

両眼立体視による視覚のモデル化と形状認識

宇野公裕 † 三池秀敏 ‡

† 山口大学大学院理工学研究科電気電子工学専攻

‡ 山口大学工学部感性デザイン工学科

本研究では、2台のTVカメラの光軸を対象物体の表面で交差するように配置して得られる2枚のステレオペア画像を対象とする。この時のカメラの配置は、人間がある物体を観察するときの視線の状態と類似している。また、人間は2つの映像を見ているという意識はなく、両眼の中央に位置すると考えられる仮想的単眼で物体を観測していると理解できる。そこで、得られた画像から、人間が脳の内部で感知していると思われる映像の創生(合成)を試みる。また、この合成映像では両眼の注視点での像は明瞭であるが周囲は二重に見える。二重化の度合いはカメラからの距離が注視点の距離より離れているほど大きくなる。したがって、二重化の度合いは奥行き情報を含んでいるといえる。この二重化のずれ情報をもとに、3次元形状を復元する。

Modeling Virtual Monocular Vision and 3D-Shape Reconstruction from Binocular Images

Kimihiro Uno † and Hidetoshi Miike ‡

† Graduate School of Science and Engineering, Yamaguchi University

‡ Department of KANSEI Design and Engineering, Yamaguchi University

The subject of this study is modeling human vision. We consider a pair of stereo images captured by two cameras whose optical axes are intersected on the surface of the object. The arrangement of cameras is similar to human visual system. We assume that human vision (brain) fuses two stereo images into one virtual image. This brings a feeling of unified scene observed by a *virtual ocellus*. We try to create a virtual monocular image from binocular images by use of Affine transformation. The virtual image has double vision, except for around the attention point. Double vision has depth information. Finding the degree of double vision, we try to reconstruct 3-dimensional shape around the point.

1. はじめに

人間が物体を観察するとき、両眼から情報を得ている。しかしながら、融合された映像には2つの映像を見ているという意識はなく、単一の眼が捕らえた単一の映像として意識されている。ここでは、両眼の中央に位置する仮想的単眼で物体を見ていると仮定しよう。片目を閉じると左右では異なった映像が見える。この映像は両目で見たときの映像とも異なっている。注意深く観察すると両眼が融合した映

像は左右の単眼の映像よりサイズが大きく、また視角も異なっている。これは人間が両眼から得た2つの映像情報を用いて、脳の内部で1枚の合成映像を創っているからではないかと考えられる[1]。また、単眼による情報は2次元情報である。人間はこの2次元的な映像を複数枚用いて3次元情報を得ている。立体感を与える要因としては、表1のように両眼の視差、単眼の運動視差、水晶体の調節などの生理的要因と、物の大小、重なり、高低や、映像の明暗、陰影、テクスチャー、コントラスト、彩度などの心理・記憶的要因があげられる。本研究では、TVカメラの光軸を対象物体の表面で交差するように配置して2枚の画像を計算機に取り込む。得られた画像から人間が脳の内部で感知していると思われる1枚の映像の創生を試みる[2]。また、この映像は両眼の注視点での像は明瞭であるが周囲は二重に見える。この二重になる時のずれ情報をもとに3次元形状を復元する手法を提案する。

表1 立体感を与える要因

生理的要因	両眼視差、単眼の運動視差、水晶体の調節	
心理的要因	幾何学的要因	物体の大小、重なり、高低
	光学的要因	陰影、テクスチャー、コントラスト、彩度

2 両眼立体視

2.1 両眼立体視の基本原則

人間の眼がある物体を観測するとき、右眼による映像上の物体の位置と、左眼による映像上の物体の位置には差が生じる。この位置の差を視差といい、近くの物体に対しては大きく、遠くの物体に対しては小さくなる。したがって、視差は奥行き情報を含んでいるといえる。両眼立体視は、人間の眼のように2台のカメラを用いて得られる左右の画像から同一物体の視差を求め、三角測量の原理からその物体の奥行きを求めようとするものである。

2.2 両眼立体視の分類

両眼立体視においてこれまでに多くの研究がなされてきた[3]～[8]。これらは、カメラの配置法からおおまかに2つに分類される。1つは、光軸平行型である[3][4]。これは、図1のように左右のカメラの光軸を平行に配置する方法である。2枚の画像の対応点を求めることが最大の問題である。したがって、カメラの光軸を平行にし、水平方向に対して対応づけを行う。

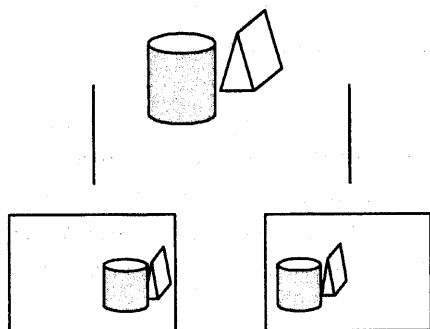


図1 光軸平行型の両眼立体視

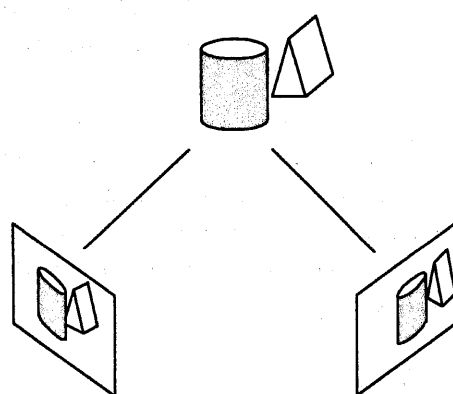


図2 光軸交差型の両眼立体視

もう1つは、光軸交差型である[5]～[8]。この場合、図2のようにカメラの光軸は対象物体のある1点で交差する。したがって、カメラの配置は人間の視覚と類似し立体視のモデル化が可能となる。しかし、対応点の探索が困難となりさまざまな方法が試みられている。

3. 原理

3.1. 両眼立体視における画像の融合

本研究では、人間の視覚機能のモデル化を実現するため光軸交差型の両眼立体視を行う。つまり、図3のように物体の表面上の点（注視点） P_0 で2台のカメラの光軸を交差する。この時、カメラから得られる映像の中央に注視点が撮像される。ここで、左右のカメラの基線に対する角度をそれぞれ α 、 β 、カメラ間距離を δx 、カメラの焦点距離を f とする。注視点の奥行き z_0 は α 、 β 、 δx から次式で求まる。

$$z_0 = \frac{\delta x \tan \alpha \tan \beta}{\tan \beta - \tan \alpha} \quad (1)$$

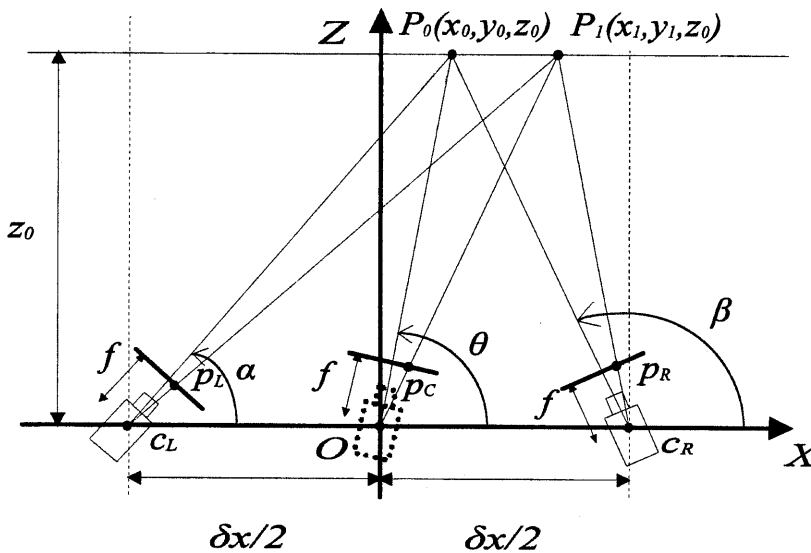


図3 仮想画像作成原理 (水平方向)

次に、得られた2枚の画像から新たにカメラの基線上の中央から見た画像をそれぞれ作成する。ここで、カメラから得られた映像は基線からの距離が z_0 で一定の同一平面上の物体であると仮定する。 $P_0(x_0, y_0, z_0)$ の近傍の点 $P_1(x_1, y_1, z_0)$ を考えると、 P_1 は各々の画像平面上的 $P_L(x_L, y_L)$ 、 $P_R(x_R, y_R)$ に投影される。同様に、人間の仮想的単眼に見立てた中央のカメラから撮影したと仮定した画像平面上では $P_C(x_C, y_C)$ に投影されたとする。したがって、 $P_C(x_C, y_C)$ と $P_L(x_L, y_L)$ 、 $P_R(x_R, y_R)$ の関係が得られれば新たな画像は作成できる。これらの関係は次式で表される。なお、垂直方向を計算するには図4を考慮する。

$$x_L = \frac{x_C(1 + \tan^2 \theta) f \tan^2 \alpha}{(1 + \tan^2 \alpha) \tan \theta (f \tan \theta - x_C) + x_C(1 + \tan^2 \theta) \tan \alpha} \quad (2)$$

$$x_R = \frac{x_C(1 + \tan^2 \theta) f \tan^2 \beta}{(1 + \tan^2 \beta) \tan \theta (f \tan \theta - x_C) + x_C(1 + \tan^2 \theta) \tan \beta} \quad (3)$$

$$y_L = \frac{\sin\left(\alpha - \tan^{-1} \frac{x_L}{f}\right)}{\sin\left(\theta - \tan^{-1} \frac{x_L}{f}\right)} y_C \quad (4)$$

$$y_R = \frac{\sin\left(\beta - \tan^{-1} \frac{x_R}{f}\right)}{\sin\left(\theta - \tan^{-1} \frac{x_R}{f}\right)} y_C \quad (5)$$

ただし、 θ は α 、 β から次式で求まる。

$$\tan \theta = \frac{2 \tan \alpha \tan \beta}{\tan \alpha + \tan \beta} \quad (6)$$

得られた2枚の画像を合成することで人間が脳の内部で感知している映像と類似した画像が得られる。
この合成画像は注視点近傍では明瞭で周辺部は二重に見える。

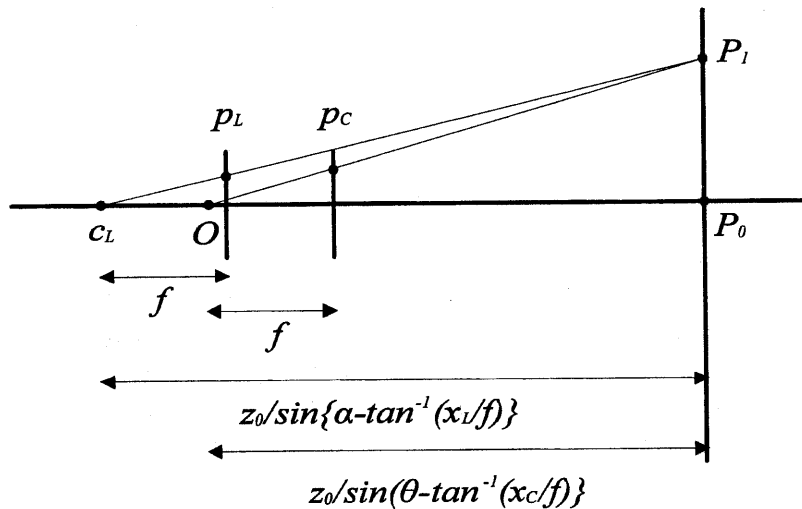


図4 仮想画像作成原理 (垂直方向)

3.2.3 次元形状の認識

前節で中央のカメラから見た画像を作成するとき、カメラから得られた映像は基線からの距離が z_0 で一定の同一平面上の物体であると仮定した。しかし、実際の物体には3次元形状が存在し、カメラで得られた映像の全領域にわたり仮定が成立することはまれである。仮定が不成立である点が合成画像の二重化となって現れる。二重化の度合いはカメラの基線からの距離が注視点の距離より離れているほど大きくなる。したがって、二重化の度合いは奥行き情報を含んでいるといえる。この二重になるずれ情報をもとに3次元形状を復元する。図6のように任意の点 $P_2(x_2, y_2, z_2)$ の奥行き情報を決定できる。前節で作成した2枚の画像に対してテンプレート・マッチングにより対応づけを行う。テンプレート・マッチングとは図5のように画像平面上にある大きさのテンプレートを用意し、一方の画像におけるテンプレートを固定して他方の画像のテンプレートを移動させながら両方のテンプレートの差分を取り、それらの和がもっとも小さい点（もっとも類似した点）を対応点とするものである。ここでは、カメラの視平線は一致するものとし水平方向についてのみ探索を行う。左カメラから作成した画像上の $P_{Cl}(d_L, y_C)$ と、右カメラから作成した画像上の $P_{Cr}(d_R, y_C)$ が対応したとき、対応点の元の取り込み画像での位置を計算する。 $P_{Cl}(d_L, y_C)$ は左カメラから取り込んだ画像の $P_L(x_L, y_L)$ 、 $P_{Cr}(d_R, y_C)$ は右カメラから取り込んだ画像の $P_R(x_R, y_R)$ に対応するとすれば、次の関係が得られる。

$$x_L = \frac{f\{(\delta x \tan \alpha - 2z_0)(f \tan \theta - d_L) + 2z_0 \tan \alpha (f + d_L \tan \theta)\}}{(2z_0 \tan \alpha + \delta x)(f \tan \theta - d_L) + 2z_0 (f + d_L \tan \theta)} \quad (7)$$

$$x_R = \frac{f\{(-\delta x \tan \beta - 2z_0)(f \tan \theta - d_R) + 2z_0 \tan \beta (f + d_R \tan \theta)\}}{(2z_0 \tan \beta - \delta x)(f \tan \theta - d_R) + 2z_0 (f + d_R \tan \theta)} \quad (8)$$

したがって、三角測量の原理から α 、 β 、 δx 、および f から $P_2(x_2, y_2, z_2)$ の奥行き z_2 が次式で求まる。

$$z_2 = \frac{\delta x}{\frac{f + x_L \tan \alpha}{f \tan \alpha - x_L} - \frac{f + x_R \tan \beta}{f \tan \beta - x_R}} \quad (9)$$

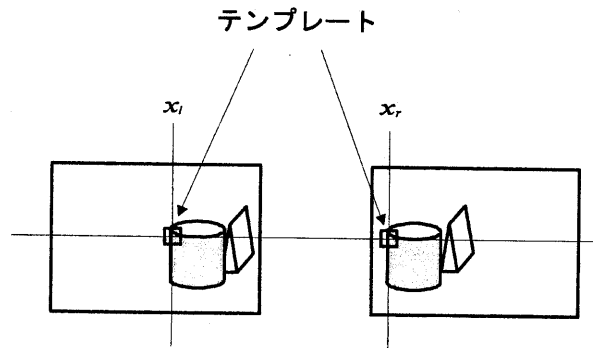


図5 テンプレート・マッチング

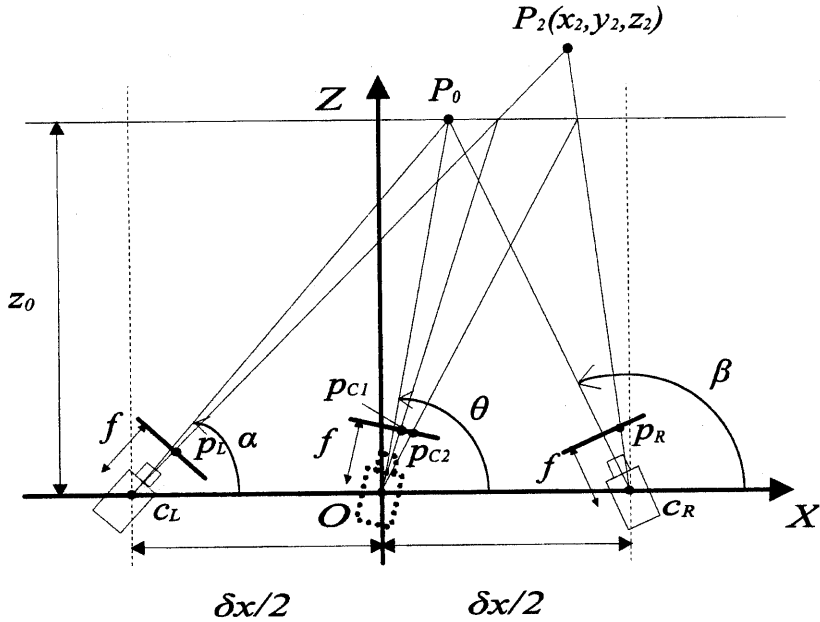


図 6 奥行き検出方法

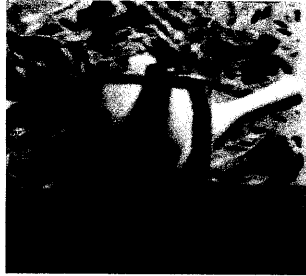
4. 解析と結果

表 2 のような条件で左右の画像を 1 枚ずつ取り込んだ。

表 2 実験に用いたカメラの条件

カメラの角度	左	$\alpha=80^\circ$
	右	$\beta=97^\circ$
カメラ間距離	$\delta x=16.0\text{cm}$	
焦点距離	$f=12.5\text{mm}$	

最初に、式(2)~(5)を用いて人間の仮想的単眼に見立てた中央のカメラから見たと仮定した映像を作成した。実際に取り込んだ画像を図 7 に、新たに作成した中央のカメラの画像を図 8 に示す。



左カメラ

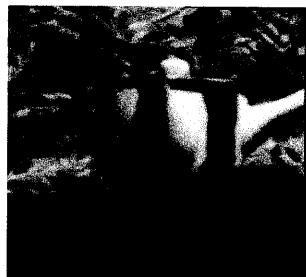


右カメラ

図 7 実画像



左補正画像



右補正画像

図 8 補正画像

補正画像の内側の領域は拡大され、外側の領域は縮小されている。つまり、左補正画像の右側と右補正画像の左側が拡大され、左補正画像の左側と右補正画像の右側が縮小されている。

次に、図 8 の補正画像を合成させて、人間が脳の内部で感知していると思われる映像の創生を試みた。この結果を図 9 に示す。



図 9 融合された画像



図 10 3次元形状

ここで得られた画像は人間が物体のある1点を注視したときの映像に類似している。注視点付近では明瞭で周辺部については二重化が起きている。この二重化が奥行き情報を与える。

次に、図8の画像に対してテンプレートのサイズを 3×3 [pixel]としてテンプレート・マッチングにより対応づけを行った。(1)式および、(7)～(9)式より、3次元形状の復元を行った。その結果を図10に示す。注視点近傍については、対応づけが行われ、3次元形状が復元できている。

5. まとめ

人間が物体を観察するとき、2つの眼を意識することはない。また、網膜上の映像は2次元情報であるにもかかわらず、3次元形状を認識している。本研究では、2台のカメラが得られた映像から人間が脳の内部で感知していると思われる映像を作成した。その結果、得られた画像上の物体は実際にカメラが捕らえた画像と違ったサイズになった。人間の視覚機能にも同じことが言える。我々が物体を注視すれば片眼で見たときと両眼で見たときでは両眼で見たときのほうが大きく見える。また、注視している物体の近傍は明瞭であり、周辺部は二重になっている。本研究では、この状態に非常に類似した映像が得られた。また、注視点近傍での3次元形状を復元できた。

実際には、人間はある一瞬では注視点近傍の大まかな形状を認識し、注視点をすばやくかつ連続的に移動させることにより、全体の3次元情報を獲得しているものと思われる。したがって、静止画のみを扱う場合には注視点とその近傍での3次元形状を復元させることで十分であるといえる。

今後、注視点を連続的に変化させていったときのアニメーション画像の作成、および注視点近傍3次元情報の蓄積による正確な形状計測法の確立を試みる予定である。

参考文献

- [1] 松田隆夫, 「視知覚」, 培風館, 1995.
- [2] K.Uno and H.Miike, *A Stereo Vision through Creating a Virtual Image using Affine Transformation*, IAPR Workshop on Machine Vision Applications, 1996, pp.526-529.
- [3] H.Z.Dan and B.Dubuisson, *String Matching for Stereo Vision*, Pattern Recognition Letters 9, Elsevier Science Publishers B.V., North-Holland, 1989, pp.117-126.
- [4] Z.Zhang, *Estimating Motion and Structure from Correspondences of Line Segments between Two Perspective Images*, IEEE Fifth International Conference on Computer Vision, 1995, pp.257-262.
- [5] Y.Xiong and S.A.Shafer, *Hypergeometric Filters for Optical Flow and Affine Matching*, IEEE Fifth International Conference on Computer Vision, 1995, pp.771-776.
- [6] A.L.Abbott and B.Zheng, *Active Fixation using Attentional Shifts, Affine Resampling, and Multiresolution Search*, IEEE Fifth International Conference on Computer Vision, 1995, pp.1002-1008.
- [7] B.Boufama and R.Mohr, *Epipole and Fundamental Matrix Estimation using Virtual Parallax*, IEEE Fifth International Conference on Computer Vision, 1995, pp.1030-1036.
- [8] N.C.Griswold and C.P.Yeh, *A New Stereo Vision Model Based upon the Binocular Fusion Concept*, Computer Vision, Graphics, and Image Processing 41, Academic Press, 1988, pp.153-171.