九州大学学術情報リポジトリ
Kyushu University Institutional Repository

両眼立体視における尤度計算

光藤, 宏行
九州大学大学院人間環境学研究院

https://doi.org/10.15017/18445

出版情報：九州大学心理学研究. 11, pp.57-62, 2010-03-31. Faculty of Human-Environment Studies, Kyushu University
バージョン：
権利関係：

Kyushu University Psychological Research
2010, Vol.11, 57-62

両眼立体視における尤度計算

光藤　宏行[1)]　九州大学大学院人間環境学研究院

Likelihood computation in stereo vision

Hiroyuki Mitsudo（*Faculty of Human-Environment Studies, Kyushu University*）

In the correlation model for stereo vision, local binocular disparity is determined by calculating the value of covariance for small regions called windows between the images of the two eyes. When applying the Bayesian approach to stereo vision, it is important to calculate the likelihood of binocular correspondence–the degree to which both eyes are likely to see a single visual scene given a pair of stereo images. This study presents a theoretical idea to compute the likelihood of binocular correspondence in a correlation model. For several types of stereograms, such as random-dot, grating, and more complex patterns, the likelihood of correspondence was calculated as a function of the correlation-window size and the range of disparity computation. The results showed that the present model's outputs are consistent with human percepts, suggesting that the idea proposed here is useful in modeling human stereo vision.

Key Words: stereo vision, binocular disparity, correlation model, Bayesian approach

現在の視覚心理学では，見るということは，与えられた網膜像やその他の情報から，視覚系が外界の事物の物理的特徴（e.g., 物体の形状，物体表面の反射率など）を推定する過程とみなされる。その際，理論的な枠組みとしてよく用いられるのはベイズ推定であり（Knill, Kersten, & Yille, 1996），色の恒常性（Brainard & Freeman, 1997），多感覚統合（Ernst & Banks, 2002），2 次元運動方向の解釈（Weiss, Simoncelli, & Adelson, 2002）など，人間の種々の知覚を説明する際に利用されている。この枠組みでは，外界の物理的特徴の表現に関するパラメタ（母集団特性値）に関する事前情報と得られた感覚的情報（視覚の場合には網膜像）に基づいて，パラメタの関数としての事後確率分布が計算される。視覚系が事後確率分布を評価することで，パラメタの推定値が求まり，知覚内容が一意に定まるとみなされる。

本論文では，見ることに関する重要な過程の一部である両眼立体視（binocular stereopsis）を取り上げる。両眼立体視とは，右眼と左眼の網膜像の間で対応する要素を見つけ，対応する要素の位置ずれをもとに，外界の詳細な 3 次元構造を知覚することをいう。右眼と左眼の網膜像のずれは，両眼網膜像差（binocular retinal disparity）と呼ばれる。視覚系が両眼網膜像差を視野の大部分に渡って自動的に計算することで，我々は外界の詳細かつ正確な 3 次元構造を容易に知覚することができると考えられている。両眼立体視の数学的なモデル化においてベイズ的アプローチを用いる場合，両眼網膜像の情報と外界の 3 次元構造に関する事前分布から，知覚対象となる 3 次元構造を含む m 個のパラメタ $\boldsymbol{\Theta} = \{\theta_1, \theta_2, \cdots, \theta_m\}$ について事後分布を計算し外界の 3 次元構造を一意に決定する，という枠組みとなる。形式的に書けば，両眼立体視は事後分布

$$p(\boldsymbol{\Theta}|\boldsymbol{D}) \propto p_{\text{prior}}(\boldsymbol{\Theta}) L(\boldsymbol{\Theta}|\boldsymbol{D}) \quad (\text{ただし } \boldsymbol{\Theta} = \{\theta_1, \theta_2, \cdots, \theta_m\}) \tag{1}$$

に基づき，外界の 3 次元形状を含めたパラメタを決定することである。ここで $\boldsymbol{D}$ はデータ（視覚情報），$p_{\text{prior}}(\boldsymbol{\Theta})$ は事前分布，$L(\boldsymbol{\Theta}|\boldsymbol{D})$ は尤度（likelihood）である。尤度とは，データが得られている場合に，パラメタの値がどの程度本当らしいかを示す度合いである。一組の画像について n 個の対応点を評価するとする。3 次元構造の解釈を決定するということは，分析画像の任意の位置で両眼間の対応があり，かつ像差が一意に決まるということである。したがって，特定の 3 次元解釈の尤度は，

$$L(\boldsymbol{\Theta}|\boldsymbol{D}) = \prod_{j=1}^{n} p_j \tag{2}$$

で表される。ここで p_j は与えられたパラメタ $\boldsymbol{\Theta}$ のもとで j 番目の対応点が特定の像差をもつ確率である。

Mitsudo, Kaneko, & Nishida（2009）はこの枠組みを拡張し，外界の 3 次元構造に関するものだけが視覚系が推定するパラメタなのではなく，像を得る装置である眼球

1) 本研究は文部科学省科学研究費補助金（若手研究 B，課題番号 20730478）の部分的な援助を受けた。草稿に関して，酒井歩さんと光藤崇子さんにコメントを頂いたことに感謝します。

の位置もまた視覚系が推定しているパラメタであるという考え方を提案した。具体的には，Mitsudo et al.（2009）は，両眼の回旋（視軸，視線方向周りの回転）によって引き起こされる立体錯視現象（Mitsudo, 2007）を説明するためにベイズ的アプローチを用いた。彼らのモデルでは，視覚系が3次元解釈を得つつ，回旋輻輳（両眼間で逆方向の視軸まわりの眼球回転）の量を推定していると考えることで，錯視の生起を説明できることが示されている。

Mitsudo et al.（2009）の説明では3次元構造の解釈を決定する原理として，両眼網膜像の局所的な相関を計算するモデル（Banks, Gepshtein, & Landy, 2004）を採用している[2)]。このモデルでは局所的な両眼対応の度合いを評価するために，ある程度の大きさの窓（相関窓）を考え，左右眼それぞれの像における窓内の情報をもとに相互相関係数を算出する。片眼の画像における窓の位置を固定し，もう片方の画像の窓の位置を変化させ，相関係数が最大となる位置ずれを求める。相関係数が最大となる像差を，その対応点についての局所的な像差量とする。この計算を画像の各位置で繰り返すことにより，視覚情報に基づく3次元解釈である“像差マップ”が得られる。

先行研究が解決していない重要な問題の一つに，両眼対応に基づく特定の3次元解釈の尤度を相関係数からどのように計算するか，という問題がある。ベイズ推定では得られた解釈の確からしさを確率に基づいて判断するため，相関係数の値自体を用いることはできない。直感的には，相関係数が高ければそれが本当に対応している（つまり，両眼が同じ対象を見ている）可能性が高いと考えることが自然である。したがって，両眼対応の確率は相関係数の単調増加関数となることが予測される。しかし具体的な関数形は自明ではなく，Mitsudo et al.（2009）では，第一近似としての線形関数

$$p=(1+r)/2 \tag{3}$$

を仮定し，尤度関数を求めている。本研究では，理論的な妥当性がより高いと考えられる尤度計算のアイデアを提案し，その妥当性を計算シミュレーションにより確認する。

理　　論

2次元の網膜像が2枚1組あり，それぞれの像上で輝度に対応する値（視覚情報）を受け取ると仮定する。この視覚情報は，画像中心を原点とした座標系によって表現できる。議論の単純化のため網膜は平面とし，像も平面とする。右眼像と左眼像の網膜座標系における特定の位置で画像の対応をとるとすると，両眼が別々の光景を見ている場合，局所的な相関係数の平均値は理論的には0となることが予想される。また，両眼が全く同一の情報を受け取る場合には，相関係数は1となるはずである。

単一の光景であるがいくらかの奥行きをもつ，生け垣のような対象を両眼で見る場合を考えてみよう。この場合，網膜座標系における同じ位置で相関係数を計算しても1とはならず，像差が生じている箇所では値は1より小さくなることが予想される。しかし，この場合でも同一の対象を見ているので，対応という観点からは対応の度合いを示す係数は1に近くなることが望ましい。そこで，相互相関モデル（Banks et al., 2004）では，前述のように両眼間の対応方向を定義した上で局所的な相関を計算し，相関係数が最大値となる位置ずれをもって両眼網膜像差の推定値としている。このような事態では，多少の像差をもつ視覚情報が与えられた場合でも，相関係数の値は1に十分近くなることが予想される。逆に，両眼が異なる光景を見ている場合は，両眼が同じ光景を見ている場合より相関係数は小さくなることが予想される。

本研究ではこのような相互相関モデルにおいて，両眼対応に基づく3次元解釈の尤度を計算する理論的アイデアを提案する。具体的には以下の通りである。まず，従来の相互相関モデルと同じく，左右眼の像の特定の位置について，ある程度の大きさの“窓”（相関窓）を定義し，相互相関係数を計算する。対応方向を定義した上で片眼の像をある方向へ移動させ，相互相関係数を再計算する。これは両眼対応の計算においてエピポーラ拘束条件（射影幾何学に基づき，特定の方位をもつ直線を用いて両眼対応を計算すること）を使っていることと同じである。各点の相関係数 r は，

$$r=\frac{\sum_{h=-c}^{c}\sum_{k=-c}^{c}(l_1(x+h,\,y+k)-\mu_1)(l_2(x+h+d,\,y+k)-\mu_2)}{\sqrt{\sum_{h=-c}^{c}\sum_{k=-c}^{c}(l_1(x+h,\,y+k)-\mu_1)^2}\sqrt{\sum_{h=-c}^{c}\sum_{k=-c}^{c}(l_2(x+h+d,\,y+k)-\mu_2)^2}} \tag{4}$$

によって与えられる。$l_1(x, y)$ および $l_2(x, y)$ は左眼および右眼像上の位置 (x, y) における輝度の階調値，d は像差，μ_1 および μ_2 はそれぞれ左眼および右眼の相関窓内の階調値の平均，c は相関窓の大きさを表す整数，h および k は整数である。この操作をある程度の像差計算の範囲内で繰り返し，相関係数が最大となる位置をその位置の像差とする（したがって，局所的な像差の推定には事前分布を仮定するベイズ的アプローチではなく，最尤法的アプローチが用いられると仮定している）。この操作を，与えられた画像の任意の位置で行う。本研究では，得られた両眼の網膜像が同一の対象から得られた確率を求めるために，同一の対象を見ていないと仮定し

[2)] Read（2002a, b）は，ベイズ的アプローチを像差エネルギーモデルに適用することで両眼立体視のモデル化を行っている。

た場合（つまり，エピポーラ拘束条件を使わない場合）の相関係数の確率分布を計算する。具体的には，両眼画像間で無関係な位置から得た相関係数の最大値分布を計算し，その累積分布をワイブル関数

$$p(r)=\exp(-[(1-r)/b]^{a}) \tag{5}$$

で近似する（Galambos, 1978）。a, b はワイブル分布の形状を決める係数である。この関数は $r=1$ のとき a, b に関わらず $p=1$ となる。この関数を微分すると，次式の確率密度関数が求まる。

$$p'=a/b^{a}(1-r)^{a-1}\exp(-[(1-r)/b]^{a}) \tag{6}$$

この関数は単峰の関数となり，両端の値の相関係数（1または−1）が得られる頻度は低くなることを意味している。得られた相関係数が1に近い場合は，それは“両眼が異なる光景を見ているとは考えにくい”事態，つまり“両眼が同じ光景を見ている”事態を反映していると考えることが自然である。両眼が異なる光景を見ているという仮定のもとで，相関係数がある値 r_0 より大きくなる確率 q_0（$r>r_0$）は，式 (6) の上側確率，つまり $q_0=1-p(r_0)$ によって与えられる。よって両眼対応の確からしさは，$1-q_0$ となり，直接，式 (5) によって評価できる。

以下に，視覚情報およびいくつかのパラメタの値を変化させたときに尤度関数がどのように変化するかを調べた結果を報告する。具体的には，(a) 画像の種類，(b) 相関窓の大きさ，(c) 像差計算の範囲が尤度に及ぼす影響を検討する。(b) と (c) に関しては，これらの値は通常は定数として扱うことが多い。定数として扱うことが妥当かどうかを検討し，定数として扱う際の注意点を考察する。

方　　法

左右像からなるステレオ画像を入力視覚情報とし，C言語のプログラムを用いて相互相関係数を式 (4) により計算した。ステレオ画像は256階調のグレースケール画像であった。ワイブル関数を決定する際には，相関を計算するための片方の画像の相関窓の位置をランダムとした。両眼対応を計算する際のエピポーラ線は画像上で常に水平であった。分析した画像は，ランダムドット画像（300×300ピクセル，約870個のドット），水平方向に階調値をサイン波変調した縦縞画像（250×250ピクセル，波長は10ピクセル），およびステレオ画像データベース（Scharstein & Szeliski, 2002: URL, http://vision.middlebury.edu/stereo/）の“barn 1”（216×190ピクセル）であった（Fig. 1A-D）。モデルのパラメタとして，像差計算の範囲および相関窓の大きさを変化させた。相関係数の分析には統計ソフトウェアRを使用し，ワイブル関数の当てはめには最小自乗法を用いた。

結　　果

モデルのパラメタを指定し，左右像間で無関係な位置の相関窓を用いて相関係数を算出した。相関係数の累積分布を，固定の区間（ここでは原則50区間）を用いて算出した。累積分布と，それに対するワイブル関数の当てはめ結果の例をFig. 2Aに示す。また，同じデータの相対頻度分布と，ワイブル関数を微分して求めた確率密度関数をFig. 2Bに示す。本研究の分析においては，用いたすべての画像およびほぼすべてのパラメタについてワイブル関数の係数は一定の条件で収束し，当てはめの妥当性は確認できた。

画像の種類と相関窓の大きさ

4.6°の水平剪断像差をもつランダムドット画像（Fig. 1A），“barn 1”（Fig. 1B），および一定の水平像差（2.5ピクセル）をもつ縦縞画像（Fig. 1C）を解析の対象とした。幾何学的には，水平剪断像差をもつステレオ刺激は水平軸まわりの奥行き方向の傾きを持つ面を作り出す。“barn 1”は，奥行きの異なる複数の面が重なり合う光景を作り出すステレオグラムである。一定の水平像差をもつ縦縞は，知覚的に不安定な奥行きを作り出す“壁紙”模様である。Fig. 1Aのランダムドット画像の最大水平像差は12ピクセル，“barn 1”の最大像差は10ピクセルであった。ランダムドット画像については，標準偏差2ピクセルの2次元ガウスフィルタを適用して，ぼけを付与した。いずれの画像も半径80ピクセルの円形領域の情報を分析の対象とし，像差の計算範囲は±15ピクセルとした。像差マップの例として，相関窓の一辺（幅および高さ）が11ピクセル，サンプリング間隔1ピクセルの場合の計算結果をFig. 1E-Gに示す。相関窓の幅を5ピクセルから21ピクセルまで変化させたときの対数尤度関数をFig. 3Aに示す。その結果，ランダムドット画像と“barn 1”については相関窓が大きくなるにつれて尤度が増加することが分かった。それに対し，縦縞画像については窓が大きくても尤度はかなり小さいままにとどまった。

この結果は，複雑な画像（ランダムドット画像と“barn 1”）については，相関窓がある程度大きい時に，モデルは両眼間の対応を確度高く計算できることを示している。これはBanks et al. (2004) の結果と一致している。相関窓を大きくしても尤度が増加しない縦縞画像については，モデルは両眼対応をうまく検出できないこと

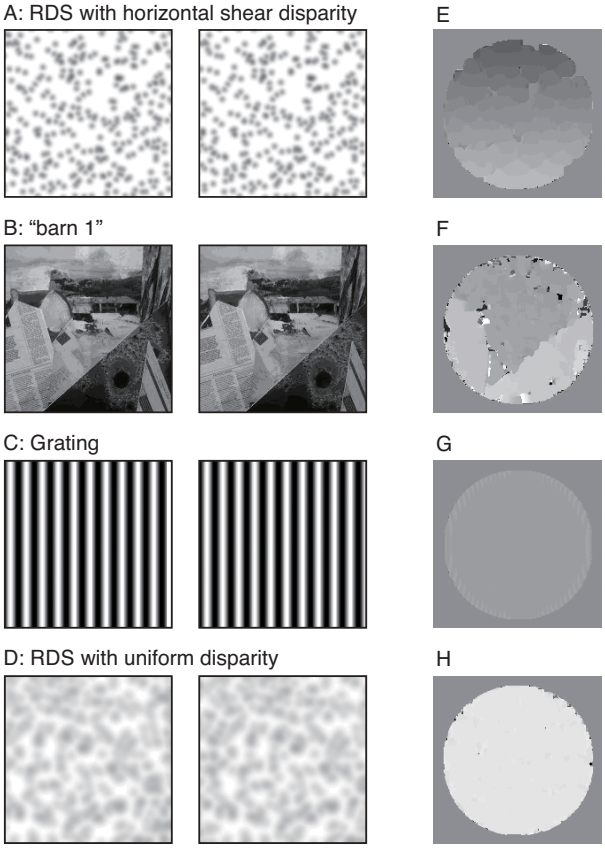

Fig. 1 分析したステレオ画像の例と像差マップの例。A は 4.6° の水平剪断像差をもつランダムドットステレオグラム（RDS），B は “barn 1”，C は縦縞刺激，D は単一の水平像差（20 ピクセル）をもつランダムドットステレオグラム。すべての画像は 190×190 ピクセルの大きさに切り取って表示し，A と D については実際の分析に用いた画像の階調値を反転して表示している。それぞれの刺激の奥行きを正しく知覚するためには，交差法によって両眼融合することが必要である。E-H はそれぞれ A-D の画像をもとに計算した像差マップである。分析対象は円形領域内の視覚情報であり，E-G については像差計算の範囲は±15 ピクセル，H については±25 ピクセルであった。背景の灰色が 0 ピクセルの像差を示し，それより明るいピクセルは交差性像差（手前），暗いピクセルは非交差性像差（奥）を示す。

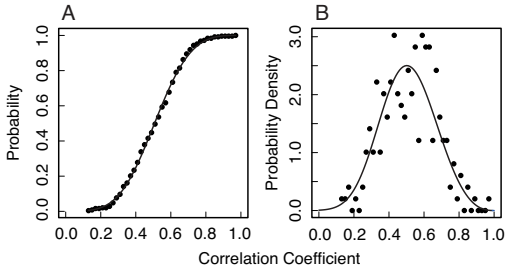

Fig. 2 両眼間対応がないと仮定して計算した場合の相関係数の分布の例。ステレオ画像 “barn 1” を解析対象とし，分析領域の半径は 80 ピクセル，像差の計算範囲は±15 ピクセル，相関窓の大きさは 21×21 ピクセルとした。A に累積分布およびワイブル関数の当てはめ結果を示し，B に相対頻度分布および当てはめたワイブル関数に基づく確率密度関数を示す。

が分かる。縦縞のように繰り返しのあるパタン（壁紙模様）の見えの奥行きは不安定であることが知られており（Howard & Rogers, 1995; Nakamizo, Ono, & Ujike, 1999），ワイブル関数を用いるモデルは人間の立体視の特性と一致していることが明らかとなった。

相関係数の値から確率を求める際にワイブル関数を用いず，Mitsudo et al.（2009）と同様に線形関数を用いた場合の解析結果を Fig. 3B に示す。その結果，尤度は窓の幅にそれほど依存しない結果が得られ，また縦縞刺激についての尤度は他の画像とは大きく変わらないことも分かった。これらの結果は，線形関数を用いたモデルは，人間の立体視の特性をうまく再現できないことを示唆する。

像差計算の範囲

前節で分析した水平剪断像差をもつランダムドット画像，“barn 1”，および一定の水平像差を設けたランダムドット画像を解析の対象とした。相関窓の大きさを一定に保ち（一辺は 9 ピクセル），像差計算の範囲を±1 ピクセルから±40 ピクセルまで変化させた。剪断像差をもつランダムドット画像の解析結果（解析領域の半径は 150 ピクセル）を Fig. 4A の実線に示す。尤度は，像差計算の範囲が解析画像の最大像差（12 ピクセル）をカバーするまでは急激に上昇し，像差計算の範囲が最大像差より大きくなると尤度は緩やかに減少することが分かる。

この傾向は，他のランダムドット画像（Fig. 1D：20 ピクセルの一定の水平像差，標準偏差 4 ピクセルの 2 次元ガウスフィルタを適用）を解析した場合でも確認された（Fig. 4A の細かい点線，解析領域の半径は 80 ピクセル）。また，刺激がランダムドットパタンでない画像，“barn 1” を対象とする解析でも同様の結果であった（Fig. 4A の荒い点線，解析領域の半径は 80 ピクセル）。

これらの結果から，両眼対応の尤度を計算する場合には，像差計算の範囲が画像の最大像差と重なっているかを確認することが重要であることが分かる。像差計算の範囲が最大像差と重なっていない場合の像差マップの例を Fig. 5 に示す。画像は 20 ピクセルの単一の水平像差をもつランダムドット画像（Fig. 1D）であり，像差計算の範囲は±15 ピクセルであった。Fig. 1H とは異なり，期待される 3 次元形状は推定できていないことが分かる。

実際のシミュレーションでは，像差計算の範囲は，画像の大きさと両眼融合の限界（視角で 1-2° 程度）を考慮して決めることが理に適っているが，それでもどの程度の値を設定するかは自由度が高い。解析結果は，融合範囲が像差計算の範囲に含まれているならば，像差計算の範囲は尤度計算にそれほど大きな影響を及ぼさないと考えて良いことを示唆する。

尤度を計算する際，式 (5) のワイブル関数の代わりに式 (3) の線形関数を用いた場合の結果を Fig. 4B に示す。結果はワイブル関数を用いる場合（Fig. 4A）と類似しているが，尤度関数の変化が小さくなる計算範囲はそれほど明確には得られないことが分かる。

考　　察

本研究では，両眼立体視のモデル化の際に重要となる両眼対応に基づく 3 次元解釈の尤度を計算する方法を提案した。提案した方法を用いる場合，視覚情報に依存する尤度の違いは人間の立体視の特性と一致していることが分かった。加えて，定数として扱われることが多い，像差計算の範囲と相関窓の大きさを変化させた場合の尤度関数を具体的に求めた。像差計算の範囲については，刺激のもつ像差が像差の計算範囲に入っているかどうかが尤度を大きく変化させる要因であることが分かった。このことは，人間の立体視過程をモデル化する際には，両眼対応計算の像差範囲が人間の処理特性と一致しているかどうかを考慮する必要があることを示している。

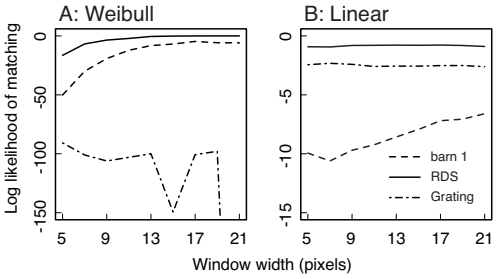

Fig. 3 相関窓の幅を変化させた場合に得られる対数尤度関数。対数の底は 10 とした。ワイブル関数 (A) および線形関数 (B) を用いた解析結果の例を示す。

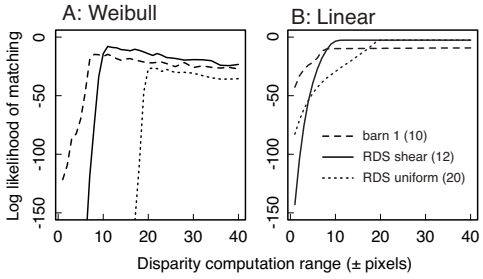

Fig. 4 像差計算の範囲を変化させた場合に得られる対数尤度関数。対数の底は 10 とした。ワイブル関数 (A) および線形関数 (B) を用いた解析結果の例を示す。

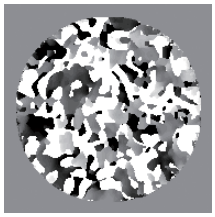

Fig. 5 像差計算の範囲が狭い場合の像差マップの例。Fig. 1D **の画像（20ピクセルの均一な水平像差をもつ画像）について，像差計算の範囲を±15 ピクセルとして計算した場合の解析結果を示す。**

また，空間的な繰り返しをもたない複雑な刺激の場合には，相関窓が大きいほど両眼対応の尤度は増加することが分かった（Fig. 3A）。つまり，両眼対応を正確に判断するためには，相関窓を大きくすることが効果的であることを示唆している。ただ相関窓を大きくしすぎると，推定される 3 次元構造の解像度が低くなるため（Banks et al., 2004），相関窓の大きさは解析したい対象の立体解像度を考慮しつつ決定する必要がある。

引 用 文 献

Banks, M. S., Gepshtein, S., & Landy, M. S. (2004). Why is spatial stereoresolution so low? *The Journal of Neuroscience*, **24**, 2077-2089.

Brainard, D. H., & Freeman, W. T. (1997). Bayesian color constancy. *Journal of the Optical Society of America A: Optics, Image Science, and Vision*, **14**, 1393-1411.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, **415**, 429-433.

Galambos, J. (1978). *The asymptotic theory of extreme order statistics. New York*: John Wiley & Sons.

Howard, I. P., & Rogers, B. J. (1995). *Binocular vision and stereopsis*. New York: Oxford University Press.

Knill, D. C., Kersten, D., & Yuille, A. (1996). Introduction: A Bayesian formulation of visual perception. In D. C. Knill, & W. Richards (Eds.), *Perception as Bayesian inference*. Cambridge: Cambridge University Press. pp. 1-21.

Mitsudo, H. (2007). Illusory depth induced by binocular torsional misalignment. *Vision Research*, **47**, 1303-1314.

Mitsudo, H., Kaneko, H., & Nishida, S. (2009). Perceived depth of curved lines in the presence of cyclovergence. *Vision Research*, **49**, 348-361.

Nakamizo, S., Ono, H., & Ujike, H. (1999). Subjective staircase: a multiple wallpaper illusion. *Perception & Psychophysics*, **61**, 13-22.

Read, J. C. A. (2002a). A Bayesian approach to the stereo correspondence problem. *Neural Computation*, **14**, 1371-1392.

Read, J. C. A. (2002b). A Bayesian model of stereopsis depth and motion direction discrimination. *Biological Cybernetics*, **86**, 117-136.

Scharstein, D., & Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, **47**, 7-42.

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience*, **5**, 598-604.