレームと最終フレームのみを用いれば十分である.

提案手法は Morris ら [7] や Perrier ら [9] の方法と同様に, 入力画像間を比較するのみで, シーンの 3 次元情報は一切必要としない. また, 判定のための何らのしきい値も必要としない.

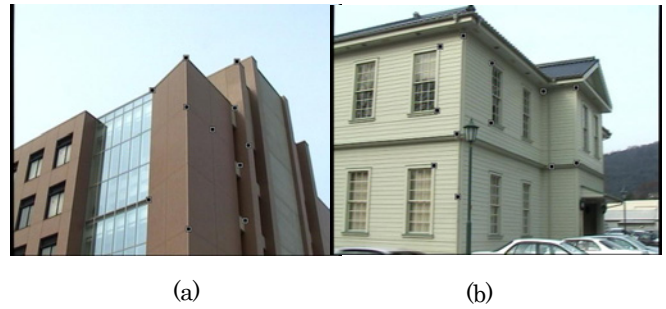


図 9: ビデオ画像例

表 2: 図 11 のビデオ画像の最適化の正解率 (%) と実行時間 (秒) . それぞれ (全フレームを使う場合の値) / (前後の 2 フレームのみを使う場合の値) .

(a)	正解率 (%)	実行時間 (秒)
Morris ら	100/81.3	12.8/0.8
Perrier ら	62.5/56.3	11.6/1.0
提案手法	100/100	25.2/1.2

(b)	正解率 (%)	実行時間 (秒)
Morris ら	81.0/81.0	12.0/0.9
Perrier ら	71.4/61.9	14.8/1.1
提案手法	100/100	31.0/1.9

参考文献

- [1] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, Cambridge, U. K., 2000.
- [2] 飯田亮介, 出口光一郎, 平面を利用したステレオ画像からの任意視点画像の合成, 情報処理学会研究報告, 99-CVIM-114-10, pp. 73-80, January 1999.
- [3] 金谷健一, 三島等, 未校正カメラによる 2 画像からの 3 次元復元とその信頼性評価, 情報処理学会論文誌: コンピュータビジョンとイメージメディア, 42-SIG 6 (2001-6) pp. 1-8.
- [4] 金澤靖, 金谷健一, 大域的な整合性を保証するロバストな画像の対応づけ, 情報処理学会論文誌: コンピュータビジョンとイメージメディア, 44-SIG 17 (2003-12), pp. 70-77.
- [5] 金澤靖, 金谷健一, コンピュータビジョンのための画像の特徴点抽出, 電子情報通信学会誌, 87-12 (2004-12), 1043-1048.
- [6] 栗原祐介, 日吉久礎, 金谷健一, 太田直哉, 画像の一致度を考慮したサーフェスモデルの構築, 情報処理学会研究報告, 2001-CVIM-127-8, pp. 51-58, May 2001.
- [7] D. D. Morris and T. Kanade, Image-consistent surface triangulation, *Proc. IEEE Conf. Comput. Vision Pattern Recog.*, Hilton Head, SC, U. S. A., Vol. 1, pp. 332-338, June 2000.
- [8] 村田正和, 中辻敦忠, 菅谷保之, 金谷健一, 画像からの形状復元に適合した三角網の生成, 情報処理学会研究報告, 2004-CVIM-145-2, pp. 9-16, September 2004.
- [9] J. S. Perrier, G. Agin, and P. Cohen, Image-based view synthesis for enhanced perception in teleoperation, in J. G. Verly (Ed.), *Enhanced and Synthetic Vision 2000*, Proc. SPIE, Vol. 4023, June 2000.
- [10] Z. Zhang, R. Deriche, O. Faugeras and Q.-T. Luong, A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry, *Artif. Intell.*, 78 (1995), pp. 87-119.