Combined 2D / 3D Face Recognition using Log-Gabor Templates

Jamie Cook, Chris McCool, Vinod Chandran and Sridha Sridharan
Image and Video Research Laboratory
Queensland University of Technology
GPO Box 2434, Brisbane Qld 4000, Australia
{j.cook, c.mccool, v.chandran, s.sridharan}@qut.edu.au

Abstract

The addition of Three Dimensional (3D) data has the potential to greatly improve the accuracy of Face Recognition Technologies by providing complementary information. In this paper a new method combining intensity and range images and providing insensitivity to expression variation based on Log-Gabor Templates is presented. By breaking a single image into 75 semi-independent observations the reliance of the algorithm upon any particular part of the face is relaxed allowing robustness in the presence of occulusions, distortions and facial expressions. Also presented is a new distance measure based on the Mahalanobis Cosine metric which has desirable discriminatory characteristics in both the 2D and 3D domains. Using the 3D database collected by University of Notre Dame for the Face Recognition Grand Challenge (FRGC), benchmarking results are presented demonstrating the performance of the proposed methods.

1 Introduction

Face as a biometric has the distinct advantage over other modalities such as fingerprint, DNA and iris recognition, in that the acquisition stage is non-intrusive and can be achieved with readily available equipment. 3D representations of the human face have the potential to overcome many of the obstacles such as pose and illumination sensitivity, which have prevented the widespread adoption of Face Recognition Technology (FRT).

Early work in 3D facial recognition emerged in the late 1980's but it wasn't until recently that substantial research databases have become available. The Face Recognition Grand Challenge [1] aims to address this issue and provides both a common dataset and experiment methodologies to enable accurate comparisons of different algorithms.

In [2] the authors present a good summary of the current research in 3D and composite 2D-3D recognition, in particular they note that while it is accepted that a combination of 2D and 3D gives greater performance, it is still unclear which modality performs better in isolation. In the following work more focus was applied to the 3D domain, however the methodology is equally applicable to combination with a more sophisticated 2D recognition engine.

In general, approaches to 3D recognition fall into 3 main categories [2]: those that use 3D correspondence matching explicitly to provide discrimination [3, 4]; those that extract 3D features such as curvature directly from the face; and those that treat the range image as a 2D image in order to extract features [5]. The latter has the advantage that a considerable number of well tested image processing algorithms can be directly applied.

Gabor filters are one such method which have been demonstrated to achieve high recognition rates in traditional 2D face recognition tasks [6, 7] and have been shown in [8] to exhibit robustness to misalignment. Techniques such as Hierarchial Graph Matching (HGM) and Elastic Bundle Graph Matching (EBGM) enhances this resilience further by adding a degree of freedom into the localisation of feature points [5].

In this paper a novel method of achieving robust face matching called Log-Gabor Templates (LGT) is presented. It is established that the use of multiple observations improves biometric performance; LGT exploits this fact by breaking a single acquisition of a subject into mulitple observations in both the spatial and frequency domains. These observations are each classified individually and recombined at the score level using linear Support Vector Machines (SVM). In this article it shall be shown that such a distributed approach is more resilient to local distortions such as expression variation.

2 Log-Gabor Filters

The Gabor family of wavelets first started gaining popularity in the field of image processing in 1980 when Daugmann first showed that the kernels exhibit useful properties

Proceedings of the IEEE International Conference
on Video and Signal Based Surveillance (AVSS'06)
0-7695-2688-8/06 $20.00 © 2006 IEEE

such as spatial localisation and orientation selectivity. Typically when used in face recognition applications [9] a family of filters is created where the filter at a scale v and orientation u is defined by,

$$\varphi_{u,v}(\mathbf{x}) = \frac{||k_{u,v}||^2}{\sigma^2} e^{-\frac{||k_{u,v}||^2||\mathbf{x}||^2}{2\sigma^2}} \left[e^{ik_{u,v}\mathbf{x}} - e^{-\frac{\sigma^2}{2}} \right] \quad (1)$$

where $k_{u,v} = k_v e^{i\phi_u}$ is the wave vector, σ is typically 2π and the final term, $e^{\frac{-\sigma^2}{2}}$, is employed to remove the D.C. offset.

In [10] Field proposes an alternate method to perform both the DC compensation and to overcome the bandwidth limitation of a traditional Gabor filter. The Log-Gabor filter has a response that is Gaussian when viewed on a logarithmic frequency scale instead of a linear one. This allows more information to be captured in the high frequency areas and also has desirable high pass characteristics. Field defines the frequency response of a Log-Gabor filter as,

$$\Phi(\mathbf{f}) = \exp - \frac{\log(\mathbf{f}/\mathbf{k})}{2\log(\sigma/\mathbf{k})}, \quad (2)$$

where $\mathbf{k} = [u_0\ v_0\ w_0 \cdots]^T$ is the centre frequency of the sinusoid and σ is a scaling factor of the bandwidth. In order to maintain constant shape ratio filters, the ratio of $\sigma/\mathbf{k}$ should be maintained constant. In the following experiments the shape parameter was chosen such that each filter had a bandwidth of approximately 2 octaves and the filter bank was constructed with a total of 6 orientations and 3 scales.

3 Face Verification

Face Verification techniques typically employ a monolithic representation of the face during recognition, however, approaches which decompose the face into subregions have shown considerable promise. Many authors [11, 12] have shown superior performance by adopting a modular representation of the face provided that face localisation is performed accurately [12].

3.1 Log-Gabor Templates

It is well established that using multiple probe images aids recognition performance, the same effect can be obtained by breaking a single face into multiple observations. After application of the 18 Log-Gabor filters, the face is broken into 25 square windows arranged in a 5x5 grid with 50% overlap in both the horizontal and vertical directions. These regions are then further decomposed by 3 scales of filter to generate 75 semi-independent observations for both the intensity and range images. An illustration of the decomposition process can be seen in Figure 1. Principal

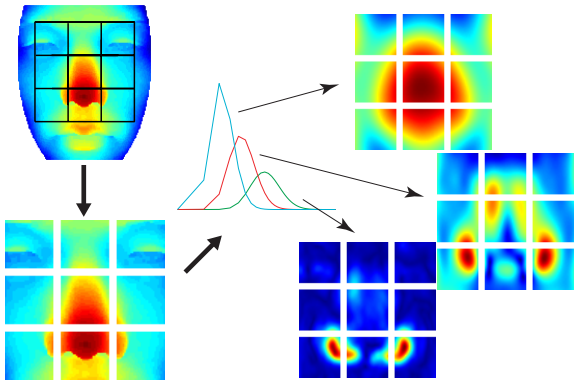

Figure 1. Decomposition of the face into subregions

Component Analysis (PCA) is used to build sub-spaces for each subregion from training data. In each region only the top 150 eigen-vectors are retained, thus each face is finally represented as 150 feature vectors each comprising 150 dimensions.

3.2 Distance Measure

The original Eigenfaces approach of Turk and Pentland used a simple Euclidean distance classifier, however, experimentation with a wide variety of distance metrics has shown that the Mahalinobis Cosine distance measure provides better performance [13]. This measure is defined as,

$$\begin{aligned} D_{MahCosine} &= -\frac{|m|\,|n|\cos\theta_{mn}}{|m|\,|n|} \\ &= -\frac{m.n}{|m|\,|n|}, \end{aligned} \quad (3)$$

where m and n are two feature vectors transformed into the Mahalanobis space. In this experimentation a new distance measure based on the Mahalanobis Cosine measure is proposed based upon the observation that the distance calculation is inherently a function of M, the number of retained eigenvectors. By averaging the angular distance between two vectors across a range of retained eigenvectors the importance of optimal selection of parameter M is reduced. The proposed distance metric is thus defined as

$$D_{MahCosAvg} = \frac{1}{M}\sum_{i=1}^{M} D^i_{MahCosine}. \quad (4)$$

3.3 Classifier Fusion

Given the multiple observations of a single face, a comparison between two faces generates 150 distance scores. In

Proceedings of the IEEE International Conference
on Video and Signal Based Surveillance (AVSS'06)
0-7695-2688-8/06 $20.00 © 2006 IEEE

this work linear Support Vector Machines (SVM)

$$\Psi(\mathbf{x}) = \mathrm{w}^T\mathbf{x} + \alpha, \quad (5)$$

are used to combine a multitude of scores back to a single value. SVMs were chosen because they combine information in a discriminatory sense, maximising the margin between client and imposter scores. Training data is used to derive sets of combination weights within each frequency band which are then concatenated to give the final weight vector w for each modality. Initial experimentation has shown that the 3D data being used performs approximately an order of magnitude better than the 2D and the combination of 2D/3D scores are weighted accordingly.

4 Dataset Description

The experiments described in this article were conducted using 3D data provided as part of the Face Recognition Grand Challenge [1]. The FRGC dataset, which contains 4007 registered texture and shape images of 466 subjects, is currently the largest publicly available database of 3D face images. The data was collected by the Computer Vision Research Laboratory at the University of Notre Dame (UND) over 3 semesters using a Minolta Vivid 900 range finder.

The 466 subjects in the database were broken into training and testing groups according to the specification of FRGC Experiment 3. Within Experiment 3 there are 3 subexperiments of increasing difficulty, all results quoted in this paper were evaluated on the hardest of these (Mask III) which is comprised of target/query pairs which are captured in different semesters. Unless otherwise stated all results are quoted as true acceptance rates at a False Acceptance Rate (FAR) of 0.1%.

Of the 4007 images in the test set 59% are captured with a neutral expression while the remainder are captured variously with expressions of surprise, happiness, sadness and disgust. Manual classification by researchers at Geometrix [4] shows that these non-neutral images are evenly distributed between mild and severe distortions. Examples of range images under various expressions are shown in Figure 2.

5 Experimentation

Before testing the robustness of the LGT method in the presence of expression variation the efficacy of the proposed distance metric and the distribution of discriminable information in both modalities are first evaluated.

5.1 Distance Metric

In Section 3.2, the Mahalanobis Cosine Average (MahCosAvg) distance measure was introduced, we now provide

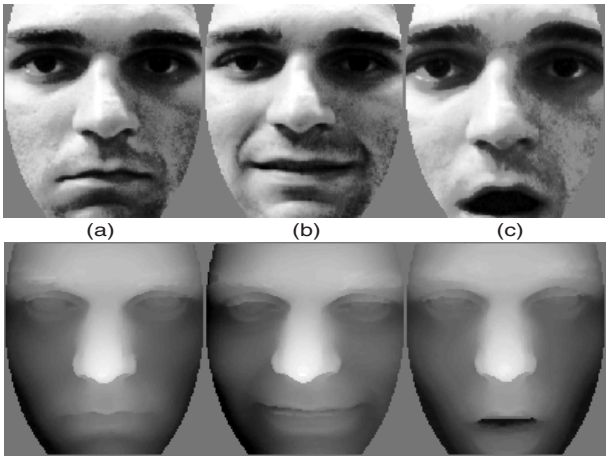

Figure 2. Examples of FRGC images with (a) neutral expression (b) small distortion and (c) large distortion

test cases evaluating the performance of the metric. The new distance measure was compared against an ensemble of other metrics on monolithic representations for both range and intensity images. Figure 3 shows the relative performance of the three top performing distance metrics. While slight improvements can be observed in the Covariance metric in the 2D modality these are completely offset by its poor performance in the 3D domain. The MahCosAvg met-

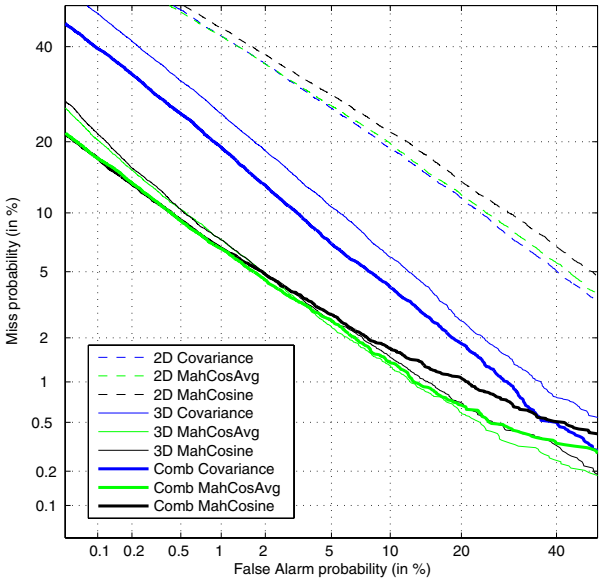

Figure 3. DET for Covariance, MahCosine and MahCosAvg distance metrics.

Proceedings of the IEEE International Conference
on Video and Signal Based Surveillance (AVSS'06)
0-7695-2688-8/06 $20.00 © 2006 IEEE

ric provides consistant performance improvment over the original MahCosine metric for both the 2D and 3D data and provides equivalent performance in the combined case. Of principal consideration is the low FAR region; in Table 1 the results at a FAR of 0.1% show that the MahCosAvg metric provides the best performance at this operating point.

	MahL2	Covariance	MahCosine	MahCosAvg
2D	37.59%	39.42%	37.40%	39.37%
3D	44.80%	51.92%	78.57%	79.66%
Comb	51.38%	60.46%	82.72%	82.72%

Table 1. Recognition rates for monolithic representations using various distance metrics.

5.2 Log-Gabor Decomposition

After dividing the face into 25 overlapping regions and calculating recognition performance in isolation, it is observed that the recognition rate:

- deteriorates with distance from the image center,
- is approximately symmetrical about the medial line,
- was better above the nose than below it,

Table 2 shows the recognition accuracy as progressively more regions surrounding the center are used in the classification process. In the 3D domain best performance is achieved by using only the centermost 5 regions whereas for 2D, best performance is achieved using the entire face.

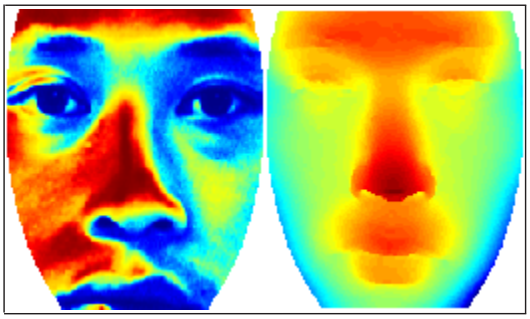

Figure 4. Example of a single normalised acquisition in both 2D and 3D.

To explain this contrast examination of the data is required. The 3D data collected for the FRGC is generally much more consistent than the corresponding 2D, which contains highly variant illumination conditions, this can be observed from Figures 2 and 4. This leads to the conclusion that the central regions of the face are sufficient to perform verification in clean 3D data, however the inclusion of extra regions is better able to compensate for the environmental conditions present in the 2D data. The overall performance of the LGT method with respect to the best performing monolithic system is shown in Figure 5 and consistent improvements can be observed in all modalities in the low FAR region. It is interesting to note that despite the significant performance gap between the 2D and 3D data the combination of both still yields an improvement.

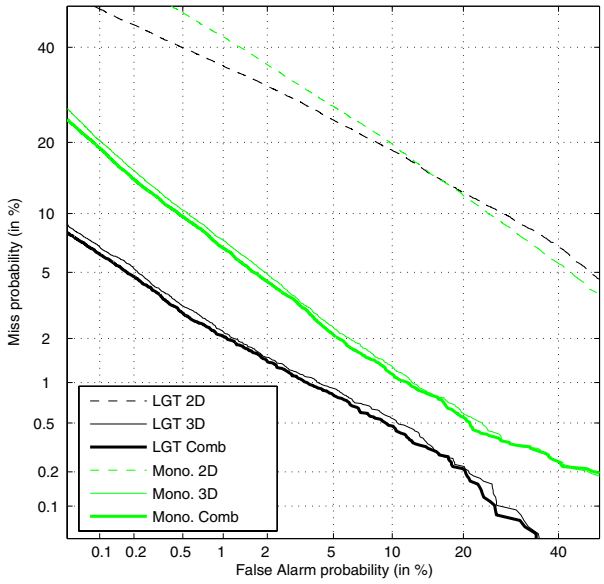

Figure 5. DET plot of LGT method (black) and Log-Gabor based monolithic classifier (red).

5.3 Expression Variation

In the previous section it was shown that best results in the 3D domain were achieved using a small closely packed set of subregions surrounding the nose while in the 2D domain best results were achieved using the entire face. In this section the effects of varying expresion levels upon each set of regions is examined and compared against the best performing monolithic approach used in previous sections.

In [4], the authors manually divide the FRGC 3D corpus into three categories based on strength of expression variation: Neutral, Small, Large. In order to test the robustness of the presented approach in the presence of expression variation the neutral images are used as the gallery and four probe sets are created containing progressively more variation. The four sets are respectively Neutral, Small, Small+Large and Large.

Proceedings of the IEEE International Conference
on Video and Signal Based Surveillance (AVSS'06)
0-7695-2688-8/06 $20.00 © 2006 IEEE

	2D				3D			
	C1	C5	C9	C25	C1	C5	C9	C25
Low	-	-	-	-	82.68	83.93	83.21	82.95
Mid	24.28	35.01	37.67	45.66	86.97	90.42	90.27	86.63
High	14.40	41.01	42.68	47.39	76.41	87.07	86.73	87.36
Combined	23.01	42.85	45.28	**50.67**	91.57	**93.16**	92.68	91.65

Table 2. Recognition rates (C5 indicates that the 5 centermost subregions were combined, C9 that the 9 centermost were combined and C25 that all subregions were combined).

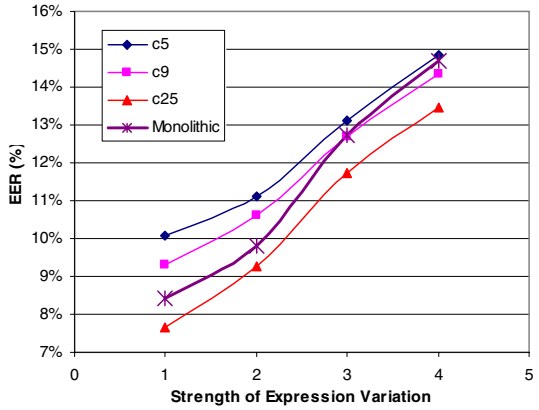

Figure 6. Equal Error Rate for increasing expression variation for 2D modality.

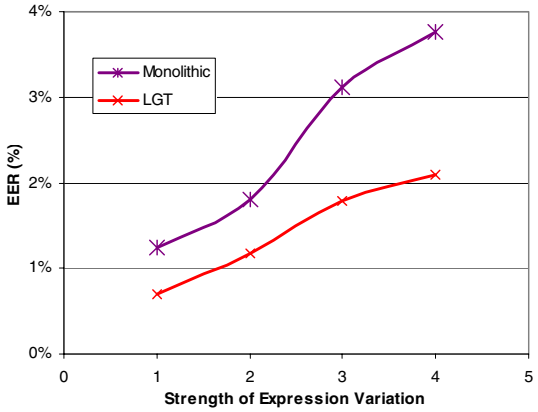

Figure 8. Equal Error Rate for increasing expression variation for combined.

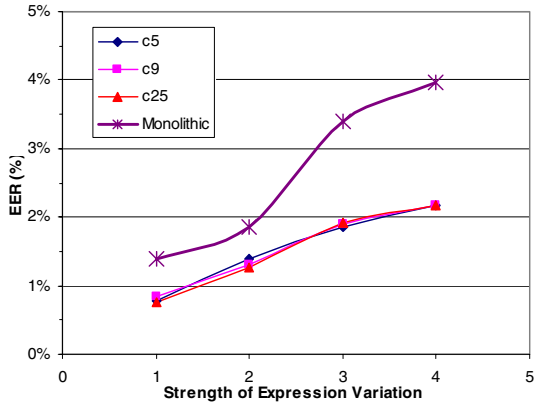

Figure 7. Equal Error Rate for increasing expression variation for 3D modality.

Figures 6, 7 and 8 show the Equal Error Rate as the strength of expression variation in the probe set is increased. As can be seen the monolithic representation suffers from a sharp degradation in performance when severe expression variation are introduced. In contrast the LGT based classifiers have a much more linear degradation over the same range.

This effect is significantly more noticeable in the 3D domain which can be attributed to the level of noise present in the two modalities. In the 2D domain the effects of expression variation can be subsumed by factors such as lighting variation, shadowing and surface reflectance. In the 3D domain expression is a much more dominant factor and thus provides a better indication of the LGT method's insensitivity to expression variation.

6 Conclusions

In this article a new distance metric is proposed for Nearest Neighbour comparison of 2D and 3D face images in a PCA based subspace. Testing has shown that the proposed metric has good discriminating power in low False Alarm regions.

Also presented is a novel and robust combined 2D/3D face recognition method. The Log-Gabor Templates method exploits the multitude of information available in

Proceedings of the IEEE International Conference
on Video and Signal Based Surveillance (AVSS'06)
0-7695-2688-8/06 $20.00 © 2006 IEEE

the human visage to construct multiple observations of a subject which are classified independently and combined with score fusion. The LGT method has been evaluated on the largest publicly available 3D face database. Results have shown that the parts based methodology adopted has better performance than an equivalent monolithic classifier and exhibits more graceful performance degradation in the presence of expression variation.

Acknowledgements

This research was supported by the Australian Research Council (ARC) through Discovery Grants Scheme, Grant DP0452676,2004-6.

References

[1] P. Jonathon Phillips, Patrick J. Flynn, Todd Scruggs, Kevin W. Bowyer, Jin Chang, Kevin Hoffman, Joe Marques, Jaesik Min, and William Worek, "Overview of the face recognition grand challenge," in *CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1*, Washington, DC, USA, 2005, pp. 947–954, IEEE Computer Society.

[2] Kevin W. Bowyer, Kyong Chang, and Patrick Flynn, "A survey of approaches and challenges in 3D and multi-modal 3D + 2D face recognition," *Computer Vision and Image Understanding*, vol. 101, no. 1, pp. 1–15, 2006, TY - JOUR.

[3] Kyong I. Chang, K.W. Bowyer, and P.J. Flynn, "Adaptive rigid multi-region selection for handling expression variation in 3d face recognition," in *Computer Vision and Pattern Recognition, 2005 IEEE Computer Society Conference on*, 2005, vol. 3, p. 157.

[4] Thomas Maurer, David Guigonis, Igor Maslov an d Bastien Pesenti, Alexei Tsaregorodtsev, David West, and Gerard Medioni, "Performance of Geometrix ActiveID^{TM}3D Face Recognition Engine on the FRGC Data," in *CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, Washington, DC, USA, 2005, IEEE Computer Society.

[5] Michael Husken, Michael Brauckmann, Stefan Gehlen, and Christoph Von der Malsburg, "Strategies and benefits of fusion of 2D and 3D face recognition," in *CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops*, Washington, DC, USA, 2005, p. 174, IEEE Computer Society.

[6] B. Duc, S. Fischer, and J. Bigun, "Face Authentication with gabor Information on Deformable Graphs," *IEEE Transactions on Image Processing*, vol. 8, no. 4, pp. 504–516, April 1999.

[7] R. Chellapa, C. L. Wilson, S.Sirohey, and C. S. Barnes, "Human and Machine Recogntion of Faces: A Survey," Tech. Rep., University of Maryland and NIST, August 1994.

[8] Shiguang Shan, Wen Gao, Yizheng Chang, Bo Cao, and Pang Yang, "Review the strength of gabor features for face recognition from the angle of its robustness to mis-alignment.," in *ICPR (1)*, 2004, pp. 338–341.

[9] C Liu and H. Wechsler, "Independent Component Analysis of Gabor Features for Face Recognition," *IEEE Transactions. Neural Networks*, vol. 14, no. 4, pp. 919–928, 2003.

[10] D. Fields, "Relations between the statistics of natural imags and the response properies of cortical cells," *Journal of Optical Society of America*, vol. 4, no. 12, pp. 2379–2394, 1987.

[11] Roberto Brunelli and Tomaso Poggio, "Face Recognition: Features versus Templates.," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 10, pp. 1042–1052, 1993.

[12] S. Lucey and Tsuhan Chen, "Face recognition through mismatch driven representations of the face," in *Visual Surveillance and Performance Evaluation of Tracking and Surveillance, 2005. 2nd Joint IEEE International Workshop on*, 2005, pp. 193–199.

[13] M. Teixeira D. Bolme, R. Beveridge and B. Draper, "The CSU Face Identification Evaluation System: Its Purpose, Features and Structure," in *International Conference on Vision Systems*. 2003, pp. 304–311, Springer-Verlag.