W. Förstner, J.M. Buhmann, A. Faber, P. Faber (Eds.),
Mustererkennung 1999, Springer, Berlin, pp. 317–324, 1999.

Räumlich–zeitliche Berechnung des optischen Flusses mit nichtlinearen flußabhängigen Glattheitstermen

Joachim Weickert und Christoph Schnörr

LS Bildverarbeitung, Mustererkennung und Computergrafik
Fakultät für Mathematik und Informatik
Universität Mannheim
D-68131 Mannheim
{Joachim.Weickert,Christoph.Schnoerr}@ti.uni-mannheim.de

Zusammenfassung Nichtquadratische Variationsansätze sind die mathematische Grundlage leistungsfähiger Algorithmen zur Berechnung stückweise glatter optischer Flußfelder in digitalen Bildfolgen. In dem vorliegenden Artikel erweitern wir diese Ansätze zu *nichtlinearen* und *flußabhängigen* Funktionalen, die über einem *räumlich–zeitlichen* Intervall definiert sind. Um den Effekt dieser Erweiterung studieren zu können, basiert unsere Implementierung ausschließlich auf dem entsprechenden System von Diffusions–Reaktions–Gleichungen. Jede Form impliziter Glättung, sei es durch einen Vorverarbeitungsschritt zur Rauschunterdrückung oder bei der Schätzung partieller Ableitungen der Bildfunktion, wurde vermieden. Numerische Ergebnisse für Realweltbildfolgen zeigen, daß unser räumlich–zeitlicher Ansatz (i) den berechneten optischen Fluß signifikant verbessert, (ii) Hintergrundrauschen stark unterdrückt und (iii) Bewegungsgrenzen deutlicher hervortreten läßt. Der zusätzliche Berechnungsaufwand macht weniger als einen Faktor 2 gegenüber einem rein räumlichen Ansatz aus.

Schlagworte: Optischer Fluß, Regularisierung, nichtlineare Diffusion, adaptive räumlich–zeitliche Filterung.

1 Einführung

Methoden zur robusten Bewegungsanalyse digitaler Bildfolgen sind ein zentrales Forschungsthema des Maschinensehens. Ein wichtiges Konzept der Bildfolgenanalyse ist die momentane Bewegungsgeschwindigkeit von Bildstrukturen in der Bildebene, der sogenannte *optische Fluß*. Man vergleiche etwa [10] für einen

Überblick über den aktuellen Stand der Forschung sowie [1] für eine Evaluierung bekannter Algorithmen zur Berechnung des optischen Flusses.

Ein wichtiger Bestandteil von Verfahren zur Berechnung des optischen Flusses sind Glättungsmechanismen, die sowohl in impliziter als auch in expliziter Form auftreten. Implizite Glättung tritt beispielsweise durch die Verwendung von FIR–Tiefpaßfiltern zur Rauschunterdrückung und zur Schätzung der partiellen Ableitungen der Bildfunktion auf. Obwohl die exakte Spezifikation dieser Glättung erheblich Auswirkungen auf die Leistungsfähigkeit eines Verfahrens haben kann, findet man diese nur selten in Veröffentlichungen angegeben.

Explizite Glättungsmechanismen treten in Form sogenannter Glattheitsterme in Variationsansätzen zur Berechnung des optischen Flusses auf, wie sie erstmals von Horn und Schunck [7] vorgeschlagen und von Nagel und Enkelmann [13] und vielen anderen erweitert worden sind. Bei diesen Ansätzen wird der optische Fluß als minimierendes Vektorfeld eines Energiefunktionals berechnet, welches neben einem Glattheitsterm einen sogenannten Datenterm umfaßt. Der Datenterm beinhaltet eine oder mehrere Erhaltungsgleichungen, welche die (postulierte) Invarianz der Grauwertfunktion oder davon abgeleiteter Größen unter einer Bewegung beschreiben.

Der Einfluß des Glattheitstermes kann *datenabhängig* an Grauwertkanten reduziert werden, um eine Glättung des optischen Flußfeldes über solche Kanten hinweg zu verhindern [13]. In der jüngeren Literatur wurden auch *flußabhängige* Glattheitsterme vorgestellt [2,3,4,9,16,19], um die Glättung des optischen Flußfeldes über Bewegungsgrenzen hinweg zu vermeiden. Mit diesen Ansätzen konnten gute Resultate erzielt werden, obwohl zumeist nur die stückweise *räumliche* Glattheit von optischen Flußfeldern modelliert wurde.

In dem vorliegenden Artikel schlagen wir die Erweiterung konvexer, aber nichtquadratischer flußabhängiger Glattheitsterme auf die *Regularisierung des optischen Flusses über den Raum und die Zeit* vor. Das Ausnutzen der zeitlichen stückweisen Glattheit von Flußfeldern ist ebenso natürlich wie die bekannte Ausnutzung der räumlichen Glattheit und führt zu Verarbeitungsprozessen, die prinzipiell kaum komplexer sind als bei einer rein räumlichen Verarbeitung. Unsere numerischen Ergebnisse zeigen zudem, daß die Robustheit der optischen Flußschätzung durch diese Maßnahme signifikant erhöht wird. Um eine eindeutige Zuordnung der Effekte zu ermöglichen, haben wir dabei auf jegliche Form impliziter Glättung (vgl. die Diskussion oben) verzichtet. Unser Ansatz reiht sich damit in eine wachsende Zahl jüngerer Arbeiten ein, die einen größeren Zeitausschnitt zur Bildfolgenanalyse heranziehen (siehe z.B. [8] und dortige Verweise).

Zu den wenigen Arbeiten, in denen Glattheitsterme über einem räumlich–zeitlichen Bildfolgenausschnitt betrachtet werden, zählt eine Arbeit von Nagel [12]. Der dort formulierte Ansatz ist jedoch datenabhängig, und numerische Auswertungsergebnisse wurden nicht präsentiert. Gegenüber den von Murray und Buxton [11] und Black und Anandan [2] vorgestellten Ansätzen unterscheidet sich unser Verfahren durch eine rotationsinvariante kontinuierliche Formulierung und die Verwendung konvexer Regularisierungsansätze. Sie erlauben neben einer guten theoretischen Fundierung eine relativ effiziente und global konvergente numerische Approximation.

Im nächsten Abschnitt geben wir eine kurze Übersicht über Variationsansätze zur Berechnung des optischen Flusses mit räumlichen Glattheitstermen. In Abschnitt 3 stellen wir unseren neuen Ansatz mit einem flußabhängigen Glattheitsterm vor, der über einem räumlich–zeitlichen Intervall definiert ist. Die Anwendung des Ansatzes auf zwei Realweltbildfolgen wird in Abschnitt 4 diskutiert.

2 Räumliche Glattheitsterme

Wir bezeichnen eine Bildfolge mit der Funktion $f(x, y, z)$, wobei (x, y) einen Punkt im Bildbereich Ω und $z \in [0, Z]$ die Zeit angibt. Unsere Variationsansätze zur Berechnung eines optischen Flußfeldes $(u, v)^T$ basieren auf zwei grundlegenden Annahmen:

1. Die Bildintensitäten sind bewegungsinvariant. Dies führt zur Bedingung [7]

$$f_x u + f_y v + f_z = 0, \tag{1}$$

   wobei tiefgestellte Indizes partielle Ableitungen bezeichnen. Zahlreiche Verallgemeinerungen zu mehrfachen Gleichungen und/oder anderen “Erhaltungsgrößen” existieren (z.B. [5,17]).
2. Das optische Flußfeld ist stückweise glatt über dem Bildausschnitt Ω. Dies wird erreicht durch Minimierung eines nichtquadratischen ***flußabhängigen Glattheitsterms*** vom Typ

$$\int_{\Omega} \Psi\left(|\nabla u|^2 + |\nabla v|^2\right)\, dx\, dy \tag{2}$$

   wobei $\Psi : \mathbb{R}^+ \to \mathbb{R}$ eine wachsende, differenzierbare Funktion ist und $\nabla := (\partial_x, \partial_y)^T$ den 2D Nabla–Operator bezeichnet. Solche Terme mit abfallender Diffusivität $\Psi'(s^2)$ wurden von Schnörr [16] und Weickert [19] untersucht. In [19] beispielsweise wurde

$$\Psi(s^2) = \lambda^2 \sqrt{1 + s^2/\lambda^2} \qquad (\lambda > 0) \tag{3}$$

   benutzt, was auf die Diffusivität $\Psi'(s^2) = (1 + s^2/\lambda^2)^{-1/2}$ führt. λ fungiert hier als ein “Kontrastparameter”, der die Sensivität des Verfahrens gegenüber Bewegungsgrenzen steuert.

Einen Kompromiß zwischen den beiden obigen Anforderungen erhält man durch Minimierung des Funktionals

$$E(u, v) := \int_{\Omega} \Big((f_x u + f_y v + f_z)^2 \; + \; \alpha \Psi(|\nabla u|^2 + |\nabla v|^2) \Big)\, dx\, dy. \tag{4}$$

Das (bei konvexem Ψ eindeutige) minimierende Flußfeld $(u,v)^T$ ergibt sich für $t \to \infty$ als stationäre Lösung des nichtlinearen Diffusions–Reaktions–Systems:

$$u_t = \nabla \cdot \left(\Psi'\left(|\nabla u|^2 + |\nabla v|^2\right) \nabla u\right) - \tfrac{1}{\alpha} f_x (f_x u + f_y v + f_z), \tag{5}$$
$$v_t = \nabla \cdot \left(\Psi'\left(|\nabla u|^2 + |\nabla v|^2\right) \nabla v\right) - \tfrac{1}{\alpha} f_y (f_x u + f_y v + f_z), \tag{6}$$

wobei t einen numerischen Evolutionsparameter darstellt, der nicht mit der Zeit z der Bildfolge verwechselt werden sollte.

3 Räumlich–zeitliche Glattheitsterme

In diesem Abschnitt erweitern wir die Definition der im letzten Abschnitt diskutierten Glattheitsterme auf ein räumlich–zeitliches Gebiet. Anstatt den optischen Fluß (u,v) als Minimum eines zweidimensionalen Funktionals (4) für jedes einzelne Bild zu einem Zeitpunkt $z \in [0,Z]$ zu berechnen, minimieren wir nun ein einzelnes dreidimensionales Funktional und erhalten den optischen Fluß für *alle* Bilder eines Zeitintervalls $[0,Z]$:

$$E(u,v) := \int_{\Omega \times [0,Z]} \Big((f_x u + f_y v + f_z)^2 + \alpha \, \Psi(|\nabla_3 u|^2 + |\nabla_3 v|^2) \Big) \, dx \, dy \, dz \tag{7}$$

wobei $\nabla_3 := (\partial_x, \partial_y, \partial_z)^T$ den räumlich–zeitlichen Nabla–Operator bezeichnet. Analog zu Abschnitt 2 genügt das minimierende Flußfeld dem Diffusions–Reaktions–Gleichungssystem

$$u_t = \nabla_3 \cdot \left(\Psi'\left(|\nabla_3 u|^2 + |\nabla_3 v|^2\right) \nabla_3 u\right) - \tfrac{1}{\alpha} f_x (f_x u + f_y v + f_z), \tag{8}$$
$$v_t = \nabla_3 \cdot \left(\Psi'\left(|\nabla_3 u|^2 + |\nabla_3 v|^2\right) \nabla_3 v\right) - \tfrac{1}{\alpha} f_y (f_x u + f_y v + f_z) \tag{9}$$

für $t \to \infty$. In dem vorliegenden Artikel studieren wir diesen Prozeß für das nichtlineare Potential (3).

Der Diffusionsanteil in (8)–(9) ist eng mit nichtlinearen Diffusionsfiltern zur Regularisierung dreidimensionaler vektorwertiger Bilder verwandt. Solche Filter wurden erstmals von Gerig *et al.* [6] im Bereich der medizinischen Bildverarbeitung angewendet, allerdings unter Verwendung von Diffusivitäten aus [14], die instabiles Glättungsverhalten zur Folge haben können. Wir vermeiden solche Schwierigkeiten in unserem Ansatz durch die Verwendung konvexer Glattheitsterme (vgl. [15]). Für eine ausführliche Diskussion nichtlinearer Diffusionsfilter verweisen wir auf [18].

4 Numerische Aspekte und Ergebnisse

Wir approximieren das 2D Diffusions–Reaktions–System (5)–(6) bzw. dessen 3D–Erweiterung (8)–(9) mit finiten Differenzen. Partielle Ableitungen in x, y und

z werden mit zentralen Differenzen berechnet, und für die Diskretisierung in t–Richtung verwenden wir ein modifiziertes explizites Schema. Wir betonen nochmals, daß keinerlei implizite (Vor–) Glättung oder irgendeine Nachverarbeitung vorgenommen wurde.

Die numerische Schrittweite in t-Richtung betrug $\tau = 1/4$ im 2D Fall und $\tau = 1/6$ im 3D Fall. Die Iteration wurde abgebrochen, sobald die euklidische Norm des relativen Residuums unter 0.001 gefallen war.

Bemerkenswert ist, daß der Berechnungsaufwand des 3D Ansatzes nicht einmal doppelt so hoch ist wie die separate Anwendung des 2D Ansatzes auf alle Bilder des Zeitausschnittes. Der Speicheraufwand dagegen ist natürlich deutlich höher, aber bei der heutigen Rechnergeneration eigentlich kein relevantes Kriterium mehr.

In diesem Abschnitt illustrieren wir das Verhalten unseres Ansatzes anhand zweier Realweltbildfolgen (Abbildungen 1 und 3). Die zweite Bildfolge ist erhältlich unter `ftp://csd.uwo.ca` im Verzeichnis `pub/vision`.

Wir benutzten 16 aufeinanderfolgende Bilder bei der ersten Bildfolge, 20 Bilder bei der zweiten Bildfolge, und verglichen die rein räumliche 2D–Verarbeitung (Gleichungen (5)–(6)) mit der räumlich–zeitlichen 3D–Verarbeitung (Gleichungen (8)–(9)), wobei in jedem Fall die gleichen Parameterwerte benutzt wurden. Aus Gründen der Übersichtlichkeit wurden in den folgenden Abbildungen Vektoren $w = (u, v)^T$ mit $|w| < 0.2$ Pixel nicht wiedergegeben.

Eine Beschreibung und Kommentierung der Ergebnisse findet der Leser bei den jeweiligen Abbildungen.

Abbildung1. Eine Person bewegt sich auf die Kamera zu (rechts: Ausschnitt).

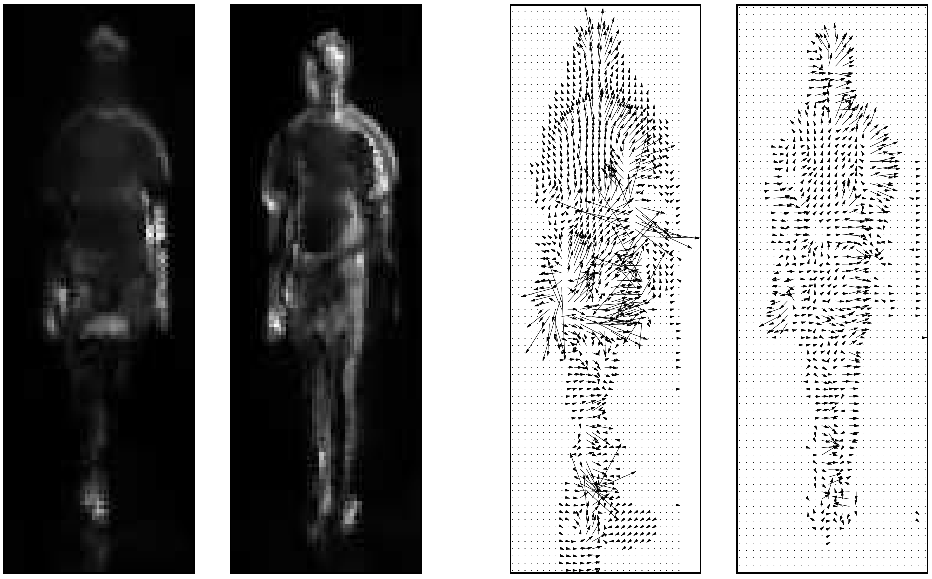

Abbildung2. Numerische Ergebnisse für den Bildausschnitt in Abbildung 1, rechts. **Von links nach rechts: Bild 1:** Die Grauwert–kodierten Beträge der Geschwindigkeitsvektoren als Ergebnis rein räumlicher 2D–Verarbeitung (Gln. (5)–(6)). Ausreißer verzerren das Gesamtergebnis erheblich. **Bild 2:** Das Ergebnis der räumlich–zeitlichen 3D–Verarbeitung (Gln. (8)–(9)) ergibt ein wesentlich kohärenteres und vollständiges Bild. **Bilder 3 und 4:** Die zu Bild 1 und Bild 2 gehörigen Geschwindigkeitsvektorfelder (zur Visualisierung mit dem Faktor 2 unterabgetastet).

Abbildung3. Hamburger Taxi–Szene.

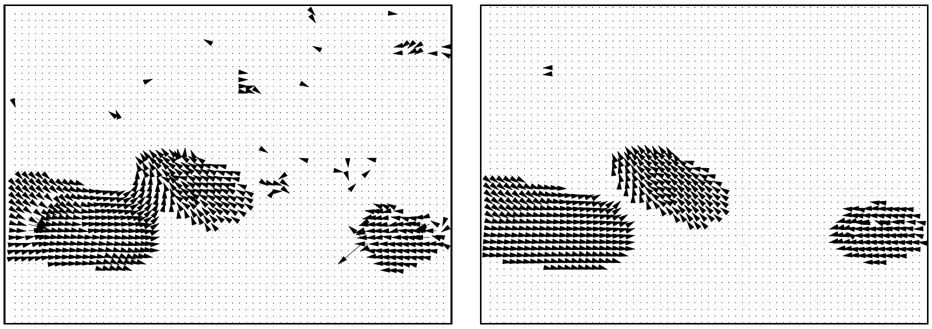

Abbildung4. Numerische Ergebnisse für die Taxi–Szene (zur Visualisierung mit dem Faktor 4 unterabgetastet). **Links:** Rein räumliche 2D–Verarbeitung. **Rechts:** Räumlich–zeitliche 3D–Verarbeitung. Ein Vergleich zeigt deutlich, daß die vorgeschlagene Erweiterung der adaptiven Glättung in Richtung der Zeitachse (i) das berechnete Vektorfeld signifikant verbessert, (ii) Hintergrundrauschen wirksam unterdrückt, und (iii) Bewegungsgrenzen deutlicher hervortreten läßt.

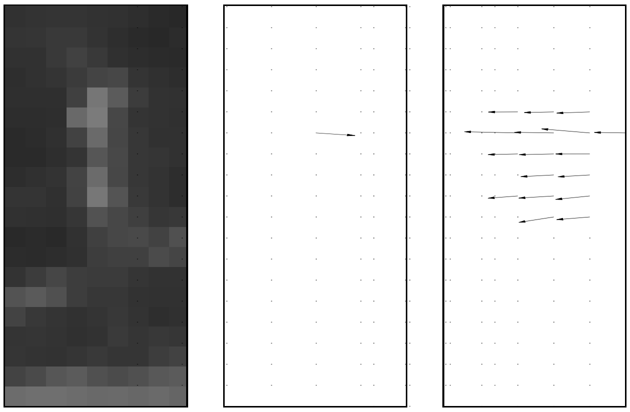

Abbildung5. Bild links: Bildausschnitt mit dem Fußgänger im oberen linken Bildbereich der Taxi–Szene aus Abbildung 3. **Mittleres Bild:** Der entsprechende Bildausschnitt aus Abbildung 4, links. Zur Visualisierung wurden die Vektorfelder mit dem Faktor 4 skaliert. Der rein räumliche 2D–Prozeß glättet die verrauschten, aus der Bildfunktion ermittelten lokalen Bewegungsdaten (Normalfluß) ohne ein Vektorfeld zu ergeben, welches die Bewegung korrekt beschreibt. **Bild rechts:** Der entsprechende Bildausschnitt aus Abbildung 4, rechts. Der 3D–Prozeß berechnet ein kohärentes Flußfeld für den Rumpf des Fußgängers. Die Bereiche mit sich bewegenden Gliedmaßen werden bei dieser kleinen örtlichen Auflösung als Rauschen interpretiert.

5 Zusammenfassung und Ausblick

Ein *nichtlinearer, räumlich–zeitlicher Regularisierungsansatz* zur Berechnung stückweise glatter optischer Flußfelder wurde vorgestellt. Numerische Ergebnisse belegen einen signifikante Verbesserung gegenüber einer rein räumlichen 2D–Verarbeitung auf Kosten eines moderaten zusätzlichen Berechnungsaufwandes. Sie bestätigen, daß der Vorschlag von Nagel [12], räumlich–zeitliche Glattheitsterme zu betrachten, vielversprechend ist. Gegenstand zukünftiger Arbeiten wird eine weitere kritische Evaluierung sein (vgl. [1]) sowie die Anwendung effizienter numerischer Methoden auf den hier vorgestellten Ansatz.

Literatur

1. J.L. Barron, D.J. Fleet, S.S. Beauchemin, *Performance of optical flow techniques*, Int. J. Comput. Vision, Vol. 12, 43–77, 1994.
2. M.J. Black, P. Anandan, *Robust dynamic motion estimation over time*, Proc. IEEE Comp. Soc. Conf. on Computer Vision and Pattern Recognition (CVPR '91, Maui, June 3–6, 1991), IEEE Computer Society Press, Los Alamitos, 292–302, 1991.
3. I. Cohen, *Nonlinear variational method for optical flow computation*, Proc. Eighth Scandinavian Conf. on Image Analysis (SCIA '93, Tromsø, May 25–28, 1993), Vol. 1, 523–530, 1993.
4. R. Deriche, P. Kornprobst, G. Aubert, *Optical-flow estimation while preserving its discontinuities: A variational approach*, Proc. Second Asian Conf. Computer Vision (ACCV '95, Singapore, December 5–8, 1995), Vol. 2, 290–295, 1995.
5. D.J. Fleet, A.D. Jepson, *Computation of component image velocity from local phase information*, Int. J. Comput. Vision, Vol. 5, 77–104, 1990
6. G. Gerig, O. Kübler, R. Kikinis, F.A. Jolesz, *Nonlinear anisotropic filtering of MRI data*, IEEE Trans. Medical Imaging, Vol. 11, 221–232, 1992.
7. B. Horn, B. Schunck, *Determining optical flow*, Artif. Intell., Vol. 17, 185–203, 1981.
8. B. Jähne, *Spatio-temporal image processing*, Lecture Notes in Computer Science, Vol. 751, Springer, Berlin, 1993.
9. A. Kumar, A.R. Tannenbaum, G.J. Balas, *Optic flow: a curve evolution approach*, IEEE Trans. Image Proc., Vol. 5, 598–610, 1996.
10. A. Mitiche, P. Bouthemy, *Computation and analysis of image motion: a synopsis of current problems and methods*, Int. J. Comput. Vision, Vol. 19, 29–55, 1996.
11. D.W. Murray, B.F. Buxton, *Scene segmentation from visual motion using global optimization*, IEEE Trans. Pattern Anal. Mach. Intell., Vol. 9, 220–228, 1987.
12. H.H. Nagel, *Extending the 'oriented smoothness constraint' into the temporal domain and the estimation of derivatives of optical flow*, O. Faugeras (Ed.), Computer vision – ECCV '90, Lecture Notes in Computer Science, Vol. 427, Springer, Berlin, 139–148, 1990.
13. H.H. Nagel, W. Enkelmann, *An investigation of smoothness constraints for the estimation of displacement vector fields from images sequences*, IEEE Trans. Pattern Anal. Mach. Intell., Vol. 8, 565–593, 1986.
14. P. Perona, J. Malik, *Scale space and edge detection using anisotropic diffusion*, IEEE Trans. Pattern Anal. Mach. Intell., Vol. 12, 629–639, 1990.

15. C. Schnörr, *Unique reconstruction of piecewise smooth images by minimizing strictly convex non-quadratic functionals*, J. Math. Imag. Vision, Vol. 4, 189–198, 1994.
16. C. Schnörr, *Segmentation of visual motion by minimizing convex non-quadratic functionals*, Proc. 12th Int. Conf. Pattern Recognition (ICPR 12, Jerusalem, Oct. 9–13, 1994), Vol. A, IEEE Computer Society Press, Los Alamitos, 661–663, 1994.
17. J. Weber, J. Malik, *Robust computation of optical flow in a multi-scale differential framework*, Int. J. Comput. Vision, Vol. 14, 67–81, 1995.
18. J. Weickert, *Anisotropic diffusion in image processing*, Teubner, Stuttgart, 1998.
19. J. Weickert, *On discontinuity-preserving optic flow*, S. Orphanoudakis, P. Trahanias, J. Crowley, N. Katevas (Eds.), Proc. Computer Vision and Mobile Robotics Workshop (CVMR '98, Santorini, Sept. 17–18, 1998), 115–122, 1998.