Final print version (as of July 7, 2001)
for Proceedings Deutsche Arbeitsgemeinschaft für Mustererkennung (DAGM2001)
Munich, September 2001 (to appear in Springer Lecture Notes in Computer Science)

Illumination-invariant change detection using a statistical colinearity criterion

Rudolf Mester[1], Til Aach[2], and Lutz Dümbgen[3]

[1] Institute for Applied Physics, University of Frankfurt
Robert-Mayer-Str. 2–4, D-60054 Frankfurt, Germany
[2] Institute for Signal Processing, University of Lübeck
[3] Institute for Mathematics, University of Lübeck
Ratzeburger Allee 160, D-23538 Lübeck, Germany
mester@iap.uni-frankfurt.de, aach@isip.mu-luebeck.de,
duembgen@math.mu-luebeck.de

Abstract. This paper describes a new algorithm for illumination-invariant change detection that combines a simple multiplicative illumination model with decision theoretic approaches to change detection. The core of our algorithm is a new statistical test for linear dependence (colinearity) of vectors observed in noise. This criterion can be employed for a significance test, but a considerable improvement of reliability for real-world image sequences is achieved if it is integrated into a Bayesian framework that exploits spatio-temporal contiguity and prior knowledge about shape and size of typical change detection masks. In the latter approach, an MRF-based prior model for the sought change masks can be applied successfully. With this approach, spurious spot-like decision errors can be almost fully eliminated.

1 Introduction

In many computer vision applications, the detection and accurate delineation of moving objects forms an important first step. Many video surveillance systems, especially those employing a static or quasi-static camera, use processing algorithms that first identify regions where at least potentially a motion can be observed, before these regions are subject to further analysis steps, which might be, for instance, a quantitative analysis of motion. By this procedure, the available processing power of the system can be focused on the relevant subareas of the image plane. In applications such as traffic surveillance or video-based security systems, this focusing on the moving parts of the image typically yields a reduction of the image area to be processed in more detail to about 5-10 percent. Obviously, such a strategy is very advantageous compared to applying costly operations such as motion vector estimation to the total area of all images.

In order to let a change detection scheme be successful, an utmost level of robustness against different kinds of disturbances in typical video data is required (a very low false alarm rate), whereas any truly relevant visual event should be detected and forwarded to more sophisticated analysis steps (high sensitivity). Obviously, this presents change detection as a typical problem that should be dealt with by statistical decision and detection theory. Some early papers [2, 4] stress the importance of selecting the most efficient test statistic, which should be both adapted to the noise process and to a suitably chosen image model. A certain boost in performance has been introduced in the early nineties by employing prior models for the *typical shape and size* of the objects to be detected; this can be achieved very elegantly using *Gibbs-Markov random fields* [9, 5, 6]. These models reduce very strongly the probability of false positive and false negative alarms due to the usage of spatio-temporal context. They are superior to any kind of change mask post-processing (e.g. by morphological filters), since both the shape *and* the strength of observed signal anomaly is used. Such algorithms are in the meantime widely accepted as state of the art and integrated in multimedia standard proposals such as [11].

Despite all these advancements, certain problematic situations remain in real-life video data, and possibly the most important ones are *illumination changes*. A rapid change in illumination does cause an objectively noticeable signal variation, both visually and numerically, but it is very often not regarded as a *relevant* event. Therefore, recent investigations [10, 12, 13] have put emphasis on the desired feature of *illumination invariance*. In contrast to earlier work, the present paper aims at integrating illumination invariance in a framework that is as far as possible based on decision theory and statistical image models.

2 The image model, including illumination changes and superimposed noise

We model the recorded image brightness as the product of illumination and reflectances of the surfaces of the depicted objects. We furthermore assume that the illumination is typically a spatially slowly varying function (cf. [1]).

For change detection, we compare for two subsequent images the grey levels which lie in a small sliding window[1]. Due to the spatial low-frequency behaviour of illumination, we can assume that illumination is almost constant in each small window. Thus, if no structural scene change occurs within the window, temporal differences between observed grey levels in the window can be caused only

1. by a positive multiplicative factor k which modulates the signal and accounts for illumination variation
2. and secondly by superimposed noise which can be modelled as i.i.d. Gaussian or Laplacian noise.

Let us consider the case that this *null hypothesis* H_0 is true: if we order the grey values from the regarded windows W_1 and W_2 into column vectors $\mathbf{x}_1$ and $\mathbf{x}_2 \in \mathbb{R}^N$, these are related by $\mathbf{x}_1 = \mathbf{s} + \boldsymbol{\varepsilon}_1$ and $\mathbf{x}_2 = k \cdot \mathbf{s} + \boldsymbol{\varepsilon}_2$, where $\boldsymbol{\varepsilon}_i$, $i = 1, 2$, are additive noise vectors,

$$\mathsf{E}[\boldsymbol{\varepsilon}_1] = \mathsf{E}[\boldsymbol{\varepsilon}_2] = \mathbf{0}\,, \qquad \mathsf{Cov}[\boldsymbol{\varepsilon}_1] = \mathsf{Cov}[\boldsymbol{\varepsilon}_2] = \sigma_d^2 \cdot \mathbf{I}_N \tag{1}$$

and $\mathbf{s}$ is a signal vector. Unfortunately, the signal vector $\mathbf{s}$ is unknown. In such situations it might on the first glance appear reasonable to employ some kind of signal model

[1] or a fixed block raster, if detection and speed is of primary interest, and not so much the spatial accuracy of the detection masks

(e.g. using cubic facets). However, considering the extremely variable local structure of real video scenes, especially for outdoors applications, it is certainly not advantageous to model the signal blocks as e.g. low-pass signals, since the algorithm would work worse in this locations where the signal model incidentally differs strongly from the true signal such as in highly textured areas. In an ideal noise free case, $\mathbf{x}_1$ and $\mathbf{x}_2$ are parallel given H_0. We thus formulate change detection as testing whether or not $\mathbf{x}_1$ and $\mathbf{x}_2$ can be regarded as degraded versions of colinear vectors, with factor k and the true signal vector $\mathbf{s}$ being unknown entities (so-called *nuisance parameters*).

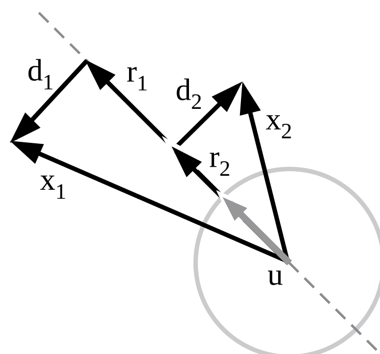


Fig. 1. Geometrical interpretation of testing the colinearity of two vectors $\mathbf{x}_1, \mathbf{x}_2$.

2.1 Derivation of the colinearity test statistic

Earlier (and simpler) approaches to the task of testing the colinearity concentrated either on the difference between the two observed vectors $\mathbf{x}_i$ or on regarding the *angular difference* between the $\mathbf{x}_i$. It is clearly not advisable to normalize both vectors prior to further processing, since this irreversibly suppresses statistically significant information. With a certain noise level σ_d^2 given, it makes a *significant* difference whether an certain angular difference is found for 'long' or 'short' signal vectors – it is simply much easier to change the direction of a 'short' signal vector. Therefore, basing the change detection decision on the angle between the observed vectors (e.g. by using the normalized correlation coefficient) cannot be recommended.

The approach we propose instead aims as much as possible on preserving any bit of statistical information in the given data. Fig. 1 illustrates the derivation of the test statistic. Given the observations $\mathbf{x}_i$, $i = 1,2$ and assuming i.i.d. Gaussian noise, a maximum likelihood (ML) estimate of the true signal 'direction' (represented by a unit vector $\mathbf{u}$) is given by minimizing the sum $D^2 = |\mathbf{d}_1|^2 + |\mathbf{d}_2|^2$ of the squared distances $\mathbf{d}_i$ of the observed vectors $\mathbf{x}_i$ to the axis given by vector $\mathbf{u}$. Clearly, if $\mathbf{x}_1$ and $\mathbf{x}_2$ are colinear, the difference vectors and hence the sum of their norms are zero [2]. The projections $\mathbf{r}_i$,

[2] Note that $\mathbf{s}$, $k\mathbf{s}$, $\varepsilon_1, \varepsilon_2$ are unknown entities, and $\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{d}_1$, $\mathbf{d}_2$ are only their estimates.

$i = 1,2$ are ML estimates of the corresponding signal vectors. Obviously, we have

$$|\mathbf{d}_i|^2 = |\mathbf{x}_i|^2 - |\mathbf{r}_i|^2 \qquad \text{for} \quad i = 1,2$$

$$|\mathbf{r}_i| = ||\mathbf{x}_i| \cdot \cos\varphi_i| = \left|\mathbf{x}_i^T \cdot \mathbf{u}\right| \qquad (|\mathbf{u}| = 1!)$$

$$\Longrightarrow \quad |\mathbf{d}_i|^2 = |\mathbf{x}_i|^2 - \left|\mathbf{x}_i^T \cdot \mathbf{u}\right|^2$$

$$\Longrightarrow \quad D^2 \stackrel{def}{=} |\mathbf{d}_1|^2 + |\mathbf{d}_2|^2 = |\mathbf{x}_1|^2 + |\mathbf{x}_2|^2 - \left|\mathbf{x}_1^T \cdot \mathbf{u}\right|^2 - \left|\mathbf{x}_2^T \cdot \mathbf{u}\right|^2$$

Let us now form the $2 \times N$ matrix $\mathbf{X}$ with

$$\mathbf{X} \stackrel{def}{=} \begin{pmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \end{pmatrix} \qquad \mathbf{X} \cdot \mathbf{u} = \begin{pmatrix} \mathbf{x}_1^T \cdot \mathbf{u} \\ \mathbf{x}_2^T \cdot \mathbf{u} \end{pmatrix}$$

$$\Longrightarrow \quad |\mathbf{X} \cdot \mathbf{u}|^2 = \mathbf{u}^T \cdot \mathbf{X}^T \cdot \mathbf{X} \cdot \mathbf{u} = \left|\mathbf{x}_1^T \cdot \mathbf{u}\right|^2 + \left|\mathbf{x}_2^T \cdot \mathbf{u}\right|^2$$

So it turns out that

$$D^2 = |\mathbf{d}_1|^2 + |\mathbf{d}_2|^2 = |\mathbf{x}_1|^2 + |\mathbf{x}_2|^2 - \mathbf{u}^T \cdot \mathbf{X}^T \cdot \mathbf{X} \cdot \mathbf{u}$$

and the vector $\mathbf{u}$ that minimizes D^2 is the same vector that maximizes

$$\mathbf{u}^T \cdot \mathbf{X}^T \cdot \mathbf{X} \cdot \mathbf{u} \longrightarrow \max \qquad \text{with} \, |\mathbf{u}| = 1$$

which is obviously an eigenvalue problem[3] with respect to matrix $\mathbf{X}^T \cdot \mathbf{X}$. Due to the special way it is constructed from just two vectors, $\mathbf{X}^T \cdot \mathbf{X}$ has maximum rank 2 and thus has only two non-vanishing eigenvalues. We are only interested in the value of the test statistic D^2, and fortunately it can be shown that D^2 is identical to the smallest non-zero eigenvalue of matrix $\mathbf{X}^T \cdot \mathbf{X}$. Beyond that, it can be shown (e.g. quite illustratively by using the *singular value decomposition* (SVD) of matrix $\mathbf{X}$) that the non-zero eigenvalues of $\mathbf{X}^T \cdot \mathbf{X}$ and $\mathbf{X} \cdot \mathbf{X}^T$ are identical. Thus, the sought eigenvalue is identical to the smaller one of the two eigenvalues of the 2×2 matrix $\mathbf{X} \cdot \mathbf{X}^T$, which can be computed in closed form without using iterative numerical techniques. So the minimum value for D^2 can be determined without explicitly computing the 'signal direction unit vector' $\mathbf{u}$. This whole derivation is strongly related to the Total Least Squares (TLS) problem (cf. [14]) and matrix rank reduction tasks in modern estimation theory.

2.2 The distribution of test statistic D^2

In order to construct a mathematically and statistically meaningful decision procedure, the distribution of the test statistic D^2 must be known at least for the null hypothesis H_0 (= colinearity). The asymptotic distribution of the test statistic D^2 can be derived on the basis of some mild approximations. For the case that the norm of the signal vector $\mathbf{s}$ is significantly larger than the expected value for the noise vector norm (which

[3] The matrix $\mathbf{X}^T \cdot \mathbf{X}$ can also be regarded as a (very coarse) estimate of the correlation matrix between the vectors $\mathbf{x}_i$, but this does not provide additional insight into the task of optimally testing the colinearity.

should hold true in almost all practical cases), it can be shown that the sum $|\mathbf{d}_1|^2 + |\mathbf{d}_2|^2$ is proportional to a χ^2 variable with $N-1$ degrees of freedom with a proportionality factor σ_d^2 according to eg. (1).

$$D^2 \quad \sim \quad \sigma_d^2 \cdot \chi^2_{N-1} \tag{2}$$

This result can be intuitively understood: assuming that the probability density function of the additive noise for each pixel is a zero-mean i.i.d. Gaussian with the same variance for all N pixels, the difference vectors $\mathbf{d}_i$ reside in a $N-1$-dimensional subspace of $\mathbb{R}^N$ which is determined by the direction vector $\mathbf{u}$. If the length $|\mathbf{s}|$ of the signal vector is large, the direction $\mathbf{u}$ is independent of the additive noise vectors ε_i. The components of $\mathbf{d}_i$ retain the property of being zero-mean Gaussians.

It might be surprising that the actual value of the multiplicative factor k does not influence the distribution of D^2, at least as long as the assumption of $|\mathbf{s}| \gg |\varepsilon_i|$ holds. This makes this decision invariant against (realistic) multiplicative illumination changes, which is of course exactly what it has been developed for.

The distribution which has been theoretically derived and described above was also exactly what we found in a set of Monte Carlo simulations for the test statistic D^2 for N varying between 4 and 64 and factor k varying between 0.2 and 5 (which corresponds, in fact, already to a quite strong multiplicative change of illumination). Figures 2 and 3 show examples of the empirical distributions obtained by these simulations.

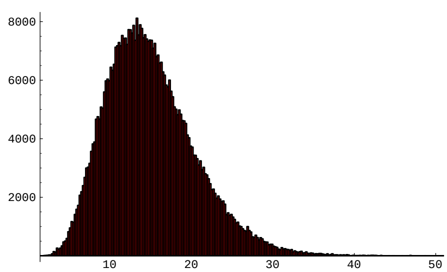


Fig. 2. Empirical distribution of D^2 for $\sigma_d^2 = 1$, $N = 16$, 100000 realizations. These empirical distributions do not noticeably change when k varies with $0.2 < k < 5$.

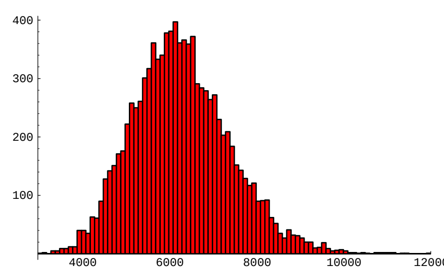


Fig. 3. Empirical distribution of D^2 for $\sigma_d^2 = 100$, $N = 64$, 10000 realizations. Note the conformity to the predicted scaling of $D^2 \sim \sigma_d^2 \cdot \chi^2_{N-1}$

Testing the null hypothesis H_0 (= colinearity) can be expressed as testing whether or not D^2 can be explained by the given noise model. On the basis of the now known distribution of D^2 under the null hypothesis, a significance test can be designed, which boils down to testing D^2 against a threshold t which has been determined in such a way that the conditional probability $Prob[D^2 > t \mid H_0] = \alpha$ with the *significance level* α.

3 Integration of the colinearity test statistic into a Bayesian MRF-based framework

To improve the (already very good) performance of this test even further, we have integrated the new test statistic into the Bayesian framework of earlier, illumination sensitive change detection algorithms [6, 8]. The Bayesian approach draws its power from using Gibbs/Markov random fields (MRF) for expressing the prior knowledge that the objects or regions to be detected are mostly compactly shaped. For integrating the new test, the conditional distribution $p(D^2|H_1)$ of the test statistic D^2 under the alternative hypothesis (H_1) has to be known at least coarsely. The resulting algorithm compares the test statistic D to an adaptive context dependent threshold, and is non-iterative. Thereby, the new approach is a illumination-invariant generalization of the already very powerful scheme [6, 8] which already was an improvement over the iterative proposal [4, 5].

Under the alternative hypothesis H_1 (vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ are *not* colinear), we model the conditional pdf $p(D^2|H_1)$ by

$$p(D^2|H_1) = \left(\frac{1}{\sqrt{2\pi}\sigma_c}\right)^N \cdot \exp\left(-\frac{D^2}{2\cdot\sigma_c^2}\right) \tag{3}$$

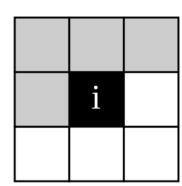
with $\sigma_c^2 \gg \sigma_d^2$ (for detail cf. [5, 8]). The assumption that this density is Gaussian does not matter very much; it is just important that the variance σ_c^2 is significantly larger than σ_d^2 and that the distribution is almost flat close to the origin. Furthermore, we model the sought change masks by an MRF such that the detected changed regions tend to be compact and smoothly shaped. From this model, *a priori* probabilities $\text{Prob}(c)$ and $\text{Prob}(u)$ for the labels c (changed) and u (unchanged) can be obtained. The *maximum a priori* (MAP) decision rule – given the labels in the neighbourhood of the regarded block – is then

$$\frac{p(D^2|H_1)}{p(D^2|H_0)} \underset{u}{\overset{c}{\gtrless}} \frac{\text{Prob}(u)}{\text{Prob}(c)} \tag{4}$$

A little algebraic manipulation yields the context adaptive decision rule

$$D^2 \underset{u}{\overset{c}{\gtrless}} T + (4-\nu_c)\cdot B \tag{5}$$

where D^2 is the introduced test statistic , and T a fixed threshold which is modulated by an integer number ν_c. The parameter ν_c denotes the number of pixels that carry the label c and lie in the 3×3-neighbourhood of the pixel to be processed (see figure). These labels are known for those neighbouring pixels which have already been processed while scanning the image raster (causal neighbourhood), as symbolized by the gray shade in the illustration.



For the pixels which are not yet processed we simply take the labels from the previous change mask (anticausal neighbourhood). Clearly, the adaptive threshold on the right hand side of (5) can only take the nine different values $\nu_c = 0, 1, \ldots, 8$. The parameter B is a positive cost constant. The adaptive threshold hence is the lower, the higher the number ν_c of adjacent pixels with label c. It is obvious that this behaviour favours the emergence of smoothly shaped changed regions, and discourages noise-like decision errors. The nine different possible values for the adaptive threshold can be precomputed and stored in a look-up table, so this procedure needs only slightly more computational effort than just applying a fixed threshold.

4 Results

Figure 4 shows some typical experimental results obtained by using the described technique. In the used image sequence there is true motion (a toy train) and a visually and numerically strong illumination change obtained by waving a strong electric torch across the scene. A comparison of image c), where a conventional change detection technique is employed, versus image d) (illumination invariant change detection) shows the advantages of the new technique very clearly.

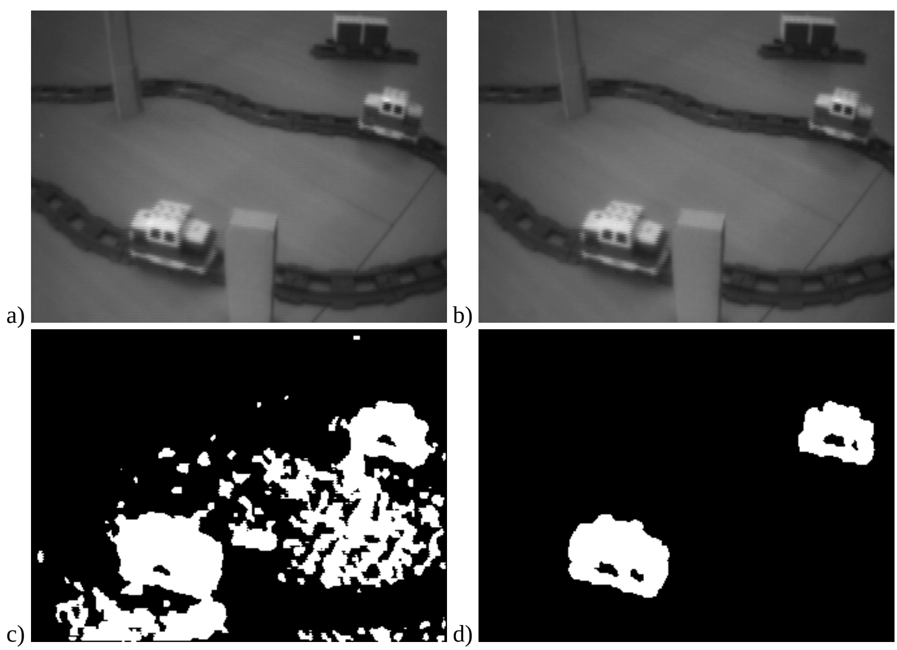


Fig. 4. a), b): Subsequent original frames from a sequence with moving toy trains. A beam of light crosses this scene quickly from left to right. c) Result of the illumination sensitive change detection algorithm in [8], mixing illumination changes with the moving locomotives. d) Result of the new illumination invariant change detection. The locomotives have been safely detected, and all erroneous detection events due to illumination changes are suppressed.

5 Conclusions

We consider the illumination-invariant change detection algorithm presented here as a significant step forward compared to earlier (already quite well performing) statistics-based approaches. For the near future, a comparison to competing approaches using homomorphic filtering (cf. [13]) remains to be performed. Furthermore, it appears to be very promising to extend the discussed approaches to change detection towards the integrated processing of more than 2 subsequent frames. We are convinced that – if again statistical modeling and reasoning is employed – a further improvement compared to the *state of the art* can be achieved.

References

1. A. V. Oppenheim, R. W. Schafer, T. G. Stockham Jr: Nonlinear filtering of multiplied and convolved signals. *Proc. IEEE* 56(8):1264–1291, 1968.
2. Y. Z. Hsu, H.-H. Nagel, G. Rekers: New likelihood test methods for change detection in image sequences. *Comp. Vis. Grap. Im. Proc* 26:73–106, 1984.
3. K. Skifstad, R. Jain: Illumination independent change detection for real world image sequences. *Computer Vision, Graphics, and Image Processing,* 46, pp. 387–399, 1989.
4. T. Aach, A. Kaup, R. Mester: A statistical framework for change detection in image sequences. *13ième Colloque GRETSI*, pp. 1149–1152, Juan-Les-Pins, France, Sept. 1991.
5. T. Aach, A. Kaup, R. Mester: Statistical model-based change detection in moving video. *Signal Processing* 31(2), pp.165–180, 1993.
6. T. Aach, A. Kaup, R. Mester: Change detection in image sequences using Gibbs random fields. *IEEE Internal. Works. Intell. Signal Processing Com. Sys.*, pp. 56–61, Sendai, Japan, Oct. 1993.
7. G. Tziritas, C. Labit: *Motion Analysis for Image Sequence Coding*. Elsevier, 1994.
8. T. Aach, A. Kaup: Bayesian algorithms for change detection in image sequences using Markov random fields. *Signal Processing: Image Communication* 7(2): pp. 147–160, 1995.
9. A. Mitiche, P. Bouthemy: Computation and analysis of image motion: A synopsis of current problems and methods. *Inl. Jour. Comp. Vis.* 19(1):29–55, 1996.
10. S.-Z. Liu, C.-W. Fu, S. Chang: Statistical change detection with moments under time-varying illumination. *IEEE Trans. Image Processing* 7(9), pp. 1258–1268, 1998.
11. ISO/IEC JTC 1/SC 29/WG 11 N2502, Final Draft International Standard, Annex F, Information Technology - Very-low bitrate audio-visual coding - Part 2: Visual, Atlantic City, 1998.
12. E. Durucan, T. Ebrahimi: Robust and illumination invariant change detection based on linear dependence for surveillance application. Proc. *EUSIPCO 2000,* pp. 1041–1044, Tampere, Finland, Sept. 3-8, 2000.
13. D. Toth, T. Aach, V. Metzler: Bayesian spatio-temporal motion detection under varying illumination European Signal Processing Conference (EUSIPCO), Tampere, Finland, 4.-8.9.2000 (M. Gabbouj, P. Kuosmanen, eds.), 2081-2084.
14. R. Mester, M. Mühlich: Improving Motion and Orientation Estimation Using an Equilibrated Total Least Squares Approach. IEEE Intern. Conf. Image Proc. 2001 (ICIP 2001) , Thessaloniki, Greece, Oktober 2001.