# ADVANCES IN THE DETECTION AND RECONSTRUCTION OF *BLOTCHES* IN ARCHIVED FILM AND VIDEO[1].

*A. C. Kokaram*[2]

## Abstract

There have been several proposals for the detection and reconstruction of missing data as it applies to the restoration of archived film and video. The essence of many of the schemes is that missing data can be characterised as impulsive distortion along motion trajectories in an image sequence. This observation, coupled with a model of an image sequence leads to the design of both detection and correction schemes. This paper presents a single probabilistic framework which unifies the principal issues in Blotch treatment. A number of schemes are examined within this framework and results are shown comparing the performance of a complete application of the framework with previous schemes. The idea is to show how a variety of different approaches are unified within one probabilistic framework, and that the implementation of the framework can be made practical.

## 1  Introduction

Missing data and noise are common forms of degradation in real film and video data. Missing data manifests as dropout in the digital stream in the case of digital broadcasting. In film the problem is caused by abrasion of the film material and the effect is called "Dirt and Sparkle".

It is possible to consider the removal of this artefact as a two stage process, first detect the missing locations [4, 16] and then reconstruct the underlying image data [5] using a spatiotemporal image sequence interpolation process. The reconstruction stage may be further specified as a motion reconstruction followed by an image reconstruction stage [2]. The motion interpolation stage is a crucial step in generating useful interpolated data since in the regions of missing data motion estimates are completely unreliable. It is possible to pose the motion reconstruction and image interpolation process as a joint problem [1], but this is an interim step towards a full specification of the problem under one framework.

These problems of missing data detection, image interpolation and motion interpolation are all interlinked within the problem of blotch removal. For instance, one cannot reliably detect the extremities of blotches unless the true motion of the missing area is known. However, the true motion of the missing area is not available since the data is of course missing.

The first steps toward the full specification were introduced in [3]. That paper employed a Bayesian framework to present a joint detection and reconstruction methodology which linked the motion reconstruction as an integral part of the process.

This paper describes a generic framework for inference in the presence of missing data. This framework is then used to relate many of the schemes that have been previously presented for blotch treatment.

The following sections illustrate the broad concepts, introducing the various priors employed for the unknowns. The book [8] contains an exhaustive discussion of the joint detection and interpolation scheme. The paper concludes by emphasising the important aspects of Blotch treatment which are revealed by the framework and in so doing points the way for future development.

## 2  A model for degradation

The observed degraded sequence may be modelled by a switching process such that the data that is the corruption is switched into the original clean data at specific sites. This switching process is monitored by the field

(or image) $b(\mathbf{x})$. A corrupted site in $b(\mathbf{x})$ is set to 1, otherwise the site is 0. The model for the observed, corrupted image $G_n(\mathbf{x})$ is thus

$$G_n(\mathbf{x}) = (1 - b(\mathbf{x}))I_n(\mathbf{x}) + b(\mathbf{x})c(\mathbf{x}) \qquad (1)$$

Where $I_n(\mathbf{x})$ is the intensity of a pixel at position $\mathbf{x}$ in the image frame $n$, and $c(\mathbf{x})$ is the corruption data that is switched in whenever $b(\mathbf{x}) = 1$.

The missing data *detection* problem is that of estimating $b(\mathbf{x})$ at each pixel site. The interpolation problem is that of estimating $I_n(\mathbf{x})$ at locations where $b(\mathbf{x}) = 1$. The replacement model was employed deterministically in [7, 6] for image sequences. This idea of employing 'replacement noise' in a Bayesian framework for image sequences was first employed by Morris and Fitzgerald [10, 12, 11].

# 3  Modelling image sequence data

It is typically assumed that the frames of an image sequence do not change much in terms of content within a single shot. This implies that each frame within a shot could be constructed by somehow *warping* the first frame in the shot. The nature of this warp depends on the motion of the camera and the objects in the scene. Assuming purely translational motion, the simplest image sequence model is therefore

$$I_n(\mathbf{x}) = I_{n-1}(\mathbf{x} + \mathbf{d}_{n,n-1}) + e(\mathbf{x}) \qquad (2)$$

Where the pixel intensity $I_n(\mathbf{x})$ in frame $n$ is predicted simply as a copy of the data in frame $n-1$ at a displaced location $\mathbf{x} + \mathbf{d}_{n,n-1}$, where $\mathbf{d}_{n,n-1}$ represents the motion of the pixel between the two frames. $e(\cdot)$ is a sample of a white Gaussian noise process which could be used to account for small deviations from this model.

In fact, this model is a special case of a much more general model; the 3D Autoregressive process. This models the image sequence as follows.

$$I_n(\mathbf{x}) = \sum_{k=1}^{P} a_k I_{n+q_k^t}(\mathbf{x} + \mathbf{q}_k^s + \mathbf{d}_{n,n+q_k^t}) + e(\mathbf{x}) \qquad (3)$$

Where the pixel intensity $I_n(\mathbf{x})$ in frame $n$ is predicted by a linear combination of $P$ pixels in a spatio-temporal neighbourhood around $I_n(\mathbf{x})$. The geometry of this 'support' is defined by the $P$ offset vectors $\mathbf{q}_k = [\mathbf{q}_k^s, q_k^t]$ where $\mathbf{q}_k^s$ is the spatial component of the vector and $q_k^t$ is the temporal component. The linear coefficients are the $a_k$, and $e(\mathbf{x}) \sim \mathcal{N}(0, \sigma_e^2)$ is an excitation or residual sequence. The motion offset between frames $n$ and the frame in which the required support pixel exists, is given by $\mathbf{d}_{n,n+q_k^t}$.

Setting $P = 1$, $\mathbf{q}_1 = [0, 0, -1]$ and $a_1 = 1.0$, yields the simple model of equation 2.

AR models are used to good effect in speech processing. In image sequence analysis it could be expected that AR processes are a good model of 'texture'. However the need for *temporally* evolving textures is limited and it is unclear whether using 3D AR models gives any advantage other than optimal interpolation in the case of fraction motion of objects. Nevertheless it is important that optimal interpolation is possible with this model.

# 4  Stating the problem

The task is ultimately to estimate $I_n$ given $G_n$. In other words, the problem is only to estimate the unknown data at the pixel sites that are missing. This implies revealing the hidden variables: motion, $\mathbf{d}_{n,n-1}(\cdot)$, the AR parameters, and the 'detection' field $b(\cdot)$. A Bayesian approach implies manipulating the joint posterior $p(I_n, I_{n-1}, I_{n+1}, G_n, \mathbf{d}, b, \mathbf{d})$ (for instance) to yield $p(I_n, b, \mathbf{d}, \theta_a | I_{n-1}, I_{n+1}, G_n)$. This is done through

$$p(I_n, b, \mathbf{d}, \theta_a | \cdot) \propto p(G_n | I_n, b, c) p(I_n | I_{n-1}, I_{n+1}, \theta_a)$$
$$\times p(b) p(\mathbf{d}) p(c) p(\mathbf{a}(\mathbf{x})) \qquad (4)$$

where $\theta_a$ represents a vector of parameters for the autoregressive process.

Note that an implicit assumption is that the data is never corrupted in the same location in consecutive frames, therefore it is reasonable to assume that $I_{n-1}$ and $I_{n+1}$ are available at the required locations.

The unknowns may then be identified as those which maximize that conditional expression resulting in what will be called *maximal* estimates (encompassing both Maximum Likelihood (ML) or Maximum-a-Posteriori (MAP) estimates). Alternatively random samples for the unknowns can be generated and used as *typical* estimates. Least Squares or MMSE estimates can be related to the ML or MAP solutions.

The next sections show how various functional forms for each of these expressions can be derived, and in particular illustrates how the choice of prior distributions $p(b)$ $p(\mathbf{d})$ $p(c)$ $p(\mathbf{a}(\mathbf{x}))$ affects the overall solution.

# 5  The Corruption likelihood

The first distribution on the right hand side of equation 4 is derived from the model for degradation stated in equation 1. Thus

$$p(G_n(\mathbf{x}) | I_n(\mathbf{x}), c(\mathbf{x}), b(\mathbf{x})) \propto$$
$$\delta(G_n(\mathbf{x}) - (1 - b(\mathbf{x}))I_n(\mathbf{x}) - b(\mathbf{x})c(\mathbf{x})) \qquad (5)$$

The delta function is a direct consequence of the switching behaviour of the corruption process.

## 5.1 The original (clean) data likelihood

The second distribution in equation 4 is the likelihood of the original, clean image data at position x, given its surrounding clean image data. It is derived directly from the model statement of equation 3.

Using an AR process, the model coefficients, a(x) are estimated in a blockwise fashion, thus the allowance for non-stationarity is on a block grid and not on a pixel grid. The likelihood can be written as

$$p(i|a(x), \sigma_e^2, d(x), I_{n-1}, I_{n+1}) \propto$$

$$\frac{1}{\sqrt{2\pi\sigma_e^2}^N} \exp\left(-\frac{i^T A^T A i}{2\sigma_e^2}\right) \quad (6)$$

where A is a matrix of coefficients, and i is a vector of raster scanned image data inside a block. See [8] for more details.

Using the simpler, translation only model, the likelihood is written simply as

$$p(I_n(x)|\cdot) \propto$$

$$\exp - \left(\frac{(I_n(x) - I_{n-1}(x'))^2 + (I_n(x) - I_{n+1}(x'))^2}{2\sigma_e^2}\right) \quad (7)$$

where x' is motion compensated.

## 6 The priors

The remaining distributions encode the prior belief about the values of the various unknowns. For simplicity, a uniform prior is assigned to a. This encodes a notion of no bias on the estimation of the model coefficients, and effectively removes $p(a(x))$ from equation 4. The variance $\sigma_e^2$ is assigned a non-informative prior $p(\sigma_e^2) \propto 1/\sigma_e^2$, following [14].

### 6.1 The motion prior

The motion field is typically smooth over small image patches. This smoothness constraint is violated at motion boundaries typically occurring at the edges of moving objects. In practice for this application, it is sufficient to encode the notion of smoothness in order to achieve implicit motion interpolation.

The prior adopted for motion smoothness is as introduced by Konrad and Dubois [9] and Stiller [15]. To reduce the complexity of the final solution the motion field is block based, with one motion vector being employed for each specified block in the image.

The prior for $d_{n,n-1}(x)$, the motion vector mapping the pixel at x in frame $n$ into frame $n-1$, is as follows:

$$p_d(d_{n,n-1}(x)|S_n(x)) \propto$$

$$\exp - \left(\sum_{v \in S_n(x)} \lambda(v)[d_{n,n-1}(x) - v]^2\right) \quad (8)$$

where v is each vector in the neighbourhood represented by $S_n(x)$, and $\lambda(v)$ is the weight associated with each clique. An 8-nearest neighbourhood is used.

### 6.2 The priors for corruption and detection

In practice, each region of missing data tends to have fairly constant intensity, therefore similar priors are used on both the binary field $b(x)$ and the blotch value field $c(x)$. In addition, edge information is incorporated into these priors through the use of a zero-crossing edge detector which emphasises that edges in $c$, $b$ must correspond to the extremities of a blotch. See [8] for more information.

## 7 Solutions and Relationships

It is possible to derive many systems for blotch treatment through different combinations of priors as well as different treatment of equation 4. Recall that the idea is to solve for the unknowns given the observed images by maximizing or exploring equation 4. Only the broad concepts of various solutions are outlined here.

### 7.1 BBC Systems (1982)

Using the image model in equation 2, deleting any spatial priors for the blotches and corruption, and assuming the motion is known; yields a motion compensated version of the Dirt Detector proposed by Richard Storey[3] [16]. The prior for blotches used is $\exp(-\lambda[1 - b(x)\text{sgn}(f)\text{sgn}(b)])$. Where $f$, $b$ are motion compensated forward and backward dfd's respectively. The solution is then to examine each pixel, evaluate $p(b(x) = 0|\cdot)$, $p(b(x) = 1|\cdot)$ and choose the decision of higher probability. In effect, $\lambda$ will act (approximately) as a threshold on the forward and backward dfd's; and the prior encourages detection only when the sign of these dfd's is the same.

---

[3] Perhaps the first to design and implement a system for Dirt detection and removal, albeit non-motion compensated.

## 7.2 MCMC approaches (1994-2000)

The most all encompassing solution solves for $d, c, b, I_n$ all at the same time. The most direct way to do this is to deploy an MCMC strategy like the Gibbs sampler (Kokaram [8]). In this process, samples for each variable are drawn in turn keeping all other variables at current values. Thus the process cycles through $c \sim p(c|\cdot)$, $b \sim p(b|\cdot)$ etc. There are various issues to be considered in this solution with respect to the design of schemes for computationally effective sampling, these are explored in [8]. This full solution is called JOMBADI: JOint Model BAsed Detection and Interpolation for missing data.

Iterations are cheapest when the sampler operates on a *pixel by pixel* basis. In that situation, draws for all the variables (excepting the AR process parameters) reduce to simple multiplications and additions only. To reduce computation further, there is the interesting notion that the search for the unknowns could be restricted over a limited space. For instance, it could be assumed that the correct motion vector is actually one which is just nearby in space. Therefore 8 nearest neighbour vectors are used as candidates for solving for the motion at the current site.

The pixel based distributions required for solving for $b, i, c$ can be derived by integrating the posterior to yield:

$$p(b(x)|B,\dots) =$$
$$\begin{cases} p_c(c = g_n|\dots)p_b(b = 1|B) & \text{for } b(x) = 1 \\ p_i(i = g_n|\dots)p_b(b = 0|B) & \text{for } b(x) = 0 \end{cases} \quad (9)$$

$$p(i_n(x)|b(x),\dots) =$$
$$\begin{cases} \mathcal{N}(\hat{i}, \frac{\sigma^2}{a_u^T a_u}) & \text{for } b(x) = 1 \\ \delta(g_n(x) - i_n(x)) & \text{for } b(x) = 0 \end{cases} \quad (10)$$

$$p(c(x)|b(x), i_n(x),\dots) =$$
$$\begin{cases} \delta(g_n(x) - c(x)) & \text{for } b(x) = 1 \\ p_c(c(x)|C) & \text{for } b(x) = 0 \end{cases} \quad (11)$$

where $B$ is the set of neighbourhood of detection indicators $b(\cdot)$ surrounding the sampled location, and $a_u$ is a vector of AR coefficients. See [8] for details of the other distributions.

Morris [4] (1994) used a data likelihood approximately similar to equation 7 and also separated the problems into detection followed by removal. There was no use of $p(c)$, but the spatial part of $p(b)$ in the prior was augmented with $p(b) \propto p_b(b_{n-1})\exp(-\alpha b_{n-1})$. The idea was to introduce two indicator fields $b_{n-1}$, $b_{n+1}$ between the frames $n$, $n-1$ and $n$, $n+1$. This in a way was an occlusion detector. Blotches were cited when both the forward and backward indicator fields were set to 1 at a particular

site. At blotched sites, a simple averaging process was used for interpolation, driven in effect by the very same likelihood expression 7.

## 7.3 Bayesian Refinement 1998-2000

In order to reduce complexity further, it is possible to restrict the operation of a stochastic scheme like the Gibbs sampler or ICM, to a few *options* for the unknowns. For instance, one could use a simple, deterministic blotch detection scheme like SDIa, SDIp ([8]) to generate a superset of sites which are missing. Then within that limited set of sites, a more complete MCMC strategy could be used to solve for motion as well. In effect this limits the application of the more complex models only to those areas where it is certainly necessary. This is the kind of approach used in JOMBADI to improve the starting point for the Gibbs sampler and to reduce dramatically the number of sites visited.

A good example of this is the work by Roosmalen et al. [13]. Here, it is assumed from the outset that cut-and-paste from surrounding frames will give a good interpolation most of the time. Therefore, one could restrict the search of interpolants to this set of values and thus both reduce the computational load, and also guarantees a good reconstruction.

## 8 Some Results

Results comparing interpolation strategies can be found at www.mee.tcd.ie\~ack, as the reproduction quality is not expected to be good enough in the printed proceedings. Figure 1 shows a comparison of detection performance with several detectors applied to a real degraded image. Only the degraded image is shown in the top left quadrant, the sequence shows small vertical motion.

What is immediately apparent is that the detectors which use little or no prior information about blotches, e.g. SDIp, ROD perform the worst, with many false alarms at a scale of one pixel. These false alarms occur mainly where there are moving edges, for instance in the background of the moving leaves and the hand and flower in the foreground. At such edges, the motion field is not well defined and so apparent temporal discontinuities are introduced.

The Morris detector does incorporate spatial information about blotches, by encouraging the detection of large mainly flat areas which represent temporal discontinuities. It clearly performs better by being able to reject most of the false alarms. However, there is still a problem in the motion of the narrow petals of the flower.

JOMBADI, however solves for all unknowns and uses spatial priors on both the motion field and the detection indicator variables, $b(x)$. It therefore is able to reject all false alarms and also detects blotches which are too low contrast to excite the Morris detector.

As far as computation/performance tradeoff is concerned however, the Morris detector seems to be the best compromise. The principally temporal SDIp and ROD are too sensitive to poor motion estimation, and the JOMBADI algorithm is perhaps too heavy handed in this case of relatively low corruption level. A combination of JOMBADI and Morris would work very well here.

# 9 Final Comments

This paper has introduced a framework for the treatment of blotches and considered the relationships between a number of processes that have previously been proposed. It is to be noted however that all the expressions and models proposed here are based on pixel manipulation over a small number of frames. For severely degraded sequences, more temporal information is certainly necessary. This can be achieved by extending the temporal window of the motion priors, as well as turning to Object based analysis of image sequences.

What is interesting however, is that all the processes shown here operate successfully only if the underlying image model is correct e.g. equation 2. This implies that the motion of the objects in the sequence must be well understood. In practice, for interesting sequences typically encountered in motion picture film, there is always some motion which cannot be modelled. Typical examples of this would be motion of textiles (clothes), water, snow, fast/complicated motion. When this occurs, the results are generally catastrophic.

Current work is being directed into these very challenging areas in the EU project BRAVA (Broadcast Restoration of Archives through Video Analysis). The project coordinated by Institut National L'Audiovisuel, Paris; involves 6 partners including EEE Dept., Trinity College, Dublin; Delft University of Technology, Netherlands; Snell and Wilcox UK, RTP Lisbon, and SGT Paris.

# References

[1] S. Godsill and A. Kokaram. Joint interpolation, motion and parameter estimation for degraded image sequences with missing data. In *Signal Processing VIII*, volume I, pages 1–4, September 1996.

[2] A. Kokaram and S. Godsill. A system for reconstruction of missing data in image sequences using sampled

3D AR models and MRF motion priors. In *European Conference on Computer Vision 1996*, pages 613–624. Springer-Verlag, April 1996.

[3] A. Kokaram and S. Godsill. Joint detection, interpolation, motion and parameter estimation for image sequences with missing data. In *IEEE International Conference on Image Processing*, pages 191–194. IEEE, October 1997.

[4] A. Kokaram, R. Morris, W. Fitzgerald, and P. Rayner. Detection of missing data in image sequences. *IEEE Image Processing*, pages 1496–1508, November 1995.

[5] A. Kokaram, R. Morris, W. Fitzgerald, and P. Rayner. Interpolation of missing data in image sequences. *IEEE Image Processing*, pages 1509–1519, November 1995.

[6] A. Kokaram and P. Rayner. A system for the removal of impulsive noise in image sequences. In *SPIE Visual Communications and Image Processing*, pages 322–331, November 1992.

[7] A. C. Kokaram. *Motion Picture Restoration*. PhD thesis, Cambridge University, England, May 1993.

[8] A. C. Kokaram. *Motion Picture Restoration: Digital Algorithms for Artefact Suppression in Degraded Motion Picture Film and Video*. Springer Verlag, 1998.

[9] J. Konrad and E. Dubois. Bayesian estimation of motion vector fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(9), September 1992.

[10] R. D. Morris and W. J. Fitzgerald. Detection and correction of speckle degradation in image sequences using a 3D markov random field. In *Proceedings International Conference on Image Processing : Theory and Applications (IPTA '93)*. Elsevier, June 1993.

[11] R. D. Morris and W. J. Fitzgerald. Replacement noise in image sequences, detection and interpolation by motion field segmentation. In *IEEE International Conference on Acoustics and Signal Processing (ICASSP)*, 1994.

[12] R. D. Morris and W. J. Fitzgerald. Stochastic and deterministic methods in motion picture restoration. In *Proceedings International Workshop on Image Processing*, June 1994.

[13] P. V. M. Roosmalen, A. Kokaram, and J. Biemond. Fast high quality interpolation of missing data in image sequences using a controlled pasting scheme. In *IEEE Conference on Acoustics Speech and Signal Processing (ICASSP '99)*, volume IMDSP 1.2, pages 3105–3108, March 1999.

[14] J. J. O Ruanaidh and W. J. Fitzgerald. *Numerical Bayesian Methods Applied to Signal Processing*. Springer-Verlag, Springer Series in Statistics and Computing, 1996.

[15] C. Stiller. Motion-estimation for coding of moving video at 8kbit/sec with gibbs modelled vectorfield smoothing. In *SPIE VCIP.*, volume 1360, pages 468–476, 1990.

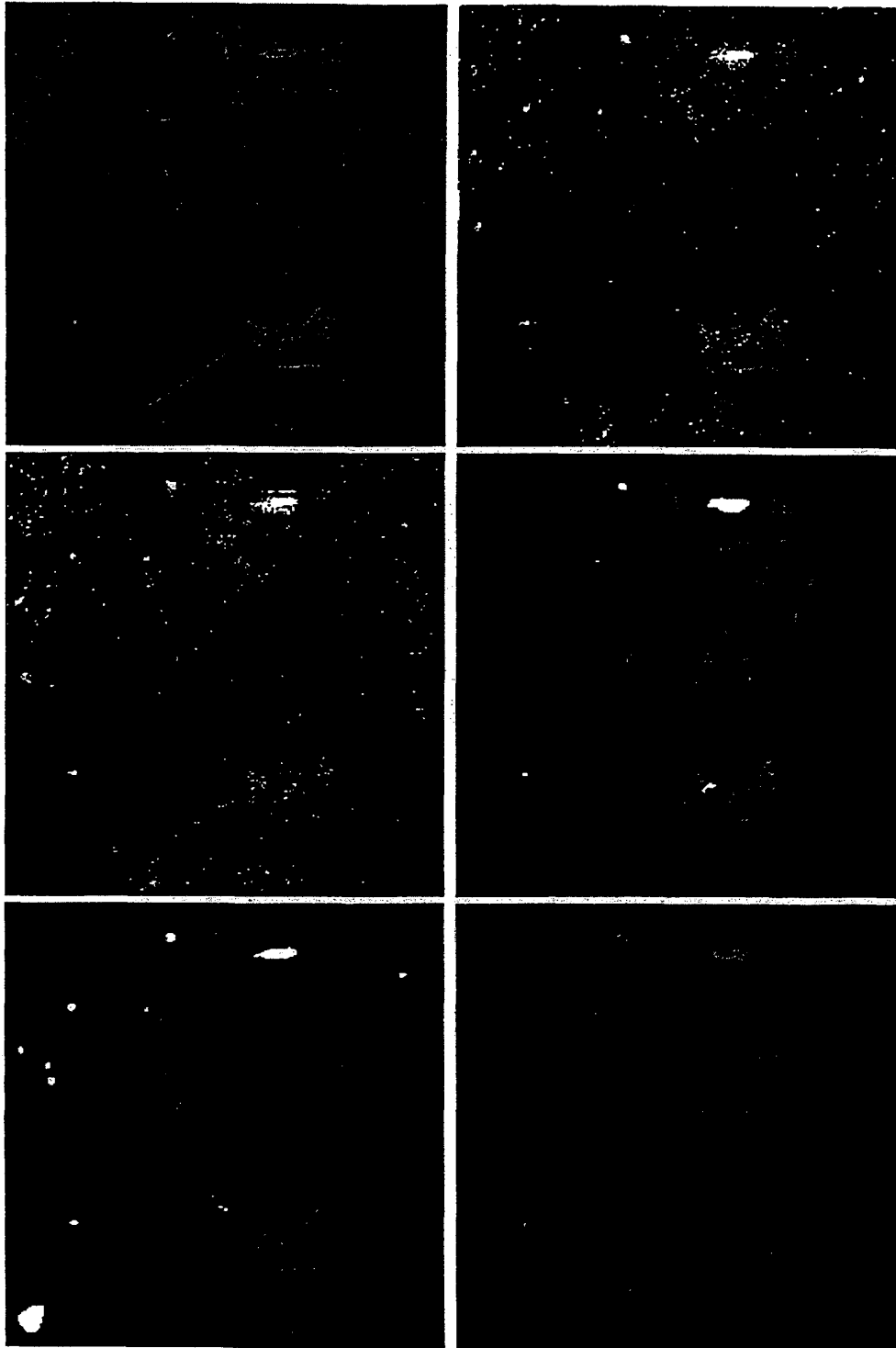[16] R. Storey. Electronic detection and concealment of film dirt. *SMPTE Journal*, pages 642–647, June 1985.

Figure 1: Top row: Real Degraded Frame, Detection with ROD, Middle Row: Detection with SDIp, Morris. Bottom Row: Detection with JOMBADI, Estimated c(x)