

第 10 章

終章

10.1 まとめ

本論文では、実世界画像の認識を実現するための研究として、主に (1) マルチエージェントによる多様な画像に対応可能な画像認識システム構成法の提案、(2) 多様な実世界画像が存在する World-Wide Web (WWW) から自動的に画像を収集とそれを用いた実世界画像の自動分類の提案、の 2 つについて述べた。

第 3 章から第 6 章では、前半部として、実世界画像に対応したシステムを構築するために、異なる多数の認識手法や知識表現の統合をマルチエージェントの考え方に基づいて実現する方法について提案した。

第 3 章では、従来の実世界画像に対する認識システムの構成法についてまとめ、従来の認識システムでは、例えば、屋外画像のみ、航空写真のみ、という様に対象を予め想定してシステム構築が行われてきたことを述べた。さらに、そのため、システムの構成が認識対象の画像の種類に依存したものになっていたり、各対象物の認識のための知識が相互に密接な関係を持っていたために、異なる種類の画像に対する知識を混在させることが困難で、様々な種類の画像が存在する実世界画像の認識には適用が難しいという問題点を指摘した。

第 4 章では、マルチエージェントによる画像認識システムの構築法を提案した。多様な認識対象に対応するために、マルチエージェントによって、独立した認識システムである 1 種類の物体のみを認識するエージェントを統合し、画像認識システムの柔軟な構築方法を実現した。マルチエージェントを用いることによって、物体毎にそれぞれに異なる知識の表現および認識手法を用いることが可能となり、システムの全体の最終的な認識結果は、エージェント間の相互作用によって求める。実験により、提案手法によって室内画像と屋外画像の両方に対応できる認識システムを構築出来ることを示した。

第 5 章では、マルチエージェントによる画像認識システムに物体間の定性的な位置関係に関する推論機構を導入することを提案し、より複雑な画像の認識を可能とした。実験では、室内画像に対するシステムを実現し、その効果を示した。実世界画像においては、物体が物体の上に載ったり、手前に位置したりして、物体が物体を隠すオクルージョンが発生する。オクルージョンがあつて物

体の一部しか画像上に現れていなくても認識可能とするために、物体間の位置関係を利用することが不可欠であるが、従来は主に画像上での物体領域同士の 2 次元的な位置関係しか利用されていなかったのに対して、物体の定性的な 3 次元情報を利用して、定性的な 3 次元位置関係の推論を行うことにより、実世界画像で問題となるオクルージョンに対処する方法を提案した。

第 6 章では、画像中に小さくしか現われていない対象の認識を高解像度画像を利用することによって認識可能とする方法を提案した。単純に高解像度画像を用いることは、認識時間の著しい増大を招くが、ここでは、多重解像度解析を導入することによって、効率的な認識を実現することを提案し、より複雑な実世界画像が認識可能となることを実験にて示した。

第 7 章から第 9 章では、後半部として、画像内容を表すテキスト情報を伴った多種多様な画像を WWW から自動収集することによって、画像認識のための知識ベースを自動構築し、同一種類でも多様な個体が含まれる実世界画像を認識可能なシステムを実現する方法について提案した。

第 7 章では、多数の学習画像を用いた実世界画像の認識について従来の研究をまとめた。そして、従来の研究では、学習画像を収集することが困難であったために、顔画像や自動車の画像などの限定された対象にしか実験が行われていなかったという問題点を指摘した。

第 8 章では、実世界画像を大量にしかも手軽に収集する方法として、WWW から自動的に大量の実世界画像を収集する方法について提案した。WWW 空間中に存在する画像は現在数億枚と言われ、様々な画像が存在している。WWW 空間中に存在する画像はその多くが画像内容を表すテキスト情報を伴っているので、テキスト情報を解析することによって、ユーザの望むあらゆる画像を WWW から収集することが可能であることを実験により示した。

第 9 章では、提案した画像収集法を用いて様々な実世界画像を自動収集し、それらを学習画像として、実世界画像に対する認識を行うことを提案した。最初に学習段階では、WWW から認識したい対象の画像を各種類 (クラス) 毎に数百枚から数千枚程度収集し、それらから色情報、テキスト情報などを画像特徴として抽出して、各クラス毎に画像特徴に関する知識ベースを構築する。次に、認識段階では認識対象の画像から同様に画像特徴を抽出し、知識ベースと照合を行い、最も確率の高いクラスに分類し認識を行った。実験では、20 クラスの場合に限定のない一般的な実世界画像に対して、約 5 割の精度で自動分類することができ、有効性が示された。

最後に、本論文の前半部分と後半部分の関係を整理する。

前半では、多数の認識モジュールの統合の手法を述べた。実験では、主にオクルージョンの多い室内画像を対象としており、エージェントの間の相互作用によって、オクルージョンがあっても正しく認識出来ることが示されている。個々の物体の認識は、人手によって構築 (hand-coding) された認識モジュールによる、物体の本質 (機能) 構造に注目した形状モデル当てはめで行っている。「机」「椅子」の様な形状が比較的単純な人工物に対しては、実験から、この方法がある程度有効であることが示された。ただし、人手による構築のため、人工物以外の物体や、形状が複雑な人工物に対しては、認識モジュールの作成が困難であるという問題点があった。

そこで、後半では、人工物の以外の物体や形状の複雑な物体など、形状モデルを人手によって構築することが困難な物体に対する認識を実現するために、学習によって実世界画像の認識を行うこ

とを試みた。人手によって認識モジュールを構築する場合は、物体の物体の本質（機能）構造に注目したモデルを用意することで、物体の形状の多様性にある程度対処することができた。ところが、学習による方法では、現状では学習画像だけから物体の本質（機能）構造を学習することが難しい。そこで、本論文では、WWW から多数の多様な画像を収集し、それを学習画像として利用することで、物体の多様性に学習画像の量で対処することとした。WWW 上の画像は通常 Web ページからリンクされており、リンク元の Web ページを解析することによって、画像内容がかなり高い精度で推定できる。そのため、キーワード入力のみで容易に多種類の学習画像を WWW から大量に自動収集することが可能である。実験では、WWW から自動収集した画像を用いて、5 割程度の精度で実世界画像が認識可能となった。

後半部で用いた手法による物体の認識手法は基本的に 2 次元画像同士の照合なので、物体の構造は考慮されていない。それに対して前半部では、物体の構造を考慮することによって、物体間の支持関係を推論し、複雑なオクルージョンがあっても認識が可能となっている。つまり、前半の研究は、「机」「椅子」などの比較的形状が簡単な人工物を対象としているが、その代り複雑なオクルージョンにも対処できる。一方、後半の研究では、画像中に対象がはっきりと写っている場合でないと認識は難しいが、モデル作成が難しい多くの種類の物体に対して適用可能である。以上の様に、前半部と後半部の研究はそれぞれ互いに相補的な関係にあると言える。

10.2 今後の課題 と 展望

本論文の前半部では対象の種類毎に適切な認識方法を用意して、それらをマルチエージェントによって統合することによって、実世界画像の認識を実現した。一方、後半部では WWW から収集した画像を学習画像として利用して、対象に依らない認識方法として画像検索手法を用いて実世界画像の自動分類を行った。今後の大きな課題としては、両方の手法を統合することが考えられる。つまり、WWW から収集した画像を用いてある特定対象についての認識モジュールを自動構築するということである。方法としては、認識モジュールを構築するために様々な画像特徴、分類方法を用意して、交差法 (cross validation) によって、最適な分類結果を残す方法を選択し、認識モジュールを自動構築することが考えられる。その際に、本論文で行った画像をそのまま 1 枚を分類するのではなく、画像の一部分を切り出して、それを分類することが必要になる。現在、単語が付与された画像を大量に用意して、それらの画像に共通に含まれる部分画像を抽出して、単語に対応する部分画像を自動的に学習する研究がいくつか行われている。こうした研究で用いられている手法を応用することによって、WWW からの収集画像に対しても単語に対応する画像の一部分を自動的に切り出すことが可能となると思われる。また、共通部分を予め単語に対応する部分画像であると学習することはせずに、単純に 2 枚の画像の一部分同士のマッチングを行うことによって、画像の一部分の切出しが実現できる可能性もある。ただし、その場合は照合する組み合わせが非常に多くなるので、高速な手法の開発や、PC クラスタ等の並列計算機システムの利用が不可欠である。

また、本論文の認識や画像分類においては、「単語（名詞）＝部分画像」という単純な対応を仮定していた。しかし、実際には、物体を表す名詞には階層構造があり、画像中に現れている実際の物

体にも部分と全体の階層構造が存在している。本研究においては、こうした階層構造を完全に無視して認識を行っているが、実際にはこうした言語の階層構造と、実際の物体の物理的な構造の階層構造を認識に利用することは可能であり、今後の課題であるといえる。

今後、ロボットのハードウェアが発達していくにつれ、実世界で行動するロボットの目としての一般の実世界画像の認識への需要が高まることが予想される。また、今後は、銀盤カメラに代わって、デジタルカメラがカメラの主流になることが予想され、大量に蓄積されたデジタル画像に対する画像の意味内容に基づく検索への需要が高まると考えられる。この場合、意味的検索を精度よく柔軟に行うためには、実世界画像の一般的な認識の実用化が不可欠である。

今後、この様に重要でかつ不可欠な技術となると考えられる実世界画像の一般的な認識の実現のためには、本論文の前半部で示した個別のモデルに基づく従来からの画像認識的な手法と、本論文の後半部で示した画像特徴と対象の関連付けを大量の学習画像に基づいて行う画像データベース的な手法 (もしくはデータマイニング的手法) をうまく融合して行くことが重要であると考ええる。

究極的には、辞書に出ている普通名詞の数だけのクラスを分類可能とすることが「実世界画像の認識」の研究の目標であり、それが可能となれば、真の「実世界画像認識システム」が実現されたと言えることができるであろう。ただし、その道のりは現状ではまだまだ長く、そして大変険しいものである。

謝辞

本論文は、私が東京大学大学院工学系研究科情報工学専攻修士課程在学中の2年間と博士課程在学中の半年間、及びその後助手として現在まで在職中である電気通信大学での5年間に行った研究をまとめたものです。その間、とても多くの方々にお世話になりながら、ここに博士論文として完成することができました。

特に、大学院時代の指導教官であった東北大学大学院情報科学研究科の出口光一郎教授には大学院進学以来、いつも研究の相談にのって頂きました。本論文の前半部分については、私が東京大学大学院に在学中から研究指導して頂いた内容であり、共著の論文を多数書かせて頂きました。本論文執筆にあたりまして、様々なご助言を頂きました。先生のお力添えなしでは博士論文は完成しなかったと思います。心から感謝申し上げます。また、予備審査、本審査の時には、お忙しい中、わざわざそのために仙台から東京までお越し下さいまして、大変有難うございました。

主査の労を取って頂きました東京大学大学院情報理工学研究科の武市正人教授に感謝致します。武市先生は8年前に私が東京大学工学部計数工学科に在学中に卒業研究についてご指導頂いて以来、お世話になっております。博士論文執筆にあたりましては多くのアドバイスを頂きました。大変有難うございました。

本論文の審査をして頂きました東京大学大学院情報理工学研究科田中英彦教授、井上博允教授、杉原厚吉教授、石塚満教授には、審査を通じて数多くの有益なご示唆を頂きました。本当に有難うございました。

山形大学工学部の本谷秀堅助教授には、私の大学院時代に当時の東大出口研究室の助手として大変お世話になりました。また、博士論文を書くにあたって、アドバイス、激励のお言葉を頂きました。大変感謝しております。

電気通信大学情報工学科の野下浩平教授、中山泰一助教授には、普段より公私問わず大変お世話になっております。研究に関しても、常にご助言を頂いて下ります。深く感謝申し上げます。

その他、旧東大出口研究室、そして電通大野下研究室の皆様には、輪講・ゼミや学会発表の練習の時などに貴重な御意見、御感想を頂戴致しました。どうも皆さん有難うございました。

最後に、論文執筆中や審査直前で忙しい時でも常に普段と変わらない態度で接してくれた家族に感謝致します。

2003年3月

柳井 啓司

参考文献

- [1] E. Rosch, C. B. Mervis, W. D. Gray, D. M. Johnson, and P. Boyes-Braem. Basic objects in natural categories. *Cognitive Psychology*, Vol. 8, pp. 382–439, 1976.
- [2] 柳井啓司, 出口光一郎. マルチエージェントによる多様な画像に対応した物体認識システムの一構成法. 情報処理学会論文誌, Vol. 39, No. 2, pp. 170–177, 1998.
- [3] K. Yanai and K. Deguchi. An architecture of object recognition system for various images based on multi-agent. In *Proc. of 14th International Conference of Pattern Recognition*, Vol. 1, pp. 278–281, 1998.
- [4] 柳井啓司, 出口光一郎. 物体間の支持関係を利用した室内画像の認識. 電子情報通信学会論文誌 D-II, Vol. 84-DII, No. 8, pp. 1741–1752, 2001.
- [5] K. Yanai and K. Deguchi. Recognition of indoor images employing supporting relation between objects. *Systems and Computers in Japan*, Vol. 33, No. 11, pp. 14–26, 2002.
- [6] K. Yanai and K. Deguchi. Recognition of indoor images employing qualitative model fitting and supporting relation between objects. In *Proc. of 15th International Conference of Pattern Recognition*, Vol. 1, pp. 964–967, 2000.
- [7] K. Yanai and K. Deguchi. A multi-resolution image understanding system based on multi-agent architecture for high-resolution images. *IEICE Transactions on Information and Systems*, Vol. E84-D, No. 12, pp. 1642–1650, 2001.
- [8] K. Yanai and K. Deguchi. A multi-resolution image understanding system based on multi-agent architecture for high-resolution images. In *Proc. of IAPR Workshop on Machine Vision and Applications 2000*, pp. 291–294, 2000.
- [9] 柳井啓司. キーワードと画像特徴を利用した WWW からの画像収集システム. 情報処理学会論文誌:データベース, Vol. 42, No. SIG10 (TOD11), pp. 79–91, 2001.
- [10] K. Yanai. Image collector: An image-gathering system from the World-Wide Web employing keyword-based search engines. In *Proc. of IEEE International Conference on Multimedia and Expo*, pp. 704–707, 2001.
- [11] K. Yanai, M. Shindo, and K. Noshita. A fast image-gathering system on the World-Wide Web using a PC cluster. In *Proc. of the 2001 International Conference on Web Intelligence*

- (Springer LNAI no.2198), pp. 324–334, 2001.
- [12] K. Yanai. Image classification by Web images. In *Proc. of the Seventh Pacific-Rim International Conference on Artificial Intelligence (Springer LNAI no.2417)*, pp. 613–614, 2002.
 - [13] K. Yanai. An experiment on generic image classification using Web images. In *Proc. of the Third IEEE Pacific-Rim Conference on Multimedia (Springer LNCS no.2532)*, pp. 303–310, 2002.
 - [14] 柳井啓司. WWWからの大量の収集画像を用いた画像認識の試み. 電子情報通信学会パターン認識・メディア理解研究会報告 PRUM2001-165, pp. 79–86, 2001.
 - [15] M. J. Swain and D. H. Ballard. Color indexing. *International Journal of Computer Vision*, Vol. 7, No. 1, pp. 11–32, November 1991.
 - [16] Y. Rubner, C. Tomasi, and L. J. Guibas. The earth mover’s distance as a metric for image retrieval. *International Journal of Computer Vision*, Vol. 40, No. 2, pp. 99–121, 2000.
 - [17] J. Z. Wang, J. Li, and G. Wiederhold. SIMPLIcity: semantics-sensitive integrated matching for picture libraries. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 9, pp. 947–963, 2001.
 - [18] 片山紀生, 佐藤真一. マルチメディア情報の大規模処理に向けた多次元インデクシング手法の応用. 電子情報通信学会論文誌 D-II, Vol. J82-D-II, No. 10, pp. 1606–1616, 1997.
 - [19] H. Murase and S. K. Nayar. Visual learning and recognition of 3-D objects from appearance. *International Journal of Computer Vision*, Vol. 14, No. 9, pp. 5–24, 1995.
 - [20] S. Ullman. *High-level Vision*. The MIT Press, 1996.
 - [21] 谷内田正彦. 画像処理から画像理解へ. 人工知能学会誌, Vol. 4, No. 1, pp. 5–12, 1989.
 - [22] C. G. Perrott and L. G. C. Hamey. Object recognition, a survey of the literature. Technical Report No.91-0065C, School of MPCE, Macquarie University, 1991.
 - [23] P. Suetens. Computation strategies for object recognition. *ACM Computing Survey*, Vol. 24, No. 1, pp. 5–61, 1992. (訳:松山隆司: 対象認識のための計算方式, 別冊 Bit ACM Computing Surveys ’92, pp. 5–59 (1994)).
 - [24] A. R. Pope. Model-based object recognition: A survey of recent research. Technical Report TR-94-04, University of British Columbia, Computer Science Department, 1994.
 - [25] A. Zisserman, D. Forsyth, J. Mundy, C. Rothwell, J. Liu, and N. Pillow. 3D object recognition using invariance. *Artificial Intelligence*, Vol. 26, No. 78, pp. 239–388, 1995.
 - [26] D. Crevier and R. Lepage. Knowledge-based image understanding systems: A survey. *Computer Vision and Image Understanding*, Vol. 67, No. 2, pp. 161–185, 1997.

- [27] J. M. Tenenbaum and H. G. Barrow. Experiments in interpretation guided segmentation. *Artificial Intelligence*, Vol. 8, pp. 241–274, 1977.
- [28] 松山隆司, ビンセントハング. 画像理解システム SIGMA. 情報処理学会論文誌, Vol. 26, No. 5, pp. 877–889, 1985.
- [29] M. Nagao, T. Matsuyama, and H. Mori. Structural analysis of complex aerial photographs. In R. Englemore and T. Morgan, editors, *Blackboard Systems*, pp. 219–242. Addison-Wesley, London, 1988.
- [30] Y. Ohta. *Knowledge-Based Interpretation of Outdoor Natural Color Scenes*. Pitman Advanced Publishing Program, Boston, 1985.
- [31] B. Draper, R. Collins, J. Brolio, A. Hanson, and E. Riseman. The schema system. *International Journal of Computer Vision*, Vol. 3, No. 2, pp. 209–250, 1989.
- [32] T. Matsuyama and V. S. Hwang. *SIGMA: A knowledge-based aerial image understanding system*. Plenum Press, New York, 1990.
- [33] R. A. Brooks. Model-based three-dimensional interpretations of two-dimensional image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 5, No. 2, pp. 140–150, 1983.
- [34] D. H. Ballard and C. M. Brown. *Computer Vision*. Prentice Hall, 1982.
- [35] D. Marr. *Vision*. Freeman, 1982. (乾, 安藤 (訳): ビジョン, 産業図書 (1985)).
- [36] J. Batlle, A. Casals, J. Freixenet, and J. Marti. A review on strategies for recognizing natural objects in colour images of outdoor scenes. *Image and Vision Computing*, Vol. 18, No. 6-7, pp. 515–530, 2000.
- [37] L. Stark and K. Bowyer. Achieving generalized object recognition through reasoning about association of function to structure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13, No. 10, pp. 1097–1104, 1991.
- [38] T. M. Strat and M. A. Fischler. Context-based vision: Recognizing objects using information from both 2-D and 3-D imagery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13, No. 10, pp. 1050–1065, 1991.
- [39] K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Morgan Kaufmann, 1990.
- [40] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification –Second Edition–*. John Wiley & Sons, 2000.
- [41] M. Turk and A.P. Pentland. Eigenfaces for recognition. *Cognitive Neuroscience*, Vol. 3, No. 1, pp. 71–96, 1991.

- [42] H.A. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 1, pp. 23–38, January 1998.
- [43] K.K. Sung and T. Poggio. Example-based learning for view-based human face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 1, pp. 39–51, January 1998.
- [44] H. Schneiderman and T. Kanade. A statistical method for 3D object detection applied to faces and cars. In *Proc. of IEEE Computer Vision and Pattern Recognition*, Vol. I, pp. 746–751, 2000.
- [45] D. A. Forsyth and M. M. Fleck. Automatic detection of human nudes. *International Journal of Computer Vision*, Vol. 32, No. 1, pp. 63–77, 1999.
- [46] D.A. Forsyth and M.M. Fleck. Body plans. In *Proc. of IEEE Computer Vision and Pattern Recognition*, pp. 678–683, 1997.
- [47] 松山隆司, 尾崎正治. LLVE: トップダウンセグメンテーションのための画像エキスパートシステム. 情報処理学会論文誌, Vol. 27, No. 2, pp. 191–204, 1986.
- [48] 長谷川純一, 久保田浩明, 鳥脇純一郎. サンプル図形提示方式による画像処理エキスパートシステム IMPRESS. 電子情報通信学会論文誌 D, Vol. J70–D, No. 11, pp. 2147–2153, 1987.
- [49] 田村秀行, 佐藤宏明, 坂上勝彦, 久保文雄. DIA-Expert システムとその知識表現方法. 情報処理学会論文誌, Vol. 29, No. 2, pp. 190–208, 1988.
- [50] V. Clement and M. Thonnat. A knowledge-based approach to integration of image processing procedures. *Computer Vision, Graphics and Image Processing*, Vol. 57, No. 2, pp. 166–184, 1993.
- [51] B. A. Draper, A. R. Hanson, and E. M. Riseman. Knowledge-directed vision: Control, learning, and integration. *Proceedings of the IEEE*, Vol. 84, No. 11, pp. 1625–1637, 1996.
- [52] R. Clouard, A. Elmoataz, C. Porquet, and M. Revenu. Borg: A knowledge-based system for automatic generation of image processing programs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 21, No. 2, pp. 128–144, 1999.
- [53] 濱田敏宏, 清水昭伸, 長谷川純一, 鳥脇純一郎. 誤り確率条件に基づく画像処理手順の自動構成の一方法と画像処理エキスパートシステム IMPRESS-Pro. 情報処理学会論文誌, Vol. 41, No. 7, pp. 1937–1947, 2000.
- [54] B.A. Draper, J. Bins, and K. Baek. ADORE: adaptive object recognition. *Videre: Journal of Computer Vision Research*, Vol. 1, No. 4, pp. 86–99, 2000.
- [55] S. Belongie, C. Carson, H. Greenspan, and J. Malik. Recognition of images in large databases using a learning framework. Technical Report 07-939, UC Berkeley CS Tech

Report, 1997.

- [56] K. Barnard and D. A. Forsyth. Learning the semantics of words and pictures. In *Proc. of IEEE International Conference on Computer Vision*, Vol. II, pp. 408–415, 2001.
- [57] 森靖英, 高橋裕信, 岡隆一. 単語群つき画像の分割クラスタリングによる未知画像からの関連単語推定. 電子情報通信学会論文誌 D-II, Vol. J84-D-II, No. 4, pp. 649–658, 2001.
- [58] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 5, pp. 530–535, 1997.
- [59] C. Harris and M.J. Stephens. A combined corner and edge detector. In *Proc. of Alvey Conference*, pp. 147–152, 1988.
- [60] D. G. Lowe. Object recognition from local scale-invariant features. In *Proc. of IEEE International Conference on Computer Vision*, pp. 1150–1157, 1999.
- [61] M. Weber, M. Welling, and P. Perona. Towards automatic discovery of object categories. In *Proc. of IEEE Computer Vision and Pattern Recognition*, Vol. I, pp. 101–108, 2000.
- [62] M. Weber, M. Welling, and P. Perona. Unsupervised learning of models for recognition. In *Proc. of European Conference on Computer Vision*, pp. 18–32, 2000.
- [63] I. Biederman. Human image understanding: Recent research and a theory. *Computer Vision, Graphics and Image Processing*, Vol. 32, No. 1, pp. 29–73, 1985.
- [64] D. G. Lowe. *Perceptual Organization and Visual Recognition*. Kluwer Academic Publishers, 1985.
- [65] S. Ullman and R. Basri. Recognition by linear combinations of models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13, No. 10, pp. 992–1006, 1991.
- [66] R. Basri. Recognition by prototypes. *International Journal of Computer Vision*, Vol. 10, No. 2, pp. 147–167, 1996.
- [67] R. Basri. Recognition by prototypes. Technical Report A.I.Memo No.1391, MIT A.I.Lab., 1992.
- [68] 村瀬洋, V. V. Vinod. ヒストグラム特徴を用いた高速物体探索法—アクティブ探索法. 電子情報通信学会論文誌 D-II, Vol. J81-D-II, No. 9, pp. 2035–2042, 1998.
- [69] 川西隆仁, 村瀬洋. 色ヒストグラム特徴とパン・チルト・ズームカメラを用いた高速物体探索法—動的アクティブ探索法. 電子情報通信学会論文誌 D-II, Vol. J84-D-II, No. 8, pp. 1722–1730, 2001.
- [70] B. Schiele and J. L. Crowley. Recognition without correspondence using multidimensional receptive field histograms. *International Journal of Computer Vision*, Vol. 36, No. 1, pp. 31–50, 2000.

- [71] L. Stark. Functionality in object recognition. *Computer Vision and Image Understanding*, Vol. 62, No. 2, pp. 145–146, 1995.
- [72] J. J. Gibson. *The Ecological Approach to Visual Perception*. Houghton Mifflin Company, 1979. (古崎敬, 古崎愛子ら (訳): ギブソン生態学的視覚論, サイエンス社 (1985)).
- [73] L. Stark and K. Bowyer. Function-based generic recognition for multiple object categories. *Computer Vision and Image Understanding*, Vol. 50, No. 1, pp. 1–21, 1994.
- [74] T. M. Strat. Employing contextual information in computer vision. Technical report, SRI International Artificial Intelligence Center, 1992. <http://www.ai.sri.com/radius/sri/baa-reports/context-paper.ps>.
- [75] J. Marti, J. Freixenet, J. Batlle, and A. Casals. A new approach to outdoor scene description based on learning and top-down segmentation. *Image and Vision Computing*, Vol. 19, No. 14, pp. 1041–1055, 2001.
- [76] 大田友一, 金出武雄, 坂井利之. ボトムアップ制御とトップダウン制御の組合わせによる領域解析. 情報処理学会論文誌, Vol. 21, No. 2, pp. 116–124, 1980.
- [77] R. S. Englemore, A. J. Morgan, and H. P. Nii. Hearsay-ii. In R. Englemore and T. Morgan, editors, *Blackboard Systems*, pp. 25–86. Addison-Wesley, London, 1988.
- [78] 石田亨, 桑原和宏. 分散人工知能 (1): 協調問題解決. 人工知能学会誌, Vol. 7, No. 6, pp. 945–954, 1992.
- [79] 桑原和宏, 石田亨. 分散人工知能 (2): 交渉と均衡化. 人工知能学会誌, Vol. 8, No. 1, pp. 17–25, 1993.
- [80] 所真理雄. マルチエージェントシステム研究の目指すもの. コンピュータソフトウェア, Vol. 12, No. 1, pp. 78–84, 1995.
- [81] Y. Demazeau and J. P. Muller. Decentralized artificial intelligence. In Y. Demazeau and J. P. Muller, editors, *Decentralized A.I. (Proc. of the 1st European Workshop on Modelling Autonomous Agents in a Multi-Agent World (MAAMAW))*, pp. 3–13. Elsevier Science Publishers, 1990.
- [82] 松山隆司. AI マップ—ビジョン研究から見た統合アーキテクチャ—. 人工知能学会誌, Vol. 10, No. 6, pp. 888–894, 1995.
- [83] 松山隆司, 大田友一編. 小特集「並列・分散協調画像理解」. 人工知能学会誌, Vol. 9, No. 5, pp. 629–643, 1994.
- [84] 松山隆司. 分散協調処理による画像理解. 計測と制御, Vol. 31, No. 11, pp. 1149–1154, 1992.
- [85] 渡辺正規, 大田友一, 池田克夫. 画像理解のための並列処理機構. 電子情報通信学会論文誌 D-II, Vol. J73, No. 6, pp. 820–829, 1990.

- [86] O. Boissier and Y. Demazeau. A distributed artificial intelligence view on general purpose vision systems. In E. Werner and Y. Demazeau, editors, *Decentralized A.I.3 (Proc. of the 3rd European Workshop on Modelling Autonomous Agents in a Multi-Agent World (MAAMAW))*, pp. 311–330. Elsevier Science Publishers, 1992.
- [87] Y. Ohta, M. Watanabe, and Y. Sumi. Approaches to parallel computer vision. *IEICE Transactions on Information and Systems*, Vol. E74, No. 2, pp. 417–426, 1991.
- [88] 渡辺正規, 大田友一. 複数ステレオ対応探索アルゴリズムの協調的統合. 電子情報通信学会論文誌 D-II, Vol. J73, No. 10, pp. 1666–1676, 1990.
- [89] 角保志, 大田友一. 多様な入力を許容する顔画像解析システムの一構成法. 電子情報通信学会論文誌 D-II, Vol. J75, No. 2, pp. 236–245, 1992.
- [90] 松山隆司, 長尾真. 航空写真の構造解析. 情報処理学会論文誌, Vol. 21, No. 5, pp. 468–480, 1980.
- [91] M. Nagao and T. Matsuyama. *A Structure of Complex Aerial Photographs*. Plenum Press, New York, 1980.
- [92] B. Draper, R. Collins, J. Brolio, A. Hanson, and E. Riseman. Issues in the development of a blackboard-based schema system for image understanding. In R. Englemore and T. Morgan, editors, *Blackboard Systems*, pp. 189–218. Addison-Wesley, London, 1988.
- [93] F. V. Jensen, H. I. Christensen, and J. Nielsen. Bayesian methods for interpretation and control in multi-agent vision systems. In *Proc. SPIE 1708, Applications of Artificial Intelligence X: Machine Vision and Robotics*, pp. 536–548, 1994.
- [94] O. Boissier and Y. Demazeau. Asic : An architecture for social and individual control and its application to computer vision. In *Proc. of European Workshop on Modelling Autonomous Agents in a Multi-Agent World (MAAMAW) (LNAI no.1069)*, pp. 107–118, 1994.
- [95] O. Boissier and Y. Demazeau. Mavi: A multi-agent system for visual integration. In *Proc. of IEEE Conference on Multisensor Fusion and Integration for Intelligent Systems*, pp. 731–738, 1994.
- [96] T. Graf and A. Knoll. A multi-agent approach to self-organizing vision systems. In *Proc. of the International Conference on Intelligent Agent Technologies*, 1999.
- [97] T. Graf and A. Knoll. Driving vision systems by communication. In *Proc. of the Int. Conf. on Information, Intelligence, and Systems*, 1999.
- [98] T. Graf and A. Knoll. A multi-agent system architecture for distributed computer vision. *International Journal on Intelligent Agent Tools*, Vol. 9, No. 2, pp. 305–319, 2000.

- [99] 和田俊和, 野村圭弘, 松山隆司. 分散協調処理による画像の領域分割法. 情報処理学会論文誌, Vol. 36, No. 4, pp. 879–891, 1995.
- [100] T. Matsuyama and T. Wada. Cooperative spatial reasoning for image understanding. *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 11, No. 1, pp. 205–227, 1997.
- [101] J. Liu and Y. Y. Tang. Distributed autonomous agents for chinese document image segmentation. *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 12, No. 1, pp. 97–118, 1998.
- [102] N. Bianchi, P. Bottoni, P. Mussio, C. Spinu, and C. Garbay. Situated image understanding in a multiagent framework. *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 12, No. 5, pp. 595–624, 1998.
- [103] M. Takatsuka, T. M. Caelli, G. A. W. West, and S. Venkatesh. An application of “agent-oriented” techniques to symbolic matching and object recognition. *Pattern Recognition Letters*, Vol. 23, No. 4, pp. 419–429, 2002.
- [104] M. Minsky. *The Society of Mind*. Simon & Schuster, New York, 1986. (安西 (訳): 心の社会, 産業図書 (1990)).
- [105] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, Vol. 1, No. 4, pp. 321–331, 1988.
- [106] S. Edelman and D. Weinshall. A self-organizing multiple-view representation of 3D object. *Biological Cybernetics*, Vol. 64, pp. 209–219, 1991.
- [107] R. Fagin, J. Y. Halpern, Y. Moses, and M. Y. Vardi, editors. *Reasoning about knowledge*. The MIT Press, 1995.
- [108] R. W. Topor. Termination detection for distributed computations. *Information Processing Letters*, Vol. 18, No. 1, pp. 33–36, 1984.
- [109] 富士通研究所. AP1000 使用手引書, 1994.
- [110] P. R. Cooper, L. A. Birnbaum, and M. E. Brand. Causal scene understanding. *Computer Vision and Image Understanding*, Vol. 62, No. 2, pp. 215–231, 1995.
- [111] M. Brand. Physics-based visual understanding. *Computer Vision and Image Understanding*, Vol. 65, No. 2, pp. 192–205, 1997.
- [112] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 5, No. 2, pp. 140–150, 1986.
- [113] D. Kim and R. Nevatia. Recognition and localization of generic objects for indoor navigation using functionality. *Image and Vision Computing*, Vol. 16, No. 11, pp. 729–743, 1998.

- [114] A. Geist, A. Beguelin, J. Dongarra, W. Jiang, R. Manchek, and V. Sunderam. *PVM: Parallel Virtual Machine*. The MIT Press, 1994.
- [115] Z. Li and L. Uhr. Pyramid vision using key features to integrate image-driven bottom-up and model-driven top-down process. *IEEE Transactions on System, Man and Cybernetics*, Vol. 17, No. 2, pp. 250–263, 1987.
- [116] C. L. Tan and W. N. Martin. An analysis of a distributed multi-resolution vision system. *Pattern Recognition*, Vol. 22, No. 3, pp. 257–265, 1989.
- [117] J. Y. Aloimonos. Preface of special issue on purposive, qualitative, active vision. *Computer Vision and Image Understanding*, Vol. 56, No. 1, pp. 1–3, 1992.
- [118] A. L. Ratan and W. E. L. Grimson. Training templates for scene classification using a few examples. In *Proc. of IEEE International Workshop on Content-Based Access of Image and Video Libraries*, pp. 90–97, 1997.
- [119] 小野敦史, 天野督士, 斗谷充宏, 佐藤隆, 坂内正夫. 状態遷移モデルとシーン記述言語による自動キーワード付与機能をもつ画像データベースとその評価. 電子情報通信学会論文誌 D-II, Vol. J79-D-II, No. 4, pp. 476–483, 1996.
- [120] P. Lipson, W. E. L. Grimson, and P. Sinha. Configuration based scene classification and image indexing. In *Proc. of IEEE Computer Vision and Pattern Recognition*, pp. 1007–1013, 1997.
- [121] J. R. Smith and C. S. Li. Image classification and querying using composite region templates. *Computer Vision and Image Understanding*, Vol. 75, No. 1/2, pp. 165–174, 1999.
- [122] O. Maron and A. L. Ratan. Multiple-instance learning for natural scene classification. In *Proc. of 15th International Conference on Machine Learning*, pp. 341–349, 1998.
- [123] A. L. Ratan, O. Maron, W.E.L. Grimson, and T. Lozano-Perez. A framework for learning query concepts in image classification. In *Proc. of IEEE Computer Vision and Pattern Recognition*, pp. 423–429, 1999.
- [124] V. N. Gudivada and V. V. Raghavan. Content-based image retrieval-systems. *IEEE Computer*, Vol. 28, No. 9, pp. 18–22, 1995.
- [125] 串間和彦, 赤間浩樹, 紺谷精一, 山室雅司. 色や形状等の表層的特徴量に基づく画像内容検索記述. 情報処理学会論文誌, Vol. 40, No. SIG3(TOD 1), pp. 171–184, 1999.
- [126] 栗田多喜夫, 加藤俊一, 福田郁美, 板倉あゆみ. 印象語による絵画データベースの検索. 情報処理学会論文誌, Vol. 33, No. 11, pp. 1373–1383, 1992.
- [127] 木本晴夫. 感性語による画像検索とその精度評価. 情報処理学会論文誌, Vol. 40, No. 3, pp. 886–898, 1999.

- [128] T. P. Minka and R. W. Picard. Vision texture for annotation. *ACM/Springer Journal of Multimedia Systems*, Vol. 3, pp. 3–14, 1995.
- [129] 森靖英, 高橋裕信, 保科雅洋, 野崎俊輔, 岡隆一. WWW 上の文書・画像混在データのクロスメディア検索. 第 15 回情報統合研究会資料 SIG-CII-2001-MAR, 2001.
- [130] C. Y. Fung and K. F. Loe. Learning primitive and scene semantics of images for classification and retrieval. In *Proc. of ACM International Conference Multimedia*, pp. 9–12, 1999.
- [131] C. Carson, S. Belongie, H. Greenspan, and J. Malik. Region-based image querying. In *Proc. of IEEE International Workshop on Content-Based Access of Image and Video Libraries*, 1997.
- [132] C. Carson, M. Thomas, S. Belongie, J. M. Hellerstein, and J. Malik. Blobworld: A system for region-based image indexing and retrieval. In *Proc. of Third International Conference on Visual Information Systems*, 1999.
- [133] C. Carson, S. Belongie, H. Greenspan, and J. Malik. Blobworld: Image segmentation using expectation-maximization and its application to image querying. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 8, pp. 1026–1038, 2002.
- [134] K. Barnard, P. Duygulu, and D.A. Forsyth. Clustering art. In *Proc. of IEEE Computer Vision and Pattern Recognition*, pp. II:434–441, 2001.
- [135] P. Duygulu, K. Barnard, J. F. G. de Freitas, and D. A. Forsyth. Object recognition as machine translation: Learning a lexicons for a fixed image vocabulary. In *Proc. of European Conference on Computer Vision*, 2002.
- [136] J. Shi and J. Malik. Normalized cuts and image segmentation. In *Proc. of IEEE Computer Vision and Pattern Recognition*, pp. 731–737, 1997.
- [137] A. B. Benitez and S. F. Chang. Semantic knowledge construction from annotated image collections. In *Proc. of IEEE International Conference on Multimedia and Expo*, 2002.
- [138] A. B. Benitez and S. F. Chang. Perceptual knowledge construction from annotated image collections. In *Proc. of IEEE International Conference on Multimedia and Expo*, 2002.
- [139] C. Fellbaum, editor. *WordNet: An Electronic Lexical Database*. The MIT Press, 2000.
- [140] R. Zhao and W. I. Grosky. From features to semantics: Some preliminary results. In *Proc. of IEEE International Conference on Multimedia and Expo*, 2000.
- [141] R. Zhao and W. I. Grosky. Negotiating the semantic gap: from feature maps to semantic landscapes. *Pattern Recognition*, Vol. 35, No. 3, pp. 593–600, 2002.

- [142] S. C. Deerwester, S. T. Dumais, T. K. Landauer, G. W. Furnas, and R. A. Harshman. Indexing by latent semantic analysis. *Journal of the American Society of Information Science*, Vol. 41, No. 6, pp. 391–407, 1990.
- [143] J. Huang, S. R. Kumar, and R. Zabih. An automatic hierarchical image classification scheme. In *Proc. of ACM International Conference Multimedia*, pp. 219–228, 1998.
- [144] A. Vailaya, A.K. Jain, and H.J. Zhang. On image classification: City images vs. landscapes. *Pattern Recognition*, Vol. 31, No. 12, pp. 1921–1935, 1998.
- [145] A. Vailaya, A.K. Jain, and H.J. Zhang. On image classification: city vs. landscape. In *Proc. of IEEE International Workshop on Content-Based Access of Image and Video Libraries*, pp. 3–8, 1998.
- [146] M. Szummer and R.W. Picard. Indoor-outdoor image classification. In *Proc. of IEEE International Workshop on Content-Based Access of Image and Video Databases*, pp. 42–51, 1998.
- [147] V. Athitsos, M. J. Swain, and C. Framkel. Distinguishing photographs and graphics on the World Wide Web. In *Proc. of IEEE International Workshop on Content-Based Access of Image and Video Libraries*, pp. 10–17, 1997.
- [148] Q. Iqbal and J. K. Aggarwal. Applying perceptual grouping to content-based image retrieval: Building images. In *Proc. of IEEE Computer Vision and Pattern Recognition*, pp. I: 42–48, 1999.
- [149] Q. Iqbal and J. K. Aggarwal. Retrieval by classification of images containing large man-made objects using perceptual grouping. *Pattern Recognition*, Vol. 35, No. 7, pp. 1463–1479, 2002.
- [150] A. Vailaya, M. Figueiredo, A. K. Jain, and H. J. Zhang. Content-based hierarchical classification of vacation images. In *Proc. of IEEE International Conference on Multimedia Computing and Systems*, 1999.
- [151] A. Vailaya, M. A. T. Figueiredo, A. K. Jain, and H. J. Zhang. Image classification for content-based indexing. *IEEE Transactions on Image Processing*, Vol. 10, No. 1, pp. 117–130, January 2001.
- [152] A. Hartmann and R. Lienhart. Automatic classification of images on the web. In *Proc. SPIE 4676, Storage and Retrieval for Media Databases*, pp. 31–40, 2002.
- [153] R. Lienhart and A. Hartmann. Classifying images on the web automatically. *Journal of Electronic Imaging*, Vol. 11, No. 4, pp. 445–454, 2002.
- [154] L. Wenyin, S. T. Dumais, Y. F. Sun, H. J. Zhang, M. P. Czerwinski, and B. Field. Semi-automatic image annotation. In *Proc. of the Eighth Conference on Human Computer Interaction – Interact 2001*, 2001.

- [155] <http://images.google.com/>.
- [156] <http://www.ditto.com/>.
- [157] <http://gallery.yahoo.com/>.
- [158] <http://multimedia.lycos.com/>.
- [159] <http://images.altavista.com/>.
- [160] R. C. Nelson and A. Selinger. Learning 3D recognition models for general objects from unlabeled imagery: An experiment in intelligent brute force. In *Proc. of International Conference on Pattern Recognition*, Vol. I, pp. 1–8, 2000.
- [161] S. Duvdevani-Bar and S. Edelman. Visual recognition and categorization on the basis of similarities to multiple class prototype. *International Journal of Computer Vision*, Vol. 33, No. 3, pp. 201–228, 1999.
- [162] 柳井啓司. WWW からの高速画像収集 と 収集画像を用いた画像認識の試み. 第 15 回人工知能学会全国大会講演論文集, pp. 3E1–05, 2001.
- [163] C. Framkel, M. J. Swain, and V. Athitsos. WebSeer: An image search engine for the World Wide Web. Technical Report TR-96-14, University of Chicago, 1996.
- [164] J. R. Smith and S. F. Chang. Visually searching the Web for content. *IEEE Multimedia*, Vol. 4, No. 3, pp. 12–20, 1997.
- [165] S. Sclaroff, M. LaCascia, S. Sethi, and L. Taycher. Unifying textual and visual cues for content-based image retrieval on the World Wide Web. *Computer Vision and Image Understanding*, Vol. 75, No. 1/2, pp. 86–98, 1999.
- [166] J. R. Smith and S. F. Chang. Visual SEEk: A fully automated content-based image query system. In *Proc. of ACM International Conference on Multimedia 1996*, pp. 87–93, 1996.
- [167] X. S. Zhou and T. S. Huang. Unifying keywords and visual contents in image retrieval. *IEEE Multimedia*, Vol. 9, No. 2, pp. 23–33, 2002.
- [168] X. S. Zhou and T. S. Huang. Unifying keywords and contents in image retrieval: Joint querying, relevance feedback, and pseudoclassification. In *Proc. of International Workshop on Content-Based Multimedia Indexing*, 2001.
- [169] S. Mukherjea, K. Hirata, and Y. Hara. AMORE: A World Wide Web image retrieval engine. *World Wide Web*, Vol. 2, No. 3, pp. 115–132, 1999.
- [170] B. Bianchi-Berthoz and T. Kato. Towards a comprehensive integration of subjective parameters in database browsing. In *Proc. of Advanced Database Systems for Integration of Media and User Environments'98*, Vol. 9, pp. 227–232, 1998.

- [171] N. Inder, B. Bianchi-Berthoze, and T. Kato. K-dime: A software framework for kansei filtering of internet material. In *Proc. of IEEE International Conference on Systems, Man and Cybernetics*, Vol. 6, pp. 241–246, 1999.
- [172] H. T. Shen, B. C. Ooi, and K. L. Tan. Giving meanings to WWW images. In *Proc. of ACM International Conference Multimedia*, 2000.
- [173] R. Lempel and A. Soffer. PicASHOW: pictorial authority search by hyperlinks on the Web. In *Proc. of the Tenth International World Wide Web Conference*, pp. 438–448, 2001.
- [174] R. Lempel and A. Soffer. PicASHOW: pictorial authority search by hyperlinks on the Web. *ACM Transactions on Information Systems*, Vol. 20, No. 1, pp. 1–24, 2002.
- [175] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, Vol. 46, No. 5, pp. 604–632, 1999.
- [176] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. In *Proc. of the Seventh International World Wide Web Conference*, 1998.
- [177] C. C. Yiu and I. C. Yin. Image ranking schemes using link-structure analysis algorithm. In *Proc. of the Eleventh International World Wide Web Conference*, 2002.
- [178] N.C. Rowe and B. Frew. Automatic caption localization for photographs on World-Wide Web pages. *Information Processing and Management*, Vol. 34, No. 1, pp. 95–107, 1998.
- [179] 柳井晴夫, 高木廣文. 多変量解析ハンドブック. 現代数学社, 1986.
- [180] 串間和彦, 佐藤路恵, 赤間浩樹, 山室雅司. 大量画像の閲覧を目的とする階層的分類支援機能 – 画像目録の実装と評価 –. 情報処理学会論文誌, Vol. 41, No. SIG3(TOD 5), pp. 54–63, 2000.
- [181] 高木幹雄, 下田陽久. 画像解析ハンドブック. 東京大学出版会, 1991.
- [182] 小林光夫. 絵画における色彩美の数理的分析の研究. 博士論文, 東京大学大学院工学系研究科, 2000.
- [183] J. Hafner, H. S. Sawhney, W. Equitz, M. Flickner, and W. Niblack. Efficient color histogram indexing for quadratic form distance functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 17, No. 7, pp. 729–736, 1995.
- [184] <http://www.google.co.jp/>.
- [185] <http://www.goo.ne.jp/>.
- [186] <http://www.infoseek.co.jp/>.
- [187] <http://www.lycos.co.jp/>.
- [188] <http://www.ocn.ne.jp/search/>.

- [189] <http://www.excite.co.jp/>.
- [190] N. Otsu and T. Kurita. A new scheme for practical flexible and intelligent vision systems. In *Proc. of IAPR Workshop on Computer Vision*, pp. 431–435, 1988.
- [191] 大津展之, 栗田多喜夫, 関田巖. パターン認識-理論と応用-. 朝倉書店, 1996.
- [192] Y. Deng, B.S. Manjunath, and H.D. Shin. Color image segmentation. In *Proc. of IEEE Computer Vision and Pattern Recognition*, pp. 446–451, 1999.
- [193] Y. Deng and B. S. Manjunath. Unsupervised segmentation of color-texture regions in images and video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 8, pp. 800–810, 2001.
- [194] F. L. Hitchcock. The distribution of a product from several sources to numerous localities. *Journal of Mathematical Physics*, Vol. 20, pp. 224–230, 1941.
- [195] T. Tokuyama and J. Nakano. Efficient algorithms for the hitchcok transportation problem. *SIAM Journal on Computing*, Vol. 24, No. 3, pp. 563–578, 1995.
- [196] Y. Rubner, J. Puzicha, C. Tomasi, and J. M. Buhmann. Empirical evaluation of dissimilarity measures for color and texture. *Computer Vision and Image Understanding*, Vol. 84, No. 1, pp. 25–43, 2001.
- [197] P. F. Felzenszwalb and D. P. Huttenlocher. Image segmentation using local variation. In *Proc. of IEEE Computer Vision and Pattern Recognition*, pp. 98–104, 1998.
- [198] M. Gersho. Asymptotically optimum block quantization. *IEEE Transactions on Information Theory*, Vol. 25, No. 4, pp. 373–380, 1979.
- [199] 谷田川英治. 色と構図に基づく画像検索ブラウザの提案. 修士論文, 東京大学大学院工学系研究科 計数工学専攻, 2002.
- [200] H. Greenspan, G. Dvir, and Y. Rubner. Region correspondence for image matching via emd flow. In *Proc. of IEEE International Workshop on Content-Based Access of Image and Video Libraries*, pp. 27–31, 2000.
- [201] 片山紀生, 佐藤真一. SR-Tree: 高次元点データに対する最近接検索のためのインデックス構造の提案. 電子情報通信学会論文誌 D-I, Vol. J80-D-I, No. 8, pp. 703–717, 1997.
- [202] N. Katayama and S. Satoh. The SR-tree: An index structure for high-dimensional nearest neighbor queries. In *Proc. of the 1997 ACM SIGMOD International Conference on Management of Data*, pp. 369–380, 1997.
- [203] K. S. Goh, E. Chang, and K. T. Cheng. Support vector machine pairwise classifiers with error reduction for image classification. In *Proc. of ACM International Conference Multimedia*, 2001.

- [204] K. S. Goh, E. Chang, and K. T. Cheng. Svm binary classifier ensembles for multi-class image classification. In *Proc. of ACM International Conference on Information and Knowledge Management*, pp. 395–402, 2001.
- [205] O. Chapelle, P. Haffner, and V. Vapnik. SVMs for histogram-based image classification. *IEEE Transactions on Neural Network*, Vol. 10, No. 5, pp. 1055–1064, 1999.
- [206] S. Tong and E. Chang. Support vector machine active learning for image retrieval. In *Proc. of ACM International Conference Multimedia*, 2001.
- [207] P. Bosch, A. van Ballegooij, A. de Vries, and M. Kersten. Exact matching in image database. In *Proc. of IEEE International Conference on Multimedia and Expo*, 2001.

付録 A

50 クラスの分類結果

50 クラスの画像を収集，分類した場合の結果について示す。

50 個のクラスの画像を第 8 章で述べた画像収集システムを用いて，合計 33950 枚 WWW より収集した．50 個のクラスは，表 A.1 に示す 10 個のジャンルについてそれぞれ 5 種類ずつクラスキーワードを選んだ．選んだキーワードは大部分が普通名詞であるが，固有名詞もいくつか含まれている．表 A.2 にそれぞれのクラスの収集枚数を示す．「運動会」「田植え」「富士山」「滝」の 4 クラスの画像がそれぞれ 1000 枚以上と飛び抜けて多く，少ないものでは「新幹線」が 214 枚であった．収集した画像例を図 A.1 に示す．

表 A.1 50 クラスの内訳

ジャンル	クラス名 (キーワード)
自然物：動物	ゾウ, 牛, 犬, ライオン, コアラ
自然物：水性動物	ペンギン, 熱帯魚, クジラ, アザラシ, ハクチョウ
自然物：植物	盆栽, クリスマスツリー, バラ, チューリップ, サボテン
自然物：人物	イチロー, タイガーウッズ, モーニング娘, 獅子舞, 赤ちゃん
人工物：乗物	新幹線, パトカー, レーシングカー, 蒸気機関車, 飛行機
人工物：建造物	観覧車, ベイブリッジ, 東京タワー, 金閣寺, 大仏
人工物：小物	ノートパソコン, 椅子, デジタルカメラ, マウンテンバイク, ヘルメット
人工物：食物	ラーメン, りんご, みかん, すし, フランス料理
風景：行事	入学式, 結婚式, 花見, 運動会, 田植え
風景：名所	富士山, エアーズロック, 夕焼け, 滝, 露天風呂

分類実験では，このうち各クラス最大 600 枚までを用いて，合計 22725 枚の画像を学習画像として利用した．この 22725 枚の中の収集時の評価の高かった各クラス上位 100 枚の画像合計 5000 枚をテスト画像として用いた．評価はテスト画像 1 枚以外をすべて学習画像として，それを 5000 枚分繰り返す leave-one-out 法によって行った．なお，50 クラスの場合は，収集枚数が多過ぎて，第 9 章の実験 2, 実験 5 のように人手で適合画像のみを選びだすことは困難であったので，適合画像のみの実験は行っていない．つまり，ここでの実験は，不適合画像が含まれている収集画像を学習にも評価にも用いていることになる．

分類結果を表 A.2, 図 A.2に示す. F 値の平均は, 方法 2 による結果が最もよく, 40.3 であった. これは, 同様の実験である 10 クラスの場合の実験 1 の結果 34.3 よりよく, 20 クラスの場合の実験 4 の 42.3 を若干下回る程度であった. このことから, 本研究で行った数十クラス程度の分類実験の場合, クラスの数よりも, 分類対象クラスの中でどの程度互いにその画像が視覚的に類似しているクラスがあるかどうかの方が, 分類精度に影響するといえる. つまり, 多様な 50 クラスを分類するよりも, 似たような画像, 例えば動物ばかり 10 クラスを分類する方が難しいと考えられる.

最後に, 方法 2 による分類結果の confusion matrix を表 A.3 に示す. 表の縦方向はテスト画像のクラス名を示し, 横方向は分類先のクラスを示す. この表から, 「滝」に多くの画像が誤分類されていることが分かるが, 誤分類されているのは, 緑色の要素をもっていると考えられるクラスが多い. また, 互いに同じジャンルである「モーニング娘」が「赤ちゃん」に, 「ノートパソコン」が「デジタルカメラ」に誤分類されている.



図 A.1 50 クラスの収集画像の例.

表 A.2 50 クラスの分類結果.

no.	クラス名	収集 枚数	方法 1			方法 2			方法 3			方法 4		
			再現	適合	F 値	再現	適合	F 値	再現	適合	F 値	再現	適合	F 値
1	夕焼け	493	49.0	54.4	51.6	73.0	66.4	69.5	51.0	53.1	52.0	51.0	54.3	52.6
2	盆栽	436	63.0	73.3	67.7	62.0	77.5	68.9	64.0	40.3	49.4	65.0	42.5	51.4
3	エアーズロック	395	41.0	65.1	50.3	53.0	84.1	65.0	46.0	59.7	52.0	39.0	42.9	40.8
4	赤ちゃん	993	79.0	42.0	54.9	94.0	47.2	62.9	85.0	49.4	62.5	83.0	56.1	66.9
5	運動会	1550	43.0	30.1	35.4	57.0	51.4	54.0	59.0	23.5	33.6	45.0	23.8	31.1
6	モーニング娘	470	36.0	40.9	38.3	52.0	54.2	53.1	30.0	50.8	37.7	30.0	43.5	35.5
7	チューリップ	811	28.0	27.7	27.9	47.0	58.0	51.9	19.0	28.4	22.8	18.0	18.6	18.3
8	ラーメン	622	55.0	32.5	40.9	65.0	42.8	51.6	50.0	41.7	45.5	35.0	30.2	32.4
9	田植え	1235	61.0	16.2	25.6	73.0	39.7	51.4	53.0	23.9	32.9	32.0	19.8	24.4
10	金閣寺	461	39.0	43.8	41.3	47.0	51.6	49.2	35.0	37.2	36.1	30.0	27.5	28.7
11	熱帯魚	306	28.0	80.0	41.5	31.0	93.9	46.6	41.0	60.3	48.8	26.0	44.8	32.9
12	パトカー	412	28.0	54.9	37.1	34.0	70.8	45.9	32.0	35.6	33.7	31.0	34.8	32.8
13	バラ	491	29.0	37.7	32.8	39.0	54.9	45.6	27.0	33.3	29.8	16.0	19.0	17.4
14	アザラシ	350	22.0	36.7	27.5	37.0	54.4	44.0	36.0	33.3	34.6	31.0	27.0	28.8
15	椅子	357	19.0	61.3	29.0	35.0	55.6	42.9	35.0	42.7	38.5	30.0	30.0	30.0
16	レーシングカー	348	11.0	61.1	18.6	30.0	71.4	42.3	27.0	29.7	28.3	18.0	27.7	21.8
17	花見	845	67.0	19.6	30.4	80.0	26.0	39.2	55.0	29.9	38.7	43.0	21.2	28.4
18	みかん	639	28.0	28.3	28.1	35.0	40.2	37.4	22.0	27.5	24.4	12.0	12.8	12.4
19	デジタルカメラ	576	37.0	14.6	21.0	62.0	23.1	33.7	44.0	23.7	30.8	33.0	20.0	24.9
20	飛行機	460	8.0	16.3	10.7	29.0	39.2	33.3	18.0	14.6	16.1	20.0	17.1	18.4
21	ライオン 動物	215	7.0	87.5	13.0	19.0	100.0	31.9	21.0	44.7	28.6	13.0	21.7	16.2
22	ハクチョウ	289	8.0	44.4	13.6	21.0	61.8	31.3	13.0	19.7	15.7	15.0	17.4	16.1
23	りんご	609	25.0	18.2	21.1	31.0	31.6	31.3	27.0	28.1	27.6	12.0	9.8	10.8
24	クジラ	693	36.0	19.8	25.5	45.0	22.6	30.1	32.0	26.2	28.8	31.0	27.7	29.2
25	マウンテンバイク	472	12.0	18.8	14.6	25.0	37.3	29.9	17.0	20.2	18.5	12.0	11.2	11.6
26	入学式	646	27.0	16.3	20.3	40.0	23.5	29.6	25.0	23.8	24.4	24.0	21.4	22.6
27	富士山	1247	31.0	14.6	19.8	42.0	21.5	28.5	24.0	22.4	23.2	21.0	20.2	20.6
28	獅子舞	496	27.0	21.3	23.8	25.0	32.1	28.1	23.0	23.0	23.0	7.0	8.3	7.6
29	滝	1245	55.0	17.1	26.1	74.0	17.0	27.7	55.0	29.3	38.2	28.0	16.7	20.9
30	ベイブリッジ	207	3.0	100.0	5.8	17.0	73.9	27.6	18.0	31.0	22.8	11.0	22.4	14.8
31	クリスマスツリー	614	15.0	14.7	14.9	23.0	31.1	26.4	16.0	19.8	17.7	12.0	13.5	12.7
32	コアラ	601	9.0	9.1	9.0	24.0	28.2	25.9	21.0	21.6	21.3	16.0	12.5	14.0
33	蒸気機関車	482	13.0	12.1	12.6	44.0	18.1	25.7	26.0	14.8	18.8	17.0	12.7	14.5
34	すし	701	29.0	14.6	19.5	24.0	25.3	24.6	14.0	13.6	13.8	19.0	19.4	19.2
35	結婚式	346	5.0	50.0	9.1	15.0	55.6	23.6	19.0	27.1	22.4	13.0	20.3	15.9
36	フランス料理	230	11.0	78.6	19.3	13.0	86.7	22.6	15.0	27.8	19.5	14.0	21.5	17.0
37	観覧車	274	2.0	66.7	3.9	12.0	85.7	21.1	19.0	36.5	25.0	13.0	25.0	17.1
38	牛	681	17.0	12.7	14.5	26.0	17.3	20.8	11.0	11.8	11.4	11.0	10.5	10.7
39	ペンギン	347	7.0	30.4	11.4	13.0	46.4	20.3	19.0	23.8	21.1	11.0	13.3	12.0
40	ヘルメット	301	2.0	15.4	3.5	12.0	60.0	20.0	12.0	21.4	15.4	15.0	26.8	19.2
41	ノートパソコン	522	9.0	9.6	9.3	20.0	19.8	19.9	23.0	21.3	22.1	20.0	19.2	19.6
42	サボテン	584	18.0	13.6	15.5	21.0	18.4	19.6	10.0	10.9	10.4	10.0	8.2	9.0
43	東京タワー	300	5.0	71.4	9.3	11.0	73.3	19.1	12.0	16.7	14.0	13.0	20.3	15.9
44	露天風呂	738	22.0	8.4	12.2	28.0	11.2	16.0	12.0	8.0	9.6	8.0	6.1	6.9
45	ゾウ	680	15.0	11.2	12.8	21.0	11.9	15.2	23.0	19.2	20.9	6.0	4.9	5.4
46	大仏	341	1.0	20.0	1.9	9.0	47.4	15.1	2.0	3.2	2.5	10.0	13.5	11.5
47	タイガーウッズ	317	7.0	58.3	12.5	8.0	66.7	14.3	6.0	35.3	10.3	6.0	20.0	9.2
48	犬	255	1.0	100.0	2.0	6.0	75.0	11.1	10.0	16.1	12.3	9.0	11.2	10.0
49	新幹線	214	2.0	100.0	3.9	4.0	100.0	7.7	13.0	28.9	17.9	16.0	27.6	20.3
50	イチロー	221	2.0	100.0	3.9	1.0	50.0	2.0	12.0	44.4	18.9	7.0	38.9	11.9
	平均値	33950	23.9	39.8	29.9	34.2	49.0	40.3	27.6	28.6	28.1	22.2	22.7	22.4

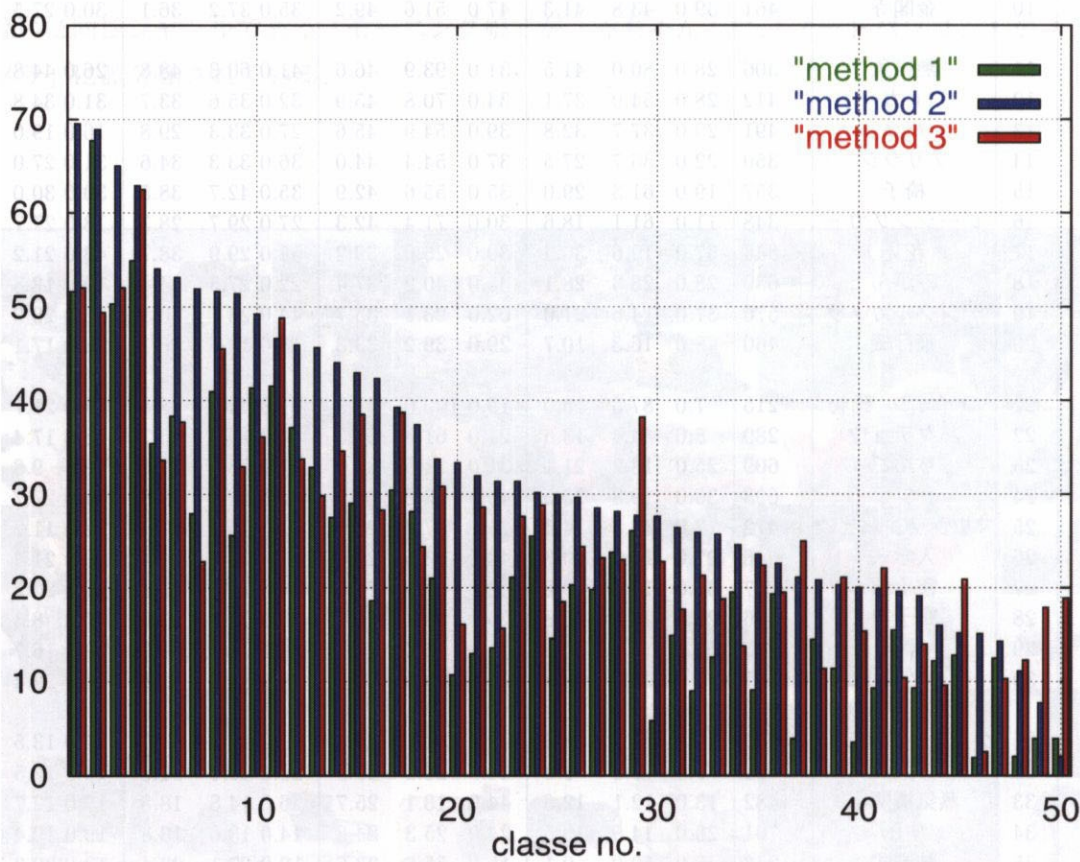


図 A.2 50 クラスの分類結果.

付録 B

画像分類実験に用いた評価用画像

図 B.1～B.20に第 9 章の実験 8 において使用した各クラス 50 枚ずつ合計 20 クラス 1000 枚の評価用画像のすべてを示す。実験 8 において正しく分類された画像を赤い枠で囲っており、誤って分類された画像については誤分類先のクラスを各画像の下に引いた 2 色の下線の色の組合わせで表現している。表 B.1に各クラスに対応する下線の色の組合わせを示す。

図 no.	クラス名	下線の色	
		左	右
B.1	りんご	青	青
B.2	クマ	青	緑
B.3	マウテンバイク	青	紫
B.4	琵琶湖	青	水色
B.5	乗用車	青	黄
B.6	ネコ	緑	青
B.7	入学式	緑	緑
B.8	住宅	緑	紫
B.9	イチロー	緑	水色
B.10	観覧車	緑	黄
B.11	金閣寺	紫	青
B.12	ライオン	紫	緑
B.13	モアイ	紫	紫
B.14	ノートパソコン	紫	水色
B.15	新幹線	紫	黄
B.16	公園	水色	青
B.17	ペンギン	水色	緑
B.18	ラーメン	水色	紫
B.19	結婚式	水色	水色
B.20	槍ヶ岳	水色	黄

表 B.1 各クラスを表す色の組合わせ

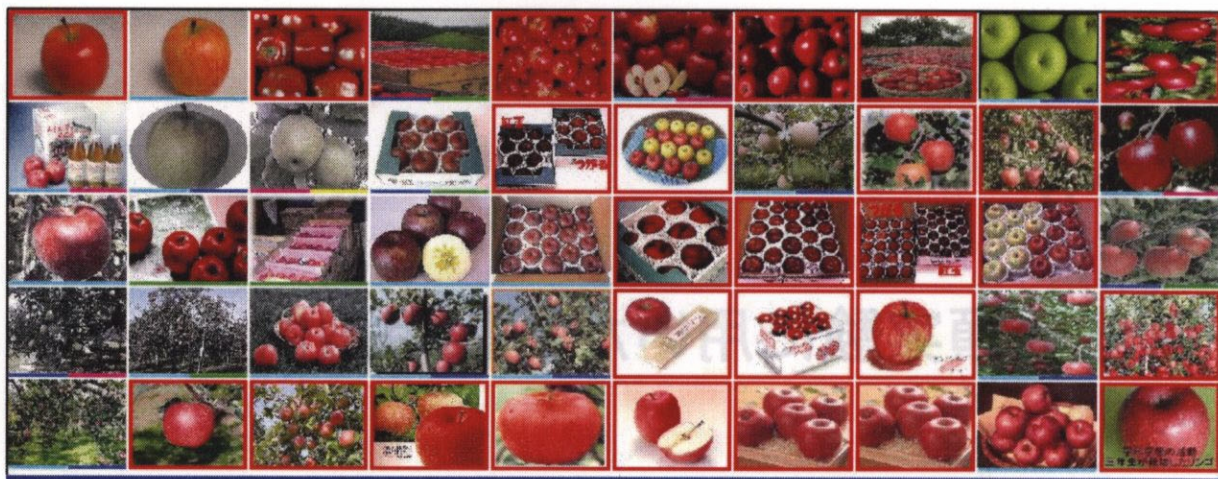


図 B.1 「りんご」の典型画像.



図 B.2 「クマ」の典型画像.

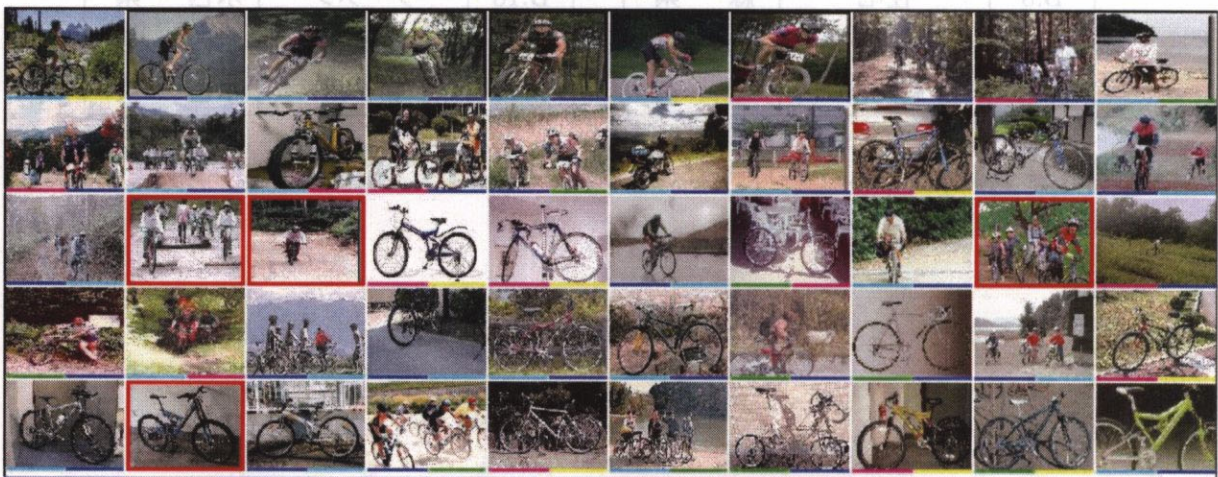


図 B.3 「マウンテンバイク」の典型画像.



図 B.4 「琵琶湖」の典型画像.

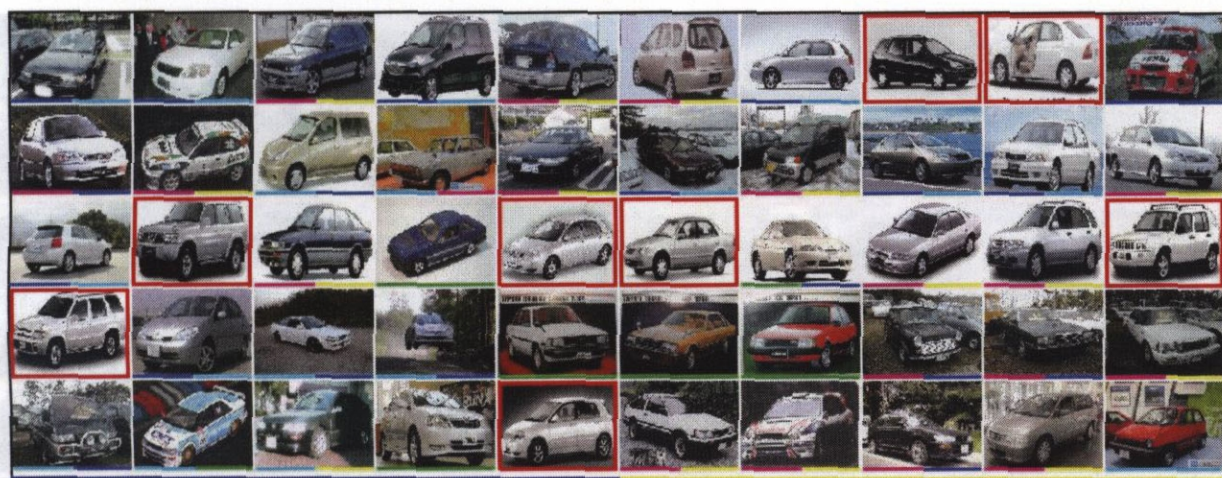


図 B.5 「乗用車」の典型画像.



図 B.6 「ネコ」の典型画像.



図 B.7 「入学式」の典型画像.

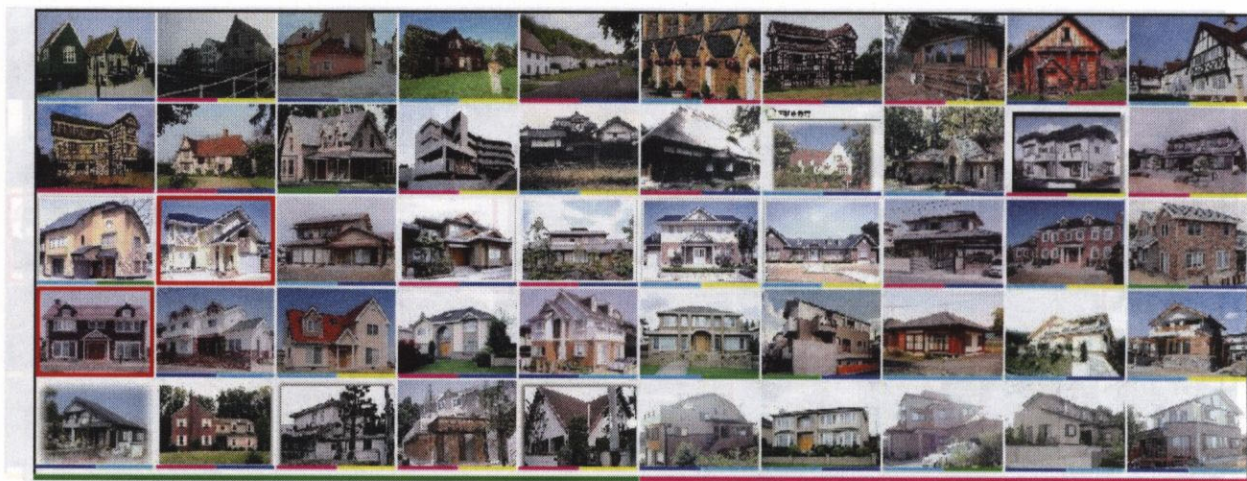


図 B.8 「住宅」の典型画像.

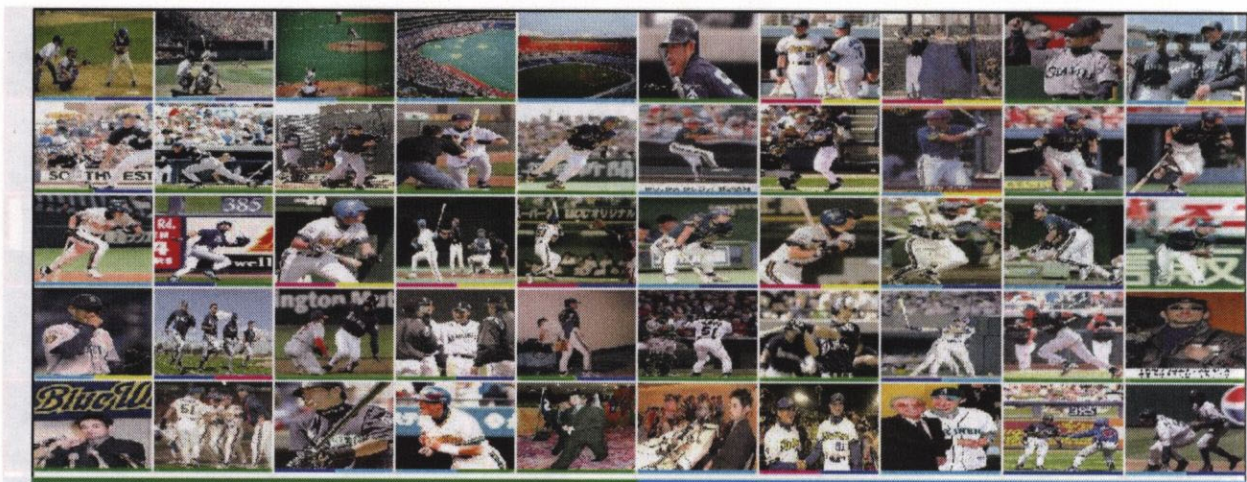


図 B.9 「イチロー」の典型画像.

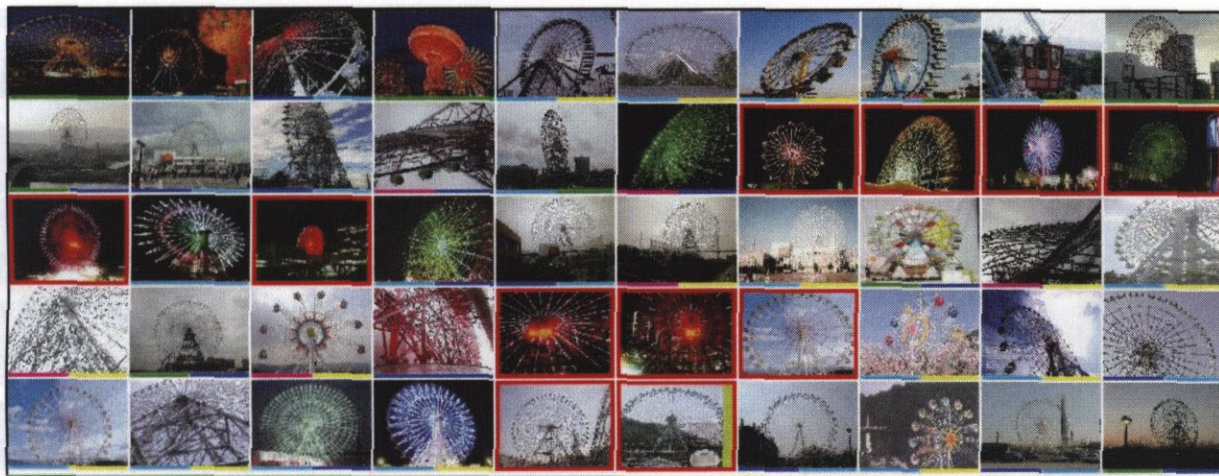


図 B.10 「観覧車」の典型画像.



図 B.11 「金閣寺」の典型画像.



図 B.12 「ライオン」の典型画像.