Reading Group Submission
University of Cambridge

Rob Anderson
Ignas Budvytis
Neill Campbell
Krisada Chaiyasarn
Yu Chen
Paul Mcroy
Fabio Viola
Tsz-Ho Yu

vid. Sect. CVI. *Cube, Sphere, Table,* are Words he has known applied to Things perceivable by Touch, but to Things perfectly Intangible he never knew them apply'd. Those Words, in their wonted application, always mark'd out to his Mind Bodies, or Solid Things which were perceiv'd by the Resistance they gave. But there is no Solidity, no Resistance or Protrusion perceiv'd by Sight. In short, the *Ideas* of Sight are all new Perceptions, to which there be no Names annex'd in his Mind; he cannot, therefore, understand what is said to him concerning them. And to ask, of the two Bodies he saw placed on the Table, which was the Sphere, which the Cube? Were, to him, a Question down right Bantering and Unintelligible: Nothing he sees being able to suggest to his Thoughts, the *Idea* of Body, Distance, or, in general, of any thing he had already known.

'Tis

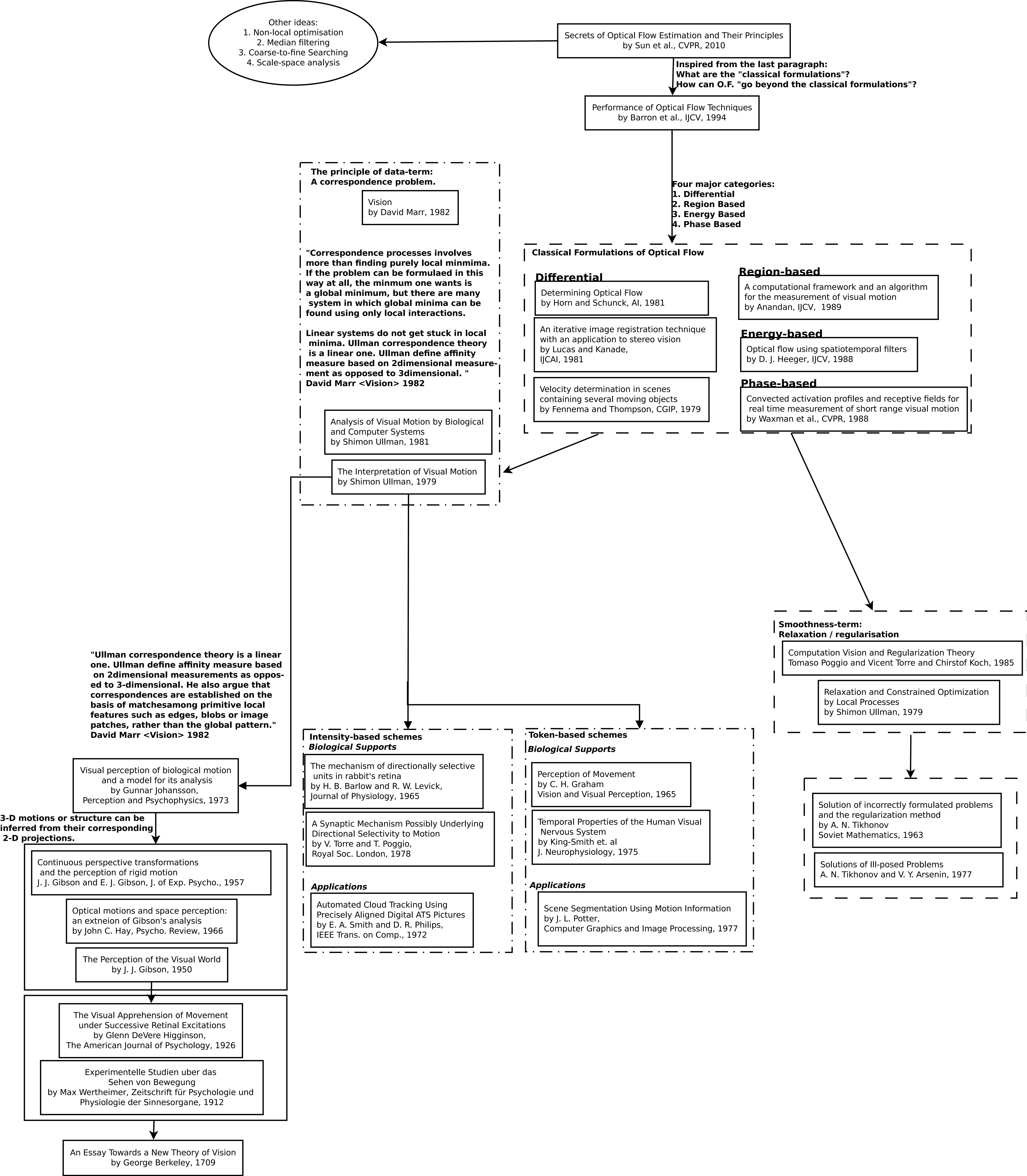
CXXXVI *The same thing doth not affect both Sight and Touch.*

'Tis a Mistake, to think the same thing affects both Sight and Touch. If the same Angle or Square which is the *Object* of Touch, be also the *Object* of Vision: What shou'd hinder the Blind Man, at first Sight, from knowing it? For tho' the manner wherein it affects the Sight, be different from that wherein it affected his Touch; yet, there being, beside this Manner or Circumstance, which is new and unknown, the Angle or Figure which is old and known, he cannot chuse but discern it.

CXXXVII *The same Idea of Motion not common to Sight and Touch.*

Visible Figure and Extension having been demonstrated, to be of a Nature intirely Different and Heterogeneous, from Tangible Figure and Extension, it remains that we inquire concerning *Motion*. Now, that Visible Motion is not of the same Sort with Tangible Motion, seems to need no farther Proof,

it


Other ideas:
1. Non-local optimisation
2. Median filtering
3. Coarse-to-fine Searching
4. Scale-space analysis
Secrets of Optical Flow Estimation and Their Principles by Sun et al., CVPR, 2010
Inspired from the last paragraph: What are the "classical formulations"? How can O.F. "go beyond the classical formulations"?
Performance of Optical Flow Techniques by Barron et al., IJCV, 1994
Four major categories:
1. Differential
2. Region Based
3. Energy Based
4. Phase Based
The principle of data-term: A correspondence problem.
Vision by David Marr, 1982
"Correspondence processes involves more than finding purely local minmima. If the problem can be formulaed in this way at all, the minmum one wants is a global minimum, but there are many system in which global minima can be found using only local interactions.
Linear systems do not get stuck in local minima. Ullman correspondence theory is a linear one. Ullman define affinity measure based on 2dimensional measurement as opposed to 3dimensional. " David Marr <Vision> 1982
Analysis of Visual Motion by Biological and Computer Systems by Shimon Ullman, 1981
The Interpretation of Visual Motion by Shimon Ullman, 1979
Classical Formulations of Optical Flow
Differential
Determining Optical Flow by Horn and Schunck, AI, 1981
An iterative image registration technique with an application to stereo vision by Lucas and Kanade, IJCAI, 1981
Velocity determination in scenes containing several moving objects by Fennema and Thompson, CGIP, 1979
Region-based
A computational framework and an algorithm for the measurement of visual motion by Anandan, IJCV, 1989
Energy-based
Optical flow using spatiotemporal filters by D. J. Heeger, IJCV, 1988
Phase-based
Convected activation profiles and receptive fields for real time measurement of short range visual motion by Waxman et al., CVPR, 1988
Smoothness-term: Relaxation / regularisation
Computation Vision and Regularization Theory Tomaso Poggio and Vicent Torre and Chirstof Koch, 1985
Relaxation and Constrained Optimization by Local Processes by Shimon Ullman, 1979
Solution of incorrectly formulated problems and the regularization method by A. N. Tikhonov Soviet Mathematics, 1963
Solutions of Ill-posed Problems A. N. Tikhonov and V. Y. Arsenin, 1977
"Ullman correspondence theory is a linear one. Ullman define affinity measure based on 2dimensional measurements as opposed to 3-dimensional. He also argue that correspondences are established on the basis of matchesamong primitive local features such as edges, blobs or image patches, rather than the global pattern." David Marr <Vision> 1982
Intensity-based schemes
Biological Supports
The mechanism of directionally selective units in rabbit's retina by H. B. Barlow and R. W. Levick, Journal of Physiology, 1965
A Synaptic Mechanism Possibly Underlying Directional Selectivity to Motion by V. Torre and T. Poggio, Royal Soc. London, 1978
Applications
Automated Cloud Tracking Using Precisely Aligned Digital ATS Pictures by E. A. Smith and D. R. Philips, IEEE Trans. on Comp., 1972
Token-based schemes
Biological Supports
Perception of Movement by C. H. Graham Vision and Visual Perception, 1965
Temporal Properties of the Human Visual Nervous System by King-Smith et. al J. Neurophysiology, 1975
Applications
Scene Segmentation Using Motion Information by J. L. Potter, Computer Graphics and Image Processing, 1977
Visual perception of biological motion and a model for its analysis by Gunnar Johansson, Perception and Psychophysics, 1973
3-D motions or structure can be inferred from their corresponding 2-D projections.
Continuous perspective transformations and the perception of rigid motion J. J. Gibson and E. J. Gibson, J. of Exp. Psycho., 1957
Optical motions and space perception: an extneion of Gibson's analysis by John C. Hay, Psycho. Review, 1966
The Perception of the Visual World by J. J. Gibson, 1950
The Visual Apprehension of Movement under Successive Retinal Excitations by Glenn DeVere Higginson, The American Journal of Psychology, 1926
Experimentelle Studien uber das Sehen von Bewegung by Max Wertheimer, Zeitschrift für Psychologie und Physiologie der Sinnesorgane, 1912
An Essay Towards a New Theory of Vision by George Berkeley, 1709

Secrets of Optical Flow Estimation and Their Principles [21]

We reckon that the main message in [21] is: *there should be no secrets of optical flow.* In the paper, the best available optimization strategies in the field are reviewed in order to find the top performing recipe for the minimization of the typical optical flow objective functions. One of the effects of this study on the future research might be normalizing the massive impact that the optimization strategy has when systems based on only slightly different models are analyzed and compared. In fact, the paper suggests that future research should aim to design brand new models: even if better optimization tricks will become available, they are expected to provide only incremental improvements.

The standard flow objective functional is composed of a data term and a prior term. In the optic flow community it is often the case that the data term component of the objective functional is not questioned. In this exercise we initially traced back in the literature to find out what other models were considered and why the one reported in [12] prevailed.

We started by looking for "review" papers and by re-reading Marr [16], searching for paths that lead to the "classical" formulation for the data term and paths that lead to dead ends.

Barron [3] identifies 4 main categories of optical flow models — Differential [6, 12, 15], Region Based [1], Energy Based [10] and Phase Based [27] — according to the way they approach the correspondence problem. Marr [16] and Ullman [24, 25] discuss the problem of correspondence in images as a 2D process, based on the analogy with biological systems, finding evidence in [2, 23]. Different interpretations can be found, see branches "intensity" [2, 23, 20] and "tokens" [9, 18, 14], and we could trace these ideas back in [13] and even earlier on in [8, 7].

It is our understanding that at this point in time the correspondence problem is tightly linked to the idea of perception - what is visual data in the first place, and how and when we perceive motion? We find seminal work by Gibson [7], Higginson[11] and the original Gestaldtist work of Wertheimer in 1912 [28]. Gibson argued in [7] that although there are similarities in the schemata used by different individuals, each man learns the "meaning of the world" for himself, partly based on his cultural background and individual's unique experience.

In Gibson we also find that it can be argued that the "seeing of space, and therefore the understanding of movement, depends, in some fundamental way, on exploring and manipulating the environment".

Two centuries earlier Bishop Berkeley, fellow at the Trinity College in Dublin, argued in [4]: "Seeing things could always be verified by touching things, and hence it was possible that solidity and depth of visual world were originally not visible but only tangible. Vision might get its spatial character from the tactile and muscular impressions which always accompany it. We learn to trust our vision of the table as being there, for instance, because we can always go over and touch it."

In this interpretation, visual perception and understanding of motion are here backed up by the information gathered via other senses and the interaction with the environment: "The extension, figure, and motions perceived by sight are specifically distinct form the idea of touch called by the same names, nor is there any such thing as one sense or kind of idea common to both senses".

Berkeley says we learn ideas - mental models of the world - by blending information form different senses, so that at times even if some sensory information is not available we can still guess about the properties around us.

> *"I can also pick the ball up, and feel that it is spherical. I can feel a smooth spherical surface. Not only can we see the shapes, sizes and motions of visible things, and feel and shapes, sizes and motion of tangible things, but the deliverance of experience seems to be that there is in general, although not invariably, a spatial coincidence of visible and tangible qualities."*

We believe that the success of richer models at solving low level vision tasks proves this point. An example for flow could be [19], where the flow field is implicitly found in the mapping from a latent image representation of the world to the input frames of a video sequence, rather than being computed explicitly as a frame to frame mapping.

Bibliography

[1] P. Anandan. A computational framework and an algorithm for the measurement of visual motion. *International Jounral of Copmuter Vision*, 2:283 – 310, 1989.

[2] H. B. Barlow and R. W. Levick. The mechanism of directionally selective units in rabbit's retina. *Journal of Physiology*, 1965.

[3] J. L. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of optical flow techniques. *IJCV*, 1994.

[4] George Berkeley. *An Essay Towards a New Theory of Vision.* George Berkeley, 1709.

[5] Michael J. Black and P. Anandan. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. In *Computer Vision and Image Understanding*, 1996.

[6] Claude L. Fennema and William B. Thompson. Velocity determination in scenes containing several moving objects. *Computer Graphics and Image Processing*, 1979.

[7] J. J. Gibson. *The Perception of the Visual World*. Hougton Mifflin, 1950.

[8] J. J. Gibson and E. J. Gibson. Continuous perspective transformations and the perception of rigid motion. *Journal of Experimental Psychology*, 54:129–138, 1957.

[9] C. H. Graham. Perception of movement. *Vision and Visual Perception*, 1965.

[10] David J. Heeger. Optical flow using spatiotemporal filters. *International Jounral of Copmuter Vision*, 1:279 – 302, 1988.

[11] Glenn DeVere Higginson. The visual apprehension of movement under successive retinal excitations. *The American Journal of Psychology*, 1:63 – 115, 1926.

[12] B. K. P. Horn and B. G. Schunck. Secrets of optical flow estimation and their principles. In *Artificial Intelligence*, 1981.

[13] Gunnar Johansson. Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics*, 14:201 – 211, 1973.

[14] P. Ewen King-Smith, L. A. Riggs, R. K. Moore, and T. W. Butler. Temporal properties of the human visual nervous system. *The Jounral of Neurophysiology*, 1975.

[15] Bruce D. Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision. In *IJCAI'81: Proceedings of the 7th international joint conference on Artificial intelligence*, pages 674–679, San Francisco, CA, USA, 1981. Morgan Kaufmann Publishers Inc.

[16] David Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information.* W. H. Freeman, 1982.

[17] Tomaso Poggio, Vicent Torre, and Chirstof Koch. Computation vision and regularisation theory. *Nature*, 1985.

[18] J. L. Potter. Scene segmentation using motion information. *Computer Graphics and Image Processing*, 1977.

[19] Alex Rav-Acha, Pushmeet Kohli, Carsten Rother, and Andrew Fitzgibbon. Unwrap mosaics: A new representation for video editing. *ACM Transactions on Graphics (SIGGRAPH 2008)*, August 2008.

[20] E. A. Smith and D. R. Philips. Automated cloud tracking using precisely aligned digital ats pictures. *IEEE Transactions on Computers*, 1972.

[21] Deqing Sun, Stefan Roth, and Michael J. Black. Secrets of optical flow estimation and their principles. In *CVPR*, 2010.

[22] A. N. Tikhonov. Solution of incorrectly formulated problems and the regularization method. *Soviet Mathematics*, 1963.

[23] V. Torre and T. Poggio. A synaptic mechanism possibly underlying directional selectivity to motion. *Royal Society of London*, 1978.

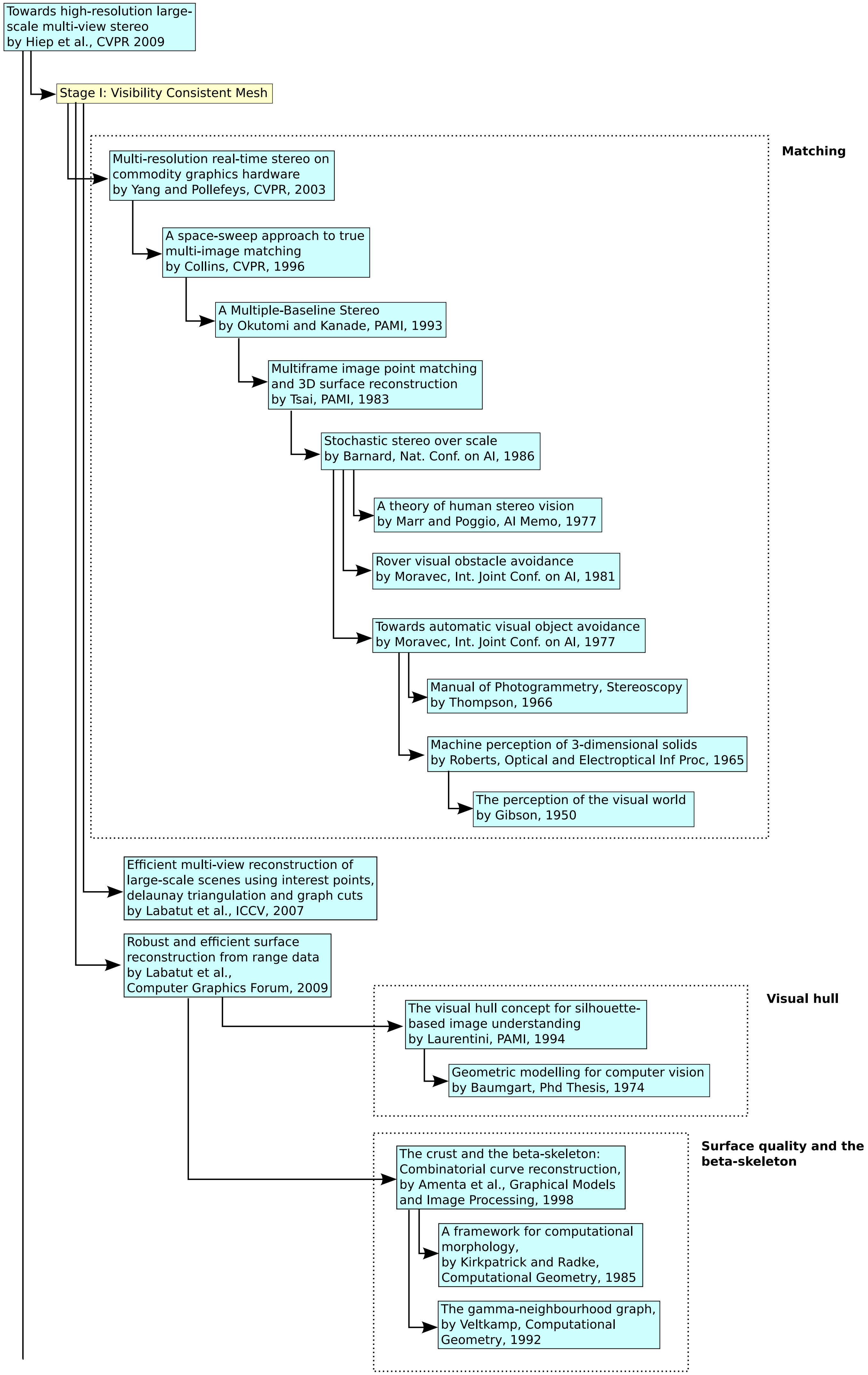
[24] Ullman. Analysis of visual motion by biological and computer systems. *Computer*, 14:57 – 69, 1983.

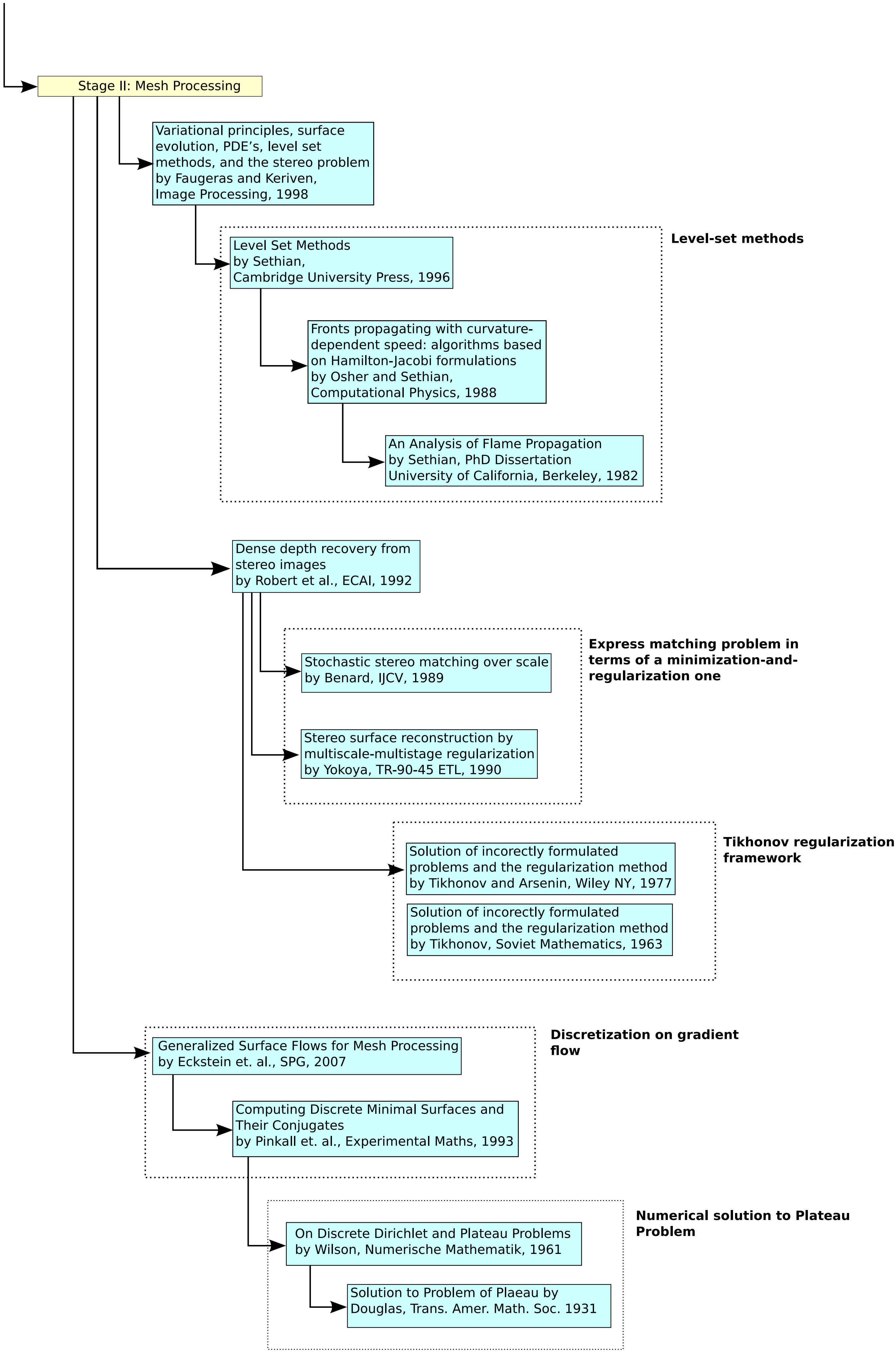
[25] Shimon Ullman. *The Interpretation of Visoin Motion.* MIT Press, 1979.

[26] Shimon Ullman. Relaxation and constrainted optimization. *Computer Graphics and Image Processing*, 1979.

[27] Allen M. Waxman, Jian Wu, and Fredrik Bergholm. Convected activation profiles and the measurement of visual motion. In *Computer Vision and Pattern Recognition*, 1988.

[28] M. Wertheimer. Experimentelle studien uber das sehen von bewegung. *Zeitschrift fr Psychologie und Physiologie der Sinnesorgane*, 61:161 – 265, 1912.


Towards high-resolution large-scale multi-view stereo by Hiep et al., CVPR 2009
Stage I: Visibility Consistent Mesh
Matching
Multi-resolution real-time stereo on commodity graphics hardware by Yang and Pollefeys, CVPR, 2003
A space-sweep approach to true multi-image matching by Collins, CVPR, 1996
A Multiple-Baseline Stereo by Okutomi and Kanade, PAMI, 1993
Multiframe image point matching and 3D surface reconstruction by Tsai, PAMI, 1983
Stochastic stereo over scale by Barnard, Nat. Conf. on AI, 1986
A theory of human stereo vision by Marr and Poggio, AI Memo, 1977
Rover visual obstacle avoidance by Moravec, Int. Joint Conf. on AI, 1981
Towards automatic visual object avoidance by Moravec, Int. Joint Conf. on AI, 1977
Manual of Photogrammetry, Stereoscopy by Thompson, 1966
Machine perception of 3-dimensional solids by Roberts, Optical and Electroptical Inf Proc, 1965
The perception of the visual world by Gibson, 1950
Efficient multi-view reconstruction of large-scale scenes using interest points, delaunay triangulation and graph cuts by Labatut et al., ICCV, 2007
Robust and efficient surface reconstruction from range data by Labatut et al., Computer Graphics Forum, 2009
Visual hull
The visual hull concept for silhouette-based image understanding by Laurentini, PAMI, 1994
Geometric modelling for computer vision by Baumgart, Phd Thesis, 1974
Surface quality and the beta-skeleton
The crust and the beta-skeleton: Combinatorial curve reconstruction, by Amenta et al., Graphical Models and Image Processing, 1998
A framework for computational morphology, by Kirkpatrick and Radke, Computational Geometry, 1985
The gamma-neighbourhood graph, by Veltkamp, Computational Geometry, 1992


Stage II: Mesh Processing
Variational principles, surface evolution, PDE's, level set methods, and the stereo problem by Faugeras and Keriven, Image Processing, 1998
Level-set methods
Level Set Methods by Sethian, Cambridge University Press, 1996
Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations by Osher and Sethian, Computational Physics, 1988
An Analysis of Flame Propagation by Sethian, PhD Dissertation University of California, Berkeley, 1982
Dense depth recovery from stereo images by Robert et al., ECAI, 1992
Express matching problem in terms of a minimization-and-regularization one
Stochastic stereo matching over scale by Benard, IJCV, 1989
Stereo surface reconstruction by multiscale-multistage regularization by Yokoya, TR-90-45 ETL, 1990
Tikhonov regularization framework
Solution of incorectly formulated problems and the regularization method by Tikhonov and Arsenin, Wiley NY, 1977
Solution of incorectly formulated problems and the regularization method by Tikhonov, Soviet Mathematics, 1963
Discretization on gradient flow
Generalized Surface Flows for Mesh Processing by Eckstein et. al., SPG, 2007
Computing Discrete Minimal Surfaces and Their Conjugates by Pinkall et. al., Experimental Maths, 1993
Numerical solution to Plateau Problem
On Discrete Dirichlet and Plateau Problems by Wilson, Numerische Mathematik, 1961
Solution to Problem of Plaeau by Douglas, Trans. Amer. Math. Soc. 1931

Towards high-resolution large-scale multi-view stereo

The root paper [11] proposes a two-stage pipeline that aims to combine the state-of-the-art in both multi-view stereo and mesh processing to provide an accurate reconstruction method that also handles large scenes.

Multi-View Stereo

The authors improve upon their previous paper [14] by using dense stero estimation using the robust NCC metric of [34]. The 'plane-sweep' stereo algorithm derives from the space-sweep method of [6]. The more general origins of window based matching across multiple views can be traced back to [21] where the authors were among the first to use multiple observations to reject outliers, having previously made use of locally adaptive windows [20]. In particular they summed window SSD scores across images as part of their SSSSD approach. Whilst others may have used multiple images they would use them to determine support rather than as a method of directly estimating the surface with the possible exception to [31]. In this work the author presents a 'Window Variance Method' for image matching and 3D object surface reconstruction using multiple perspective views. The work of [2] recognised was among the first to solve the stereo problem over scale using an explict MRF, solved using annealing. Earlier stereo methods operated on pairs and attempted to increase the robustness using coarse to fine approaches to reject outliers, for example [17] where the authors also formalise some of the constraints of the stereo problem. These constraints specifically include an appearance consistency. The use of a window based consistency metric was used by Moravec, in particular a modified NCC [19] for ease of computation, and earlier texture windows [18]. Prior to these works, the task of analysing multiple view geometry for modern computation had been in the photogrammetry community, for example [28] dedicates Chapter XI to Stereoscopy. The review on computational stereo [3] indicates that the work of [25] was one of the first to combine the fields of visual perception, for example [10], with the concept of computational stereo in the field that has become computer vision.

The visual hull concept was introduced by Laurentini [16] based on volume intersection methods first proposed by Baumgart [5].

A surface quality heuristic inspired by the graphics literature is used to filter out improbable tetrahedra. The β-skeleton method of [1] is generalised to 3D. The β-skeleton proposed by [12] is a special case of the more general γ-neighbourhood graphs of [32].

Mesh processing

The second stage of the pipeline photo-refines the noisy mesh provided by the first stage, capturing the fine detail. The authors cite [9] as a pioneering work in variational refinement for multi-view stereovision, but the roots of this idea go much deeper. An original strategy for the solution of the variational problem, the use of level set methods, has its roots in Sethian's PhD thesis [26]. Sethian first applied level set methods to the problem of crystal growth in the mathematics literature and later applied the technique to stereovision in the computer vision field [27, 22]. The alternative approach [24], which is adapted in [9], expresses the variational problem in the classical Tikhonov regularization framework [29, 30] using regularization functions capable of perserving discontinuities. Robert et al. cite [4, 35] as earlier examples of the a minimization-and-regularization approach to stereovision.

The discrete treatment on meshes for *surface flow* was first studied in [23], in which the *problem of plateau* (*i.e.* computing minimal surface) was tackled. In this paper the represention of surfaces based on triangle meshes and subsequently the discretization of *gradient flows* is achieved through the use of discrete differential operators and/or using finite element techqniques [8]. Wilson(1961) [33] published his work that demonstrated solving the problem of plateau using numerical methods, and subsequently his dissertation titled '*A computational Attack on the Problem of Plateau*'. This work was based on reformulation of the original equation based on minimizing the area function to the Direchlet integral, which was first published in Douglas (1931)[7].

Bibliography

[1] N. Amenta, M. Bern, and D. Eppstein. The crust and the beta-skeleton: combinatorial curve reconstruction. *Graphical models and image processing*, 1998.

[2] S. T. Barnard. Stochastic stereo over scale. *Nat. Conf. on AI*, 1986.

[3] S. T. Barnard and M. A. Fischler. Computational stereo. *ACM Comp. Surveys*, 1982.

[4] S.T. Barnard. Stochastic stereo matching over scale. *International Journal of Computer Vision*, 1989.

[5] B.G. Baumgart. *Geometric modeling for computer vision.* PhD Thesis, 1974.

[6] R. Collins. A space-sweep approach to true multi-image matching. *CVPR*, 1996.

[7] J. Douglas. Solution of the problem of Plateau. *Proceedings of the National Academy of Sciences*, 16:242–248, 1930.

[8] I. Eckstein, J.P. Pons, Y. Tong, C.C.J. Kuo, and M. Desbrun. Generalized surface flows for mesh processing. In *Proceedings of the fifth Eurographics symposium on Geometry processing*, page 192. Eurographics Association, 2007.

[9] O.D. Faugeras and R. Keriven. Variational principles, surface evolution, PDEs, level set methods, and the stereo problem. *IEEE Transactions on Image Processing*, 1998.

[10] J. J. Gibson. *The perception of the visual world.* 1950.

[11] V.H. Hiep, R. Keriven, P. Labatut, and J.P. Pons. Towards high-resolution large-scale multi-view stereo. *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.

[12] D.G. Kirkpatrick and J.D. Radke. A framework for computational morphology. *Computational Geometry*, 1985.

[13] P. Labatut, J-P. Pons, and R Keriven. Efficient multi-view reconstruction of large-scale scenes using interest points, delaunay triangulation and graph cuts. *ICCV*, 2007.

[14] P. Labatut, J.P. Pons, and R. Keriven. Efficient multi-view reconstruction of large-scale scenes using interest points, delaunay triangulation and graph cuts. *Proc. International Conference on Computer Vision*, 2007.

[15] P. Labatut, J.P. Pons, and R. Keriven. Robust and efficient surface reconstruction from range data. *Computer Graphics Forum*, 2009.

[16] A. Laurentini. The visual hull concept for silhouette-based image understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1994.

[17] D. Marr and T. Poggio. A theory of human stereo vision. *AI Memo*, 1977.

[18] H. P. Moravec. Towards automatic visual obstacle avoidance. *Int. Joint Conf. on AI*, 1977.

[19] H. P. Moravec. Rover visual obstacle avoidance. *Int. Joint Conf. on AI*, 1981.

[20] M. Okutomi and T. Kanade. A locally adaptive window for signal matching. *ICCV*, 1993.

[21] M. Okutomi and T. Kanade. A multiple-baseline stereo. *PAMI*, 1993.

[22] S. Osher and J.A. Sethian. Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations. *Journal of Computational Physics*, 1988.

[23] U. Pinkall and K. Polthier. Computing discrete minimal surfaces and their conjugates. *Experimental mathematics*, 2(1):15–36, 1993.

[24] L. Robert, R. Deriche, and O.D. Faugeras. Dense depth recovery from stereo images. *Proc. European Conference on Artificial intelligence*, 1992.

[25] L. G. Roberts. Machine perception of 3-dimensional solids. *Optical and Electro-Optical Information Processing*, 1965.

[26] J.A. Sethian. *Analysis of flame propagation*. PhD Thesis, 1982.

[27] J.A. Sethian. *Level set methods: Evolving interfaces in geometry, fluid mechanics, computer vision, and materials science*. Cambridge University Press Cambridge, 1996.

[28] M. M. Thompson. Chapter xi stereoscopy. *Manual of Photogrammetry*, 1966.

[29] A.N. Tikhonov. Solution of incorrectly formulated problems and the regularization method. *Soviet Mathematics*, 1963.

[30] A.N. Tikhonov, V.Y. Arsenin, and F. John. *Solutions of ill-posed problems*. Wiley New York, 1977.

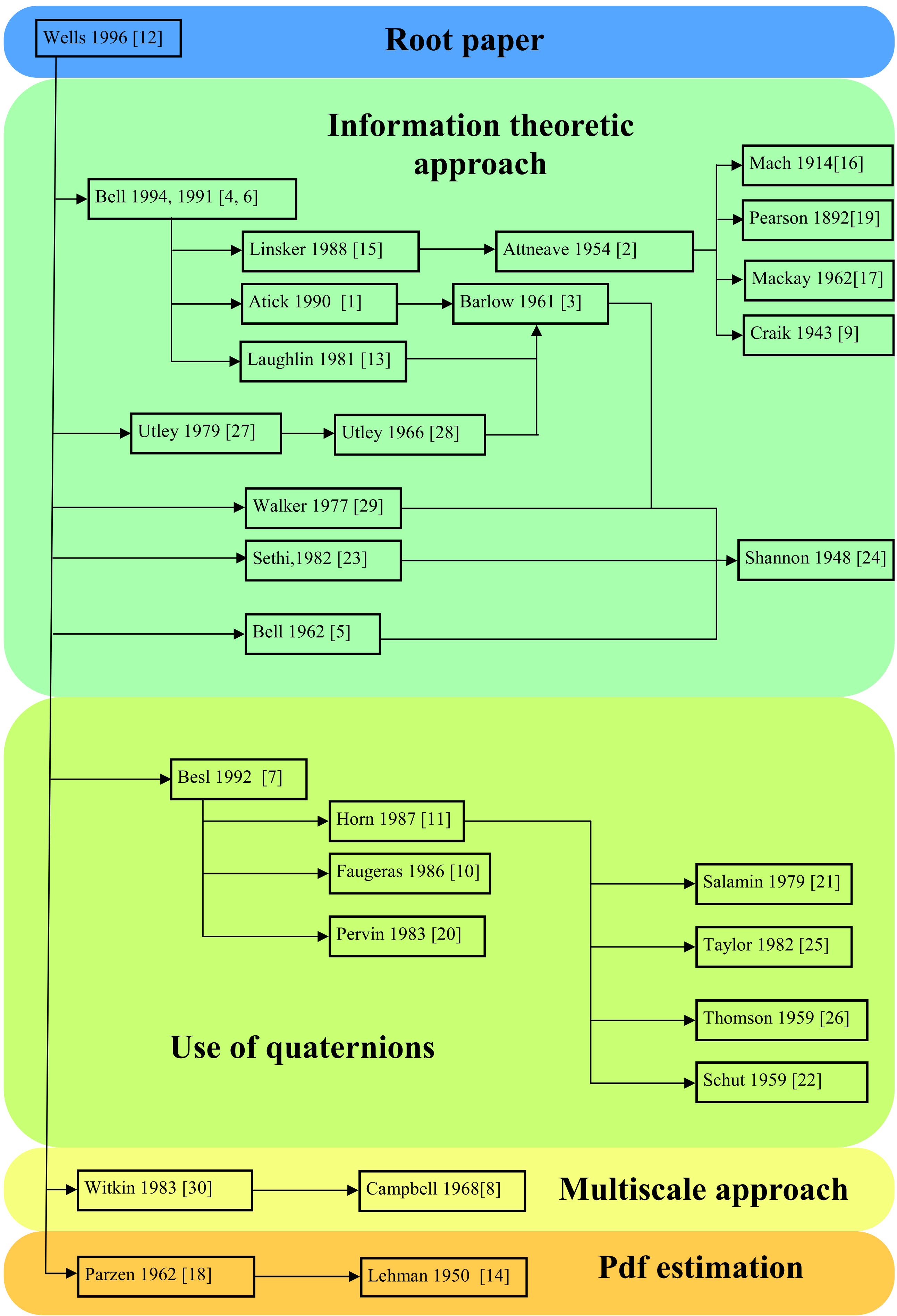
[31] R. Y. Tsai. Multiframe image point matching and 3-D surface reconstruction. *PAMI*, 1983.

[32] R.C. Veltkamp. The gamma-neighborhood graph. *Computational Geometry*, 1992.

[33] W.L. Wilson. On discrete Dirichlet and Plateau problems. *Numerische Mathematik*, 3(1):359–373, 1961.

[34] R. Yang and M. Pollefeys. Multi-resolution real-time stereo on commodity graphics hardware. *CVPR*, 2003.

[35] N. Yokoya. *Stereo surface reconstruction by mutliscale-multistage regularization*. Technical Report TR-90-45, 1990.



Root paper
Wells 1996 [12]
Information theoretic approach
Bell 1994, 1991 [4, 6]
Linsker 1988 [15]
Attneave 1954 [2]
Mach 1914[16]
Pearson 1892[19]
Mackay 1962[17]
Craik 1943 [9]
Atick 1990 [1]
Barlow 1961 [3]
Laughlin 1981 [13]
Utley 1979 [27]
Utley 1966 [28]
Walker 1977 [29]
Sethi,1982 [23]
Shannon 1948 [24]
Bell 1962 [5]
Besl 1992 [7]
Horn 1987 [11]
Faugeras 1986 [10]
Pervin 1983 [20]
Salamin 1979 [21]
Taylor 1982 [25]
Thomson 1959 [26]
Schut 1959 [22]
Use of quaternions
Witkin 1983 [30]
Campbell 1968[8]
Multiscale approach
Parzen 1962 [18]
Lehman 1950 [14]
Pdf estimation

Tree for Multi-modal volume registration by maximization of mutual information

Wells *et. al*'s paper [12] encompasses many ideas in computer vision. The two that we have focused on are first the abstract idea of using information theory as a framework for solving prolems in vision and secondly the application of quaternions for solving the rigid body registration problem. There are many other ideas that can be traced back and we briefly note a couple of these later.

Information theoretic approach to computer vision

The base paper [12] directly cited [6] and [4] which use entropy and information in vision problems. From two papers, we did some back-tracing and found quite a few relevant references (see the following list). These earlier papers tried to understand the visual processing or more general perception from information theory. The relationship among these papers are illustrated in the diagram attached.

- Atick et al. 1990 [1] – The authors of this paper studied the early processing in the mammalian visual pathway. Their model is generalised from Barlow's redundancy reduction hypothesis for perception and it hypothesizes that the goal of visual processing is to recode the sensory data in order to reduce a redundancy measure subject to a constraint that fixes the amount of average information maintained.
- Linsker 1988 [15] – While the base paper aims to maximise mutual information as a means of registration this paper aims to form M cells which preserve maximum information (the "infomax" criteria). Both papers use Shannon's definition of information and link it to the usefull information an image provides about a scene.
- Sethi 1982 [23] – Another paper which approaches a problem in vision from an information theoretic background. This paper addresses the problem of classification. While it is a different problem that this paper aims to solve it still uses the same idea as the base paper that mutual information is an appropriate tool to use in this vision based problem.
- Laughlin 1981 [13]– Laughlin studied the contrast coding in the fly visual system from an information theoretic point of view. Considering a single cell responding to a local contrast, Laughlin derived the optimal transfer function and showed that this prediction compares very well with its experimental measurement of the cell response curve. The theoretical result is that, in order to maximize the amount of information conveyed about the input signal, the cell should perform what is known in image processing as "sampling equalization".
- Walker 1977 [29] – Walker investigated the role of information theory in the perception problem. He proposed that mutual information between the

perceived object and the perceptual indicator can be used as an important criterion for claiming that some system perceives a certain object or not.

- Uttley 1966 [28] – Uttley examined the neural network for pattern recognition in which the contribution of an input is made proportional to the Shannon information between that input and the output. He named such a network "infomon" in his later papers or books [27]. He found that "infomon" framework overcomes limitations of the linear separation network. Experimental results also matched properties of neurons in the cerebral cortex and offered explanations of some known perceptual phenomena.

- Barlow 1961 [3] – Barlow proposed the efficient coding hypothesis for the general perception and sensory problem and used Shannon's information theory for the analysis purpose. In his model, the sensory pathway is treated as a communication channel. On the other hand, the neuronal spiking is an efficient code for representing sensory signals, i.e., it maximizes available channel capacity by minimizing the redundancy between representational units.

- Bell 1962 [5] – This paper is very closely related to the base paper in that it also views mutual information as an alternative to correlation for measuring the dependence of two signals. While it does not directly relate to visual perception there is a strong parallel between the more traditional measure of correlation, widely used in vision for matching, and the alternative method of mututal information which the base paper suggests as an alternative.

- Attneave 1954 [2] – This is an even earlier attempt to use information theory in visual perception. It notes that much of the visual information contained in an image is redundant which has long been understood in vision (reducing an image to an edge image and still being able to understand the scene for example) and it is this redundancy in the information provided by the two different sources of scans that allows the base paper to align them. Similar ideas that reduction of redundancy is important in the handling of sensory information were also set out in two other papers [9],[17], and even earlier books [16], [19] from a psychological point of view: concepts and hypotheses should be "economical representations of the actual" and bring order and simplicity to our complex sensory experiences in order to achieve "economy of thought".

- Shannon 1948 [24] – We would argue that this is going beyond the roots of the idea from a perspective of vision. While it is indeed true that without Shannon's underlying work on information theory the base paper's idea would not be possible, at this point it is more the roots of the analytical tool than the roots of the idea and if that path is followed then we could trace all of the way back to the roots of basic algebra. It is important to decide where to stop the tree and we believe that this is a logical termination.

Quaternions in registration

Estimating three-dimensional object position and orientation parameters from image data is an important aspect of many computer vision problems and one that the base paper [12] attacks using quaternions. The pioneering use of quaternions for representing the 3D rotations when registering 3D rigid objects is usually accredited to [7], which perhaps is the mostly cited paper in the area. [11] gave the solution to the problem of recovering the transformation between two different Cartesian coordinate systems from the coordinates of a number of points as measured in two systems. Such a solution is also provided in two other slightly earlier papers [20, 10] in the realms of computer vision. [20] is probably the first paper which clearly addressed the applications of quaternions in computer vision. However, the groundwork for the application of quaternions actually appear in even earlier literatures of robotics [21, 25] and photogrammetry [22, 26]. The last two paper can be regarded as the earliest papers which use quaternions to solve vision problems.

Other minor ideas

Virtually any idea or technique within the base paper can be traced back thorugh many years or research. The two minor branches on the tree give two examples of this. Without the idea that a probability density function can be estimated by sampling the approach proposed would not be possible. The ideas behind this can be traced directly back to Parzen 1962 [18] and even beyond to Lehmann 1950 [14] and from there it could even be argued back to the work of Gauss in the late eighteen hundreds. Similarly the idea of looking at an image over multipler scales, in the form of a Gaussian pyramid for example, is now a common one and is used in the base paper. This technique can be traced back to Witkin 1983 [30] and the importance of scale in vision even further back, for example to Campbell and Robson 1968 [8].

Bibliography

[1] J.J. Atick and A.N. Redlich. Towards a theory of early visual processing. *Neural Computation*, 2:308–320, 1990.

[2] F. Attneave. Some informational aspects of visual perception. *Psychological Review*, 61(3):183–193, May 1954.

[3] H. Barlow. Possible principles underlying the transformation of sensory messages. *Sensory communication*, pages 217–234, 1961.

[4] A. Bell and T.J. Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Advanced in Neural Information Processing*, 7, 1994.

[5] C. Bell. Mutual information and maximal correlation as measures of dependence. *The Annals of Mathematical Statistics*, 47:30–43, 1962.

[6] S. Bell and G.E. Hinton. Learning to make coherent predictions in domains with discontinuities. *Advanced in Neural Information Processing*, 4, 1991.

[7] P.J. Besl and N.D. Mckey. A method for registration of 3-D shapes. *IEEE Trans. Pattern Analysis and Machine Intelligence (PAMI)*, 14(2):239–256, 1992.

[8] F.W. Campbell and J.G. Robson. Application of fourier analysis to the visibility of gratings. *The Journal of Physiology*, pages 551–556, 1968.

[9] K.J.W. Craik. *The nature of explanation.* Cambridge University Press, 1943.

[10] O.D. Faugeras and M. Hebert. The representation, recognition, and locating of 3-D objects. *Internal Journal of Robotics Research*, 5(3):27–52, 1986.

[11] B.K.P. Horn. Closed-form solution of absolute orientation using unit quaternions. *Journal of the Optical Society of America*, 4(4):629–642, 1987.

[12] William M. Wells III, Paul Viola, Hideki Atsumi, Shin Nakajima, and Ron Kikinis. Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis*, 1(1):35 – 51, 1996.

[13] S. Laughlin. A simple coding procedure enhances a neuron's information. *Z. Naturforsh*, 3:910–912, 1981.

[14] E. Lehmann. Notes on the theory of estimation. 1950.

[15] R. Linsker. Self-organization in a perceptual network. *Computer*, 21:105–117, 1988.

[16] E. Mach. *The Analysis of Sensations*. Translated by C.M. Williams from 1st Germen ed. 1886, 1914.

[17] D.M. Mackay. Towards an information-flow model of human behaviour. *British Journal of Psychology*, 33:587–595, 1962.

[18] Emanuel Parzen. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076, September 1962.

[19] K. Pearson. *The Grammar of Science*. London, Walter Scott, 1892.

[20] E. Pervin and J.A. Webb. Quaternions in computer vision and robotics. *Int. Conf. on Computer Vision and Pattern Recognition*, pages 382–383, 1983.

[21] E. Salamin. Application of quaternions to computation with rotations. *Internal Rep. Stanford University*, 1979.

[22] G.H. Schut. Construction of orthogonal matrices and their application in analytical photogrammetry. *Photogrammetria*, 15:149–162, 1959.

[23] I.K. Sethi and G.P.R. Sarvarayudu. A mathematical theory of communication. *IEEE TRANS. PATTERN ANALY. AND MACH. INTELLIG.*, 4:441–445, 1982.

[24] C.E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, July 1948.

[25] R.H. Taylor. Planning and execution of straight line manipulator trajectories. *Robot Motion: Planning and Control*, 1982.

[26] E.H. Thompson. On exact linear solution of the problem of absolute orientation. *Photogrammetria*, 15:163–179, 1959.

[27] A.M. Uttley. *Information Transmission in the Nervous System*. Academic Press, London, 1979.

[28] H.M. Uttley. The transmission of information and the effect of local feedback in theretical and neural networks. *Brain Research*, 2:21–50, 1966.

[29] W.R. Walker. Criteria for object perception. *journal of cybernetics*, 7:169–187, 1977.

[30] A.P. Witkin. Scale-space filtering. *International Joint Conference on Artificial Intelligence*, pages 1019–1022, 1983.