

Videre: Journal of Computer Vision Research

Quarterly Journal

Summer 1999, Volume 1, Number 3

The MIT Press

Article 3

Automatic Restoration Algorithms for 35 mm Film

**Peter Schallauer
Axel Pinz
Werner Haas**

Videre: Journal of Computer Vision Research (ISSN 1089-2788) is a quarterly journal published electronically on the Internet by The MIT Press, Cambridge, Massachusetts, 02142. Subscriptions and address changes should be addressed to MIT Press Journals, Five Cambridge Center, Cambridge, MA 02142; phone: (617) 253-2889; fax: (617) 577-1545; e-mail: journals-orders@mit.edu. Subscription rates are: Individuals \$30.00, Institutions \$125.00. Canadians add additional 7% GST. Prices subject to change without notice.

Subscribers are licensed to use journal articles in a variety of ways, limited only as required to insure fair attribution to authors and the Journal, and to prohibit use in a competing commercial product. See the Journals World Wide Web site for further details. Address inquiries to the Subsidiary Rights Manager, MIT Press Journals, Five Cambridge Center, Cambridge, MA 02142; phone: (617) 253-2864; fax: (617) 258-5028; e-mail: journals-rights@mit.edu.

Automatic Restoration Algorithms for 35 mm Film

Peter Schallauer,¹ Axel Pinz,²
Werner Haas¹

Classical film restoration is based on special copying machines to improve the quality of safekeeping copies. Only a small class of defects can be removed by such a process, because the unit of manipulation is always the physical film strip. With help of digital image-processing techniques, the restoration process can be adapted for each frame or even pixel and also for many kinds of defects.

This paper presents algorithms for automatic and semiautomatic defect detection and removal in digital image sequences. First, an overview about typical defects in the 35 mm film domain is given. Then robust, full automatic algorithms to detect dust, dirt, and image vibration in noisy and disturbed environments are shown. To make possible expert user interaction with intermediate analysis data, detection and removal of defects are separated into a two-step approach. Automatic algorithms for dust and dirt removal and for image-sequence stabilization are presented as well as a semiautomatic algorithm for flicker and mold reduction. Finally, the algorithms are demonstrated and evaluated using several short film sequences.

Keywords: Digital film restoration, global motion estimation, dust, dirt, jitters, mold, flicker, local color variation, global color variation, 35 mm film

1. Joanneum Research, Institute of Information Systems & Information Management, Steyrergasse 17, A-8010 Graz—Austria. {schallauer,haas}@joanneum.ac.at

2. Tu-Graz, Institute for Computer Graphics and Vision, Münzgrabenstraße 11, A-8010 Graz—Austria. pinz@icg.tu-graz.ac.at

Copyright © 1999
Massachusetts Institute of Technology
mitpress.mit.edu/videre.html

1 Introduction

Two main reasons exist for film restoration. First, all the material in the film and TV archives (approximately 2.2 billion meters of film (UNESCO, 1992)) represents a record of the history of the artistic and cultural development of all aspects of life in this century. This historic treasure must be preserved. In order to do so, original films have to be recorded on time-proof media (modern films or digital supports), and, in order to keep a copy as close as possible to the original, a restoration phase is necessary.

The second reason is economic. The communication market is developing very fast, thanks to new communication media, such as multimedia, video on demand, and satellite and cable TV. This growth opens a new market for film and TV archives, but without a restoration phase, these archives are only partly useable. Thus, preservation as well as restoration is a key factor for film and TV archives.

Classical film restoration is based on special copying machines to improve the quality of safekeeping copies. Only a small class of defects can be removed with such a process, because the unit of manipulation is always the physical film strip. Classical film restoration equipment can handle a few defects, i.e., limited dust and dirt removal, limited scratch removal, and global color-shift reduction. The current advantage of classical film restoration is the low total cost of restoration (copying) per meter of film.

With help of digital image-processing techniques, the restoration process can be adapted for each frame—and even for each pixel—and also for each defect class. Defects that can be handled by digital film-restoration techniques are for example dust and dirt removal, noise reduction, image stabilization, scratch removal, flickering reduction, local color-shift reduction, and missing frame reconstruction.

Several current manual and semiautomatic systems for digital film restoration are the Cineon system by Kodak, the Domino system by Quantel, the Flame system by Discreet Logic, and the Matador system by Avid. Their need for a high human involvement during the restoration process leads to high costs per meter of restored film. The algorithms presented in this paper are fully automatic or semiautomatic and thus they are more suitable for cost-effective film restoration.

It has been shown that motion information is essential for the automatic detection of many defects as well as for the automatic restoration of image sequences. An overview on motion analysis and applications on image-sequence processing is given in [1] and [2]. A contribution to object tracking during an entire scene can be found in [3] and [16], which also present a robust temporal integration technique.

A Markov random-field-based motion-field segmentation technique is used in [17] and [24] to detect and remove degradations such as dirt,

scratches, and fingerprints. A 3-D autoregressive model is presented in [18] to restore dirt and sparkle areas.

A group in Switzerland (University of Basel and Federal Institute of Technology Zurich) is dealing with the restoration and reconstruction of faded color photographs. The basic idea builds on the introduction of a mathematical bleaching model [8] and uses electronic imaging methods to reconstruct the bleached originals [9]. The work has lately been extended to restoration of old movie films, the first results of which are presented in [10].

Automatic restoration demands automatic cut and scene detection within the film material. Some research has already been done: Cornell University has introduced a new algorithm, producing acceptable results (see [14, 15]). Additional research from the video-processing domain is given in [25] and [26]. In [27], four automatic techniques for detection of cuts are compared.

In 1995, the AURORA (AUtomed Restoration of ORiginal film and video Archives) project launched in the European Commissions ACTS-programme. The project's aim is to develop a method for real-time video restoration.

Following, we describe a complete, digital, film-restoration system developed by Joanneum Research (Section 2). This paper focuses on algorithms for defect detection and removal. An overview of common defects is given in Section 3, and the subsequent sections describe in detail modules for one-frame defect detection and removal (Section 4), for image vibrations detection and stabilization (Section 5), and for temporal brightness and color variation reduction (Section 6).

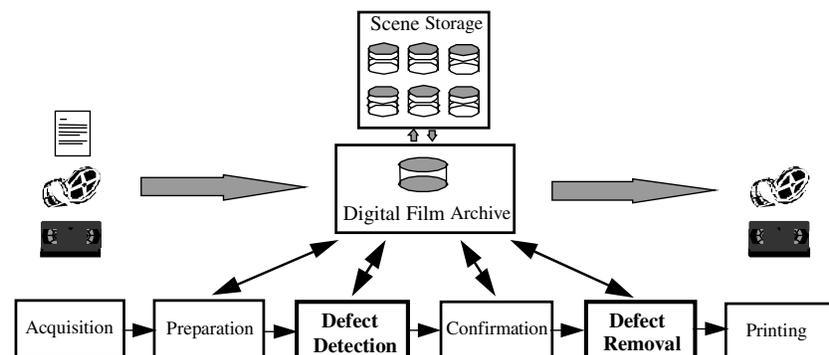
After presenting detailed experimental results, we conclude that the algorithms presented in this paper can be successfully used in a modular manner to obtain satisfactory results with minimal human operator interaction.

2 Restoration System

The digital restoration process includes everything from scanning the film, processing the image sequence, and printing the restored sequence. (See Figure 1.) This process should be automated as far as possible.

In the acquisition process, information is transformed from the analog to the digital domain. The acquired data consist of the image sequence itself, technical attributes of the scanned film (e.g., scanning parameter, calibration data), and art-historical annotations (background information like director, copyright owner, where the film was found, etc.). All these data are stored in a digital film archive (DFA), which has a purpose and functionality similar to a conventional film archive. The DFA

Figure 1. Digital restoration system.



is realized by a powerful multimedia database, together with a storage management system for large binary objects [5, 7]. Every step of the restoration process is recorded in a log (which is also stored in the DFA) in order to provide a transparent processing history for the operator.

In the preparation phase, a first semiautomatic analysis of the image sequence is obtained. This analysis includes generating an image pyramid for each frame, partitioning the image sequence into scenes (the recognition of cuts), and detecting massive frame defects. Scenes are the units for all further analysis steps. An expert is aided by the analysis software in deciding which classes of defects are analyzed for a distinct scene.

The detection of defects is the central step in the restoration process. Each scene is investigated with respect to local motion, global motion, and the defects that should be eliminated (like dust, dirt, or image vibrations). The results of this analysis are stored in the DFA and are suggested to the expert, who can follow the suggestions or can correct analysis data before the defects are removed. Defect removal depends on the defect class. From an algorithmic point of view, removing defects is an easy task in comparison to detecting them. The desired scenes of an image sequence are printed again on film or stored on a digital mass-storage medium.

The remainder of this paper deals with the tasks of defect detection and removal as emphasized in Figure 1.

3 Defects Overview

The required algorithms for analysis and restoration of image sequences depend on the kind of defect.

- **Dust and dirt** (Figure 2) are the most frequent defects in historical films. Possible reasons for dust and dirt are an improper storage environment or pollution during the duplication process. One significant characteristic of all these defects is that the pollution is local on the film material; thus, they occur in only one frame of the image sequence. The visible effect in a movie is the appearance of bright or dark spots. (See MPEG-1.)
- **Image vibrations** originate in the limited mechanical accuracy of film-transporting systems in movie cameras or duplication equipment and in the unstable camera attachment during the shooting. Image vibrations are superimposed on regular camera motions like pan, tilt, zoom, or rotation. (See MPEG-2.)
- **Flicker** (Figure 3) is visible global brightness or color variation from one frame to the next. (“Global” means that the variation is homogeneous inside a frame.) Flicker is typical for historical black-and-white film, and, in this case, it is caused by the irregular exposure time

Figure 2. Appearances of one-frame defects.



Figure 3. Global brightness variation caused by early movie cameras (flicker).

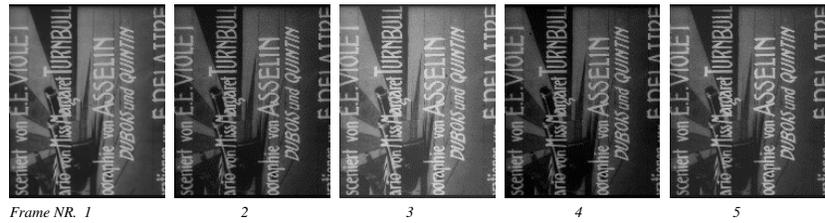


Figure 4. Local brightness variation caused by mold.

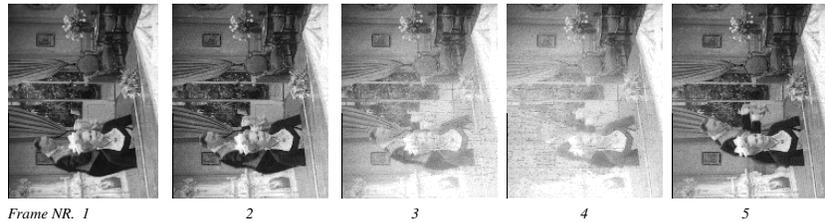
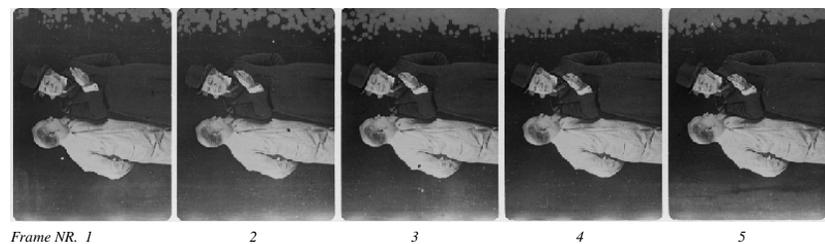


Figure 5. Local discontinuous defects caused by mold.



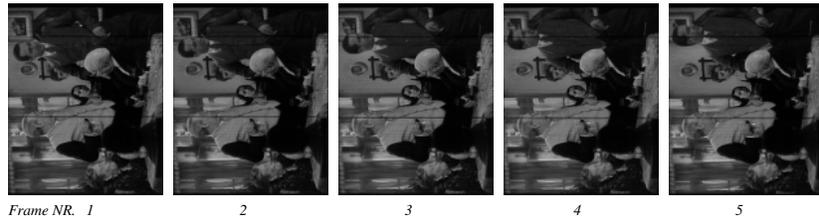
of early movie cameras (MPEG-3). Flicker can be caused in modern cameras by interferences between the lighting and the exposure during the shooting.

- **Mold** results from an improper storage environment for the film material. Very often the brightness or color variation caused by mold appears periodically every twenty to thirty frames, depending on the inhomogeneous storage environment (e.g., long-term influence of moisture from one side of the film reel). Different chemical reactions cause different types of mold: lightening by mold (Figure 4), darkening by mold, and local discontinuous defects by mold (Figure 5).

Mold causes local brightness or color variations over a few frames of an image sequence. (“Local” in this context means that the variation is inhomogeneous inside a frame.) (See MPEG-4.)

- **Scratches** (Figure 6) appear in the direction of the film strip over more than one frame of the film, and are caused by film transport or by the developing process, when there are particles in the developer’s machine. Scratches caused by film transport are exactly parallel to the film-strip perforation, but scratches caused by particles in the developer’s machine can change their position up to 5% of the film width during a few frames of the image sequence. (See MPEG-5.)
- **Heavily disturbed or missing frames** have their origin in the long storage time or the improper storage environment of the film material.
- **Captions and logos** appear on a whole scene of the movie and are created in the duplication step.

Figure 6. Scratches caused by film transport.



4 One-Frame Defects

4.1 Detection of One-Frame Defects

As can be seen from Figure 2, dust and dirt can have very different appearances. Features like shape, brightness, or size are not significant for dust and dirt. A significant characteristic of these defects is that the pollution is local on the film material; thus, they occur only in one frame of the image sequence. We call them *one-frame defects*.

Such defects can be detected automatically by observing the brightness along motion trajectories. (A motion trajectory for one point of a moving object is shown in Figure 7.) Following this trajectory, the object characteristics like brightness or color are rather constant from frame to frame. However, when a motion trajectory intersects a one-frame defect, the object characteristics change significantly. This property is the basis for the detection of such defects.

The detection scheme (Figure 8) is based on a three-frame algorithm that uses the two neighbors of the center frame I_k , for which the one-frame defects are detected. First, motion is compensated for in the neighboring frames (module *MC*). This corresponds to an estimation of the motion trajectories. Motion estimation is performed by a robust two-frame block-matching algorithm, which is presented in [20] and [21]. The necessary local motion warping scheme (Figure 9) is as follows:

Figure 7. One motion trajectory of a moving nondefected object.

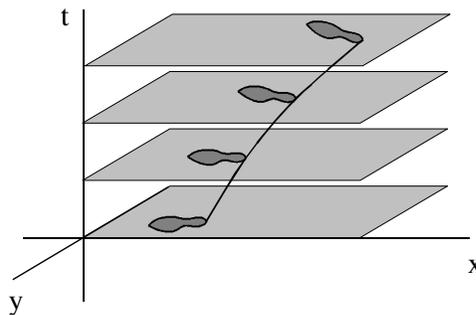


Figure 8. Detection scheme for one-frame defects.

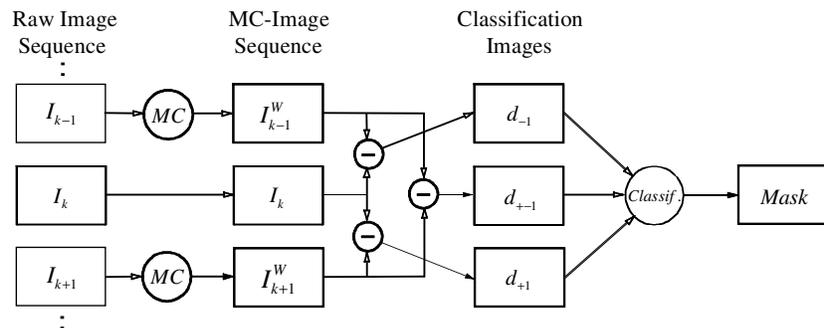
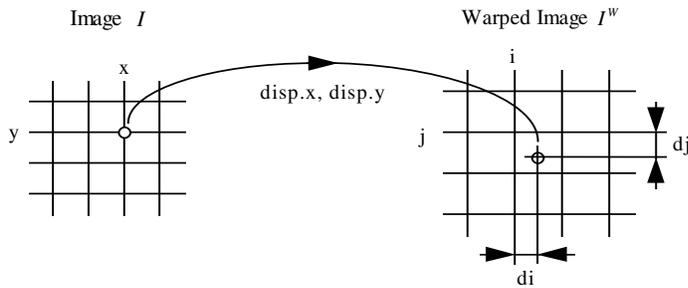


Figure 9. Local motion warping by subpixel interpolation.



Each pixel (x, y) is displaced by the row and column disparity of the estimated local motion $(disp.x, disp.y)$ and does not necessarily coincide with the sampling grid of the resulting image I^W because of the subpixel accuracy of the disparities. Interpolation between the neighboring pixels of the displaced pixel coordinate $(x + disp.x, y + disp.y)$ is performed.

The intensity value in the warped image I^W on grid coordinate (i, j) is given by

$$I^W(i, j) = \frac{\sum_{n \in A_{ij}} w_n I_n(x, y)}{\sum_{n \in A_{ij}} w_n} \quad \text{with } w_n \approx (1 - di)(1 - dj)$$

where A_{ij} is a 2×2 area centered on the (i, j) pixel in the warped image I^W , and w_n is the weight of the n th point with intensity $I_n(x, y)$ from image I .

Motion compensation of the two neighboring frames results in the motion-compensated image sequence, where the moving image parts coincide spatially with that of the center frame. This image sequence is then used to build up three classification images in the following way:

$$d_{-1}(x, y) = |I_k(x, y) - I_{k-1}^W(x, y)|$$

$$d_{+1}(x, y) = |I_k(x, y) - I_{k+1}^W(x, y)|$$

$$d_{+-1}(x, y) = |I_{k+1}^W(x, y) - I_{k-1}^W(x, y)|$$

Classification of one-frame defects is achieved pixel by pixel:

$$Mask(x, y) = d_{-1}(x, y) > th1 \quad \text{and} \quad d_{+1}(x, y) > th1 \quad \text{and} \\ d_{+-1}(x, y) < th2$$

The upper limit of the values $th1$ and $th2$ is determined by the minimum object-defect contrast that should be detected. The gray-value difference between recognizable defects in the center frame and objects in the neighboring frames was found in a range from 30 to 150, using eight-bit quantization. The lower limit of the values $th1$ and $th2$ is given by the possible gray-value change of an object along the motion trajectory caused by local lighting changes or by global brightness variations. For flicker, a frame-to-frame gray-value difference up to 15 was found. These limits determine the values $th1$ and $th2$ in a range from 15 to 30 using eight-bit quantization. Changes of $th1$ and $th2$ inside this range cause only marginal changes in the number of classified pixels.

On the classified one-frame defects, a postprocessing step is required. The real area, which is occluded by a defect, is larger than the area of the classified pixels. This is caused by the gray-value distribution of such defects. In the center area of a bright or dark spot, the gray-value

difference to the neighboring frames is significant, while, at the border, this difference decreases to zero. Therefore, the border pixels of a bright or dark spot are not classified by the method described above. To get the optimal mask size for replacement of one-frame defects, the resulting mask of the classification is enlarged by morphological dilation. It was found that the amount of dilation depends on the size of the defects. An example of the final one-frame defect mask is shown in Figure 11.

An important question is how this algorithm deals with occlusions in the image sequence. When only two consecutive frames of an image sequence are used for detecting one-frame defects, it is not possible to distinguish between defects and occluding regions because of the covered or uncovered background problem. (See [2], Chapter 5, page 78.) When using three or more consecutive frames of the image sequence, it is possible to reconstruct the whole image content of the center frame, since uncovered background can be reconstructed by image content of either the preceding or subsequent frame(s). The scheme for detecting one-frame defects sets a mask pixel only if there is no image-content similarity from the center frame I_k in forward *and* in backward directions *and* if there is an image-content similarity between the preceding frame I_{k-1} and the subsequent frame I_{k+1} . Two of these three conditions are not fulfilled for occluding areas; thus, they are suppressed in the used classification scheme of one-frame defects.

4.2 Removal of One-Frame Defects

The removal of one-frame defects has two requirements: image structure and image brightness must be estimated for the regions occluded by the defects. Missing image structures that are caused by a large defect area must be replaced. In addition, brightness must be corrected because of flickering and local brightness variations from frame to frame. Image structure estimation from the surroundings of a defect is possible only if the defect area is very small (pixel range) and therefore the probability of difficult structure in this area is low. If the defect area increases, image content of the preceding and subsequent frame must be used to estimate image structures for this region.

The first task is to estimate image structures for areas that are occluded by defects. Structure information of the preceding frame I_{k-1} and the subsequent frame I_{k+1} is used to find the image content, which is then used in frame I_k to replace dust or dirt. First, the image content of the neighboring frames must be registered with the image content of the center frame, because of image vibration and local motion. For this task, the motion-compensation (MC) scheme, presented in Section 4.1, is used.

It turns out that motion information in the region of a defect is not reliable. For the two-frame motion estimator, it is not possible to match image content of the neighbor frame with dust or dirt in the center frame. Therefore, motion disparities (motion vectors) from the neighbor frame to the center frame are pointing outside the defect; thus, the disparities are not reliable in the region of a defect. The influence of this effect grows with the size of the defect. To overcome this problem, disparities are interpolated in such areas. For this reason, a disparity mask is prepared by dilating the one-frame defect (OFD) mask with respect to the defect size. Bigger defects influence the motion disparities

in a wider range than smaller ones. Disparity interpolation for each pixel inside the area given by the disparity mask is done by an eight-direction linear interpolation of the reliable disparities outside the area given by the disparity mask.

After the disparities are corrected, the neighbor frames are registered with the center frame by applying the local motion warping scheme presented in Section 4.1, resulting in the correctly warped neighbor images I_{k-1}^W and I_{k+1}^W . These two images now contain image structures of the neighboring frames in areas where the center frame I_k is occluded by defects. A structure image is prepared by taking the average of I_{k-1}^W and I_{k+1}^W for each pixel in these images. The structure image is then used for the final replacement of defects in the center frame I_{k+1}^W .

The second requirement of an automatic algorithm for defect replacement is to estimate the brightness of replaced areas. We want to replace the defect area by motion-compensated image content of the neighbor frames (the structure image). Brightness correction is necessary because of flickering from frame to frame or local brightness variations inside a frame. Therefore, an algorithm is designed where brightness correction is done separately for each defect and in the region of the defect locally for each pixel. The task is now to fit the brightness of the structure image (which provides image content for the defect replacement) and the brightness of the defect surrounding center image I_k . When the brightness is not fitted, a gray-value discontinuity can appear between regions of replaced image content and regions of original image content. As a border of the surroundings, we choose the border indicated by the OFD mask, since defect replacement is also done for areas that are indicated by the OFD mask.

The brightness-correction image is prepared in the following manner:

First, the frame difference between the structure image and frame I_k is calculated. Then, defect areas, indicated by the OFD mask, are interpolated in this difference image. The gray-value difference between the structure image and the original I_k is estimated from the surrounding of a defect and is then interpolated to the inner part of the defect region. Interpolation is done by a scheme similar to that for disparity interpolation. The brightness-correction image is applied in the final defect-removal step.

The final replacement of one-frame defects is straightforward:

$$I_k^{Corr}(x, y) = \begin{cases} I_k(x, y) & \text{for } \{(x, y) \mid OFD(x, y) = false\} \\ \begin{aligned} &Structure\ Image(x, y) \\ &+ Brightness\ Correction(x, y) \end{aligned} & \text{for } \{(x, y) \mid OFD(x, y) = true\} \end{cases}$$

Defect areas, which are indicated by the OFD mask, are replaced by the brightness-corrected structure image; nondefect areas are taken from the original frame I_k . The final results are shown in Figure 10 and Figure 11. The upper parts show a detail in three consecutive frames of the original sequence, while, in the lower parts, defect-replaced images are visible. In Figure 10, the image-structure estimation is shown. In Figure 11, the difference between images with and without brightness correction appear mostly at the border of replaced areas, which are indicated by the OFD mask.

Figure 10. Removal of one-frame defects by image-structure estimation with the content of the neighboring frames.

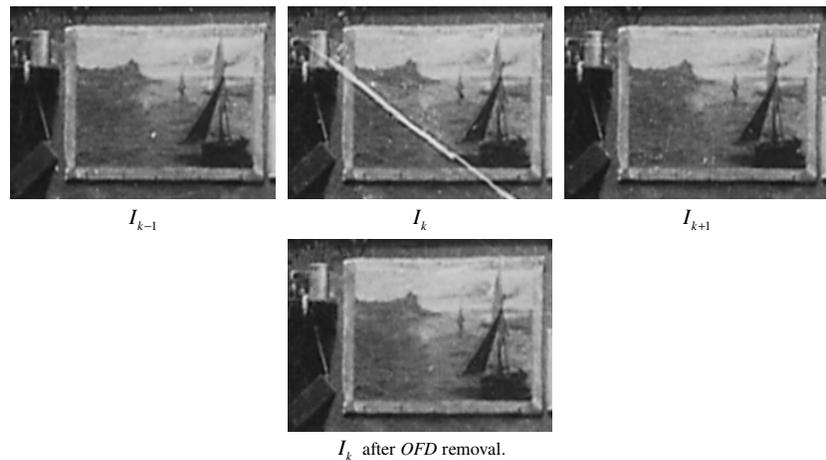
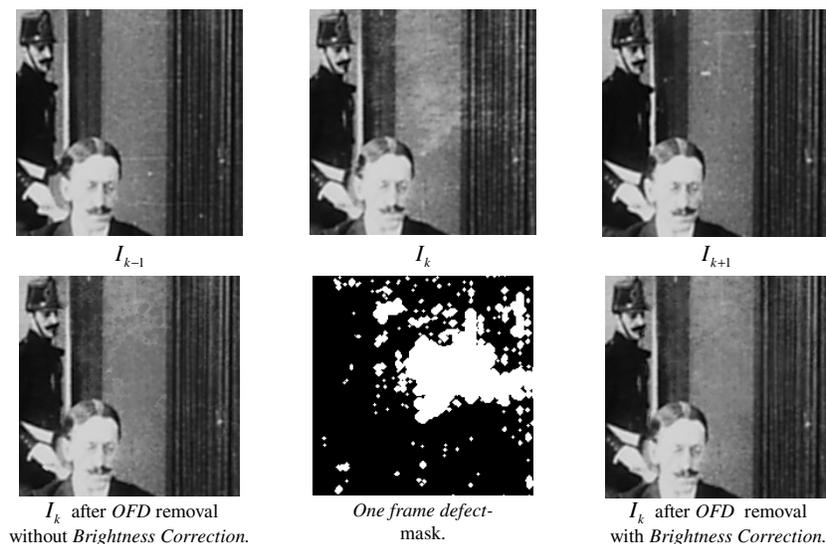


Figure 11. Detection and removal of one-frame defects. Comparison of brightness corrected and non-brightness corrected defect replacement.



5 Image Vibrations

Although image vibration is a typical defect of historical movies, it is already appearing in a reduced form in modern ones. Vibrations originate from the limited accuracy of the film transporting system in the movie camera or in the duplication machine. A second reason can be an unstable camera attachment during the shooting. The effect is that all the image content is moving from frame to frame on the image plane with a limited degree of freedom, which can be described as a 2-D motion problem. One way of describing 2-D motion involves the use of motion models, which are specified by a set of motion parameters. Motion parameters provide a very efficient means to describe 2-D motion when compared with dense motion fields (represented by one motion vector (disparity) per pixel). Our four-parameter motion model is capable of describing translation, rotation, and scaling, which fits our requirement to model pan, tilt, rotation, and zoom in an image sequence.

The basic algorithm is applied in conjunction with motion-compensated iteration and a multiresolution approach to improve the accuracy and the measuring range of the method. First, we assume that the motion parameters are the same everywhere within one frame. Then, a

local motion mask is introduced to exclude locally moving objects from the area of interest for the global parameter estimation algorithm. The result is that the dominant object with its dominant motion is tracked through an image sequence. Dominant motion includes image vibration, but camera pan, tilt, rotation, and zoom are also estimated.

Finally, image-sequence stabilization is done by filtering the dominant motion with respect to possible camera motions.

5.1 Image Vibration Detection

Two approaches can be used to estimate global motion parameters. One goes through the stage of estimating optical flow for a distinct number of points of interest, and the other one estimates the motion parameters directly from the image sequence. In the first approach, the correspondence between pixels or features in two successive frames has to be established to fit the desired motion model to the motion found for the feature points. The problem is to find good feature points, especially if image content is heavily disturbed or very noisy.

In the second approach, the correspondence problem is avoided. Motion parameters are estimated by using low-level information only. The algorithm is a differential algorithm making use of the spatial and temporal gradient of an image. In the case of defective and noisy film, this second approach is the more robust.

The global motion estimation algorithm used is proposed in [4], which defines the following four-parameter motion model:

$$\begin{aligned}x' &= a_1x - a_2y + a_3 \\y' &= a_2x + a_1y + a_4\end{aligned}\tag{1}$$

A pixel at (x, y) in frame k moves to the new position (x', y') in frame $k + 1$. The core of the parameter estimation algorithm is a differential technique based on a Taylor series expansion of the image signal resulting in a simpler first-order derivative equation.

$$\begin{aligned}I_k(x, y) - I_{k+1}(x, y) &= (a_1 - 1) (G_x(x, y)x + G_y(x, y)y) \\&\quad + a_2 (G_y(x, y)x - G_x(x, y)y) \\&\quad + a_3 G_x(x, y) \\&\quad + a_4 G_y(x, y)\end{aligned}\tag{2}$$

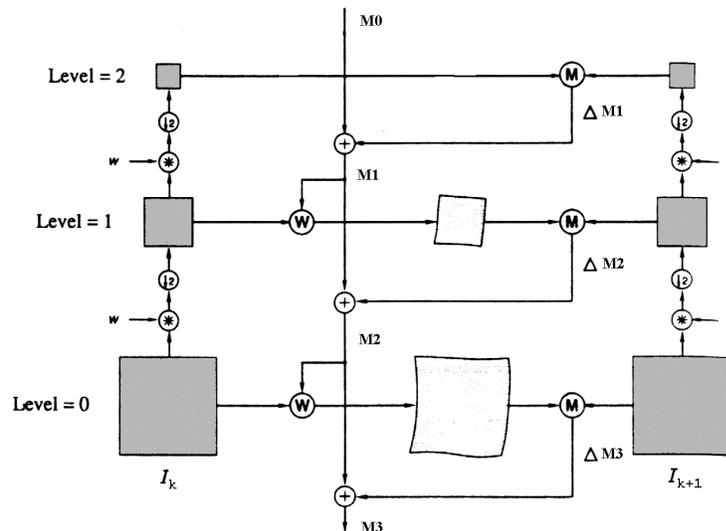
where

$$\begin{aligned}G_x(x, y) &= \frac{1}{2} \left(\frac{\partial I_{k+1}(x, y)}{\partial x} + \frac{\partial I_k(x, y)}{\partial x} \right) \\G_y(x, y) &= \frac{1}{2} \left(\frac{\partial I_{k+1}(x, y)}{\partial y} + \frac{\partial I_k(x, y)}{\partial y} \right)\end{aligned}$$

Variables a_1, \dots, a_4 are the motion parameters that we want to estimate; the rest are measurable quantities. Image gradients $G_x(x, y)$ and $G_y(x, y)$ are approximated by half the difference of the two neighboring pixels of the current pixel. Since there are four unknowns in the equation, we need at least four equations, or four pixels, to solve for the motion parameters. Usually, the algorithm is applied to a large number of pixels, and the motion parameters are estimated by linear regression.

In our implementation, we apply this algorithm to all pixels of two successive frames, except a small border to avoid possible problems. Motion parameters are estimated by a linear regression function *lfit()*,

Figure 12. Multiresolution iteration.
(Compare [6], page 890.)



adapted from *Numerical Recipes* ([22], pages 671–680). They provide linear regression by default only for the two-dimensional case and not for a multidimensional fit as in our case.

5.2 Motion-Compensated Iteration

The amount of motion must be small with respect to the image gradient for the result in Equation (2) to hold because of simplifications that were made there before in the derivation. Two approaches can be used to fulfill this condition.

The first is to suppress high-frequency parts in the images by low-pass filtering.

In the second approach, a good initial estimate must be obtainable in such a manner that the difference between an initial estimate and the true parameters is sufficiently small. To get a good initial estimate, a multiresolution iteration scheme is used. (See Figure 12.) The motion parameter estimate is first obtained from a low-resolution representation. The magnitude of motion in this case is less than in full resolution. First, a pyramid is built up for the image pair I_k and I_{k+1} . The initial motion vector $M0 = (a_1, a_2, a_3, a_4)$ is set to $(1.0, 0.0, 0.0, 0.0)$, which corresponds to no scaling, no rotation, and no translation between the image pair. The first motion vector estimate $\Delta M1$ is done at the top level of the pyramid. The initial motion vector $M0$ is now updated with $\Delta M1$ to obtain the better result $M1$. In the next-lower level (Level=1), the representation of the image I_k is warped with the parameters $M1$, so that the already found but inaccurate motion between the image pair is compensated. The following parameter estimation between the registered images, the warped image I_k and image I_{k+1} , provides only a small motion vector difference $\Delta M2$, which is used to update the motion estimate $M1$ of the previous pyramid level, resulting in a new, more accurate, motion vector $M2$. This scheme is continued until the bottom pyramid level is reached. Warping of an entire image with the motion parameters a_1, \dots, a_4 is done by calculating the displacements of the four corner points (by Equation (1)), which are then used to apply linear interpolation to all image points inside these four corner points with a standard function from [23].

During the multiresolution iteration process, an update scheme for motion parameters is used (compare [4], page 75):

The update $(\hat{a}_1, \hat{a}_2, \hat{a}_3, \hat{a}_4)$ and the initial estimate $(a_1^I, a_2^I, a_3^I, a_4^I)$ can be combined to form a better estimate $(a_1^N, a_2^N, a_3^N, a_4^N)$ by using the following equations:

$$\begin{aligned}
 a_1^N &= \hat{a}_1 a_1^I - \hat{a}_2 a_2^I \\
 a_2^N &= \hat{a}_2 a_1^I + \hat{a}_1 a_2^I \\
 a_3^N &= \hat{a}_1 a_3^I - \hat{a}_2 a_4^I + \hat{a}_3 \\
 a_4^N &= \hat{a}_2 a_3^I + \hat{a}_1 a_4^I + \hat{a}_4
 \end{aligned} \tag{3}$$

5.3 Local Motion Exclusion

In the motion-compensated iteration scheme, we assumed that the motion parameters are the same everywhere within the whole image, but in practice this assumption rarely holds. Moving objects undergo different motions. What we want to find is the object with the biggest region of support in the image (the dominant object). The task is now to prepare a local motion mask to exclude local moving objects from the region of support of the dominant object. A segmentation of the dominant object can be done in the scheme mentioned above by searching the region of support of an already found motion. This region is then used in the parameter estimation computation (2) to exclude those pixels from the estimation that are not in the region, resulting in a more accurate motion estimation for the dominant object. The region of support detection is applied in each iteration step of the pyramid algorithm; thus, the local motion mask is refined from higher levels down to the lower ones. Once a motion has been determined, we would like to identify the region belonging to the motion. (The local motion mask is given by regions not having this motion.) To simplify the problem, the two images are registered using the detected motion. The motion of the corresponding region is canceled after registration, and the tracked region is stationary in the registered images. The segmentation problem is reduced to identify stationary regions in the registered images. The absolute difference of the registered images I_k and I_{k+1} is a good base for this segmentation task.

The intensity difference caused by motion is also affected by the magnitude of the gradient in the direction of the motion. Small motion of an edge causes high gray-value difference when the motion is perpendicular to the edge. Only fast motion should be detected. Therefore, rather than using only the gray-value difference as a motion measure for classifying pixels, the gray-value difference scaled by the gradient magnitude, is used:

$$Diff_{Scaled}(x, y) = \frac{|I_k(x, y) - I_{k+1}(x, y)|}{1 + \nabla I_k(x, y)}$$

Let $I_k(x, y)$ and $I_{k+1}(x, y)$ be the intensities of pixel (x, y) of the two registered images, and let $\nabla I_k(x, y)$ be the spatial intensity gradient, which is estimated by the *sobel* operator. The constant 1 is used to avoid numerical instabilities. The local motion mask is given by:

$$Mask_{Local\ Motion}(x, y) = Smooth(Diff_{Scaled}(x, y)) > thres$$

Because of scaling by the gradient magnitude, a mask pixel is set only if the motion of a local moving object is wider than the high

gradient area of this object. Filtering is necessary because of noisy image content (which should not be detected as local motion). The threshold is given by the local-to-dominant object contrast (actually detected in the image.)

Once motion has been determined, we would like to measure the reliability of the motion parameters. Image stabilization is possible only when reliability of the already found motion parameters is high. An approach for this task is proposed in [3]. With the help of the optical flow equations, a reliability measure is derived, indicating the amount of motion between the two registered images for each pixel.

The reliability of the finally found motion parameters is given by the percentage of all image pixels that are not set in the local motion mask and for which the reliability measure is lower than 0.1 (the displacement between the registered images is lower than 0.1 pixel). Typically, the reliability is in the range from 1% (if the dominant object has few gradient areas) to 20% (for a structured dominant object).

The algorithm for detection of the dominant object described above is extended to track an object throughout long image sequences. The technique is proposed in [3] (page 9): “This is done by *temporal integration*, without assuming temporal motion constancy. For the tracked dominant object a dynamic internal representation image is constructed. This image is constructed by taking a weighted average of recent frames, registered with respect to the tracked motion (to cancel its motion). This image contains, after a few frames, a sharp image of the tracked object, and a blurred image of all the other objects. Each new frame is compared to the internal representation image of the tracked object rather than to the previous frame.”

Let I_x denote the image sequence. The internal representation image of the tracked dominant object is denoted by Av_x and is constructed as follows:

$$\begin{aligned} Av_0 &= I_0 \\ Av_{k+1} &= (1 - w) \cdot I_{k+1} + w \cdot \text{register}[Av_k, I_{k+1}] \end{aligned} \quad (4)$$

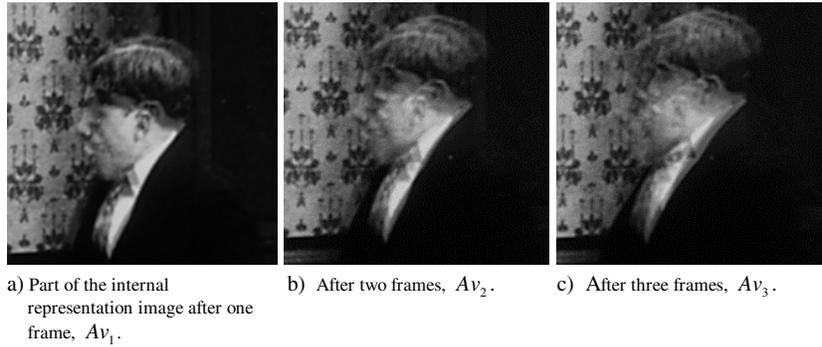
where $\text{register}[P, Q]$ denotes the registration of images P and Q according to the motion of the tracked dominant object computed between them, and $0 < w < 1$ (currently $w = 0,7$). Av_x therefore maintains sharpness of the tracked object while blurring other objects in the image. An example of the evolution of an internal representation image of the tracked dominant object is shown in Figure 13. The scene contains more objects: the dominant background object and a local moving object in the foreground. The tracked background remains sharp, while the local moving foreground object gradually blurs out. This effect enforces tracking of the already found sharp background, while local moving objects can be detected more reliably by the segmentation scheme mentioned above.

5.4 Image Vibration Stabilization

Image-vibration detection results in the vibration disparity signals. Four dominant motion parameters are always estimated between two consecutive frames in the image sequence; thus, four vibration disparity signals are generated for a whole scene. (One of them can be seen in Figure 22.)

First, the real image vibration must be found. The vibration disparity signals describe the change of vibration from one frame to the next (the derivation of vibration) through an entire scene; therefore, image

Figure 13. Evolution of the internal representation of the tracked dominant object, the background. (See MPEG-6.)



vibration is given by discrete integration of this signal. The four resulting vibration signals correspond to the real visual effect of vibration when watching a movie. The task is then to filter these vibration signals appropriately and to register the original frames with the position given by the four filtered vibration signals.

The image-vibration signal can be derived from the vibration disparity signal by integrating the vibration disparities over a whole scene by (combining the dominant motion parameters of consecutive frames by Equation (3)). In our context, the initial estimates are the four dominant motion parameters from frame I_k to I_{k+1} , the update estimates are the motion parameters from frame I_{k+1} to I_{k+2} , and the better estimates are then the combined or integrated motion parameters from frame I_k to I_{k+2} . The integrated motion parameters up to frame I_{k+2} are combined with the dominant motion parameters between I_{k+2} and I_{k+3} to get the integrated motion parameters for frame I_{k+3} , and so on. By extending this scheme from four frames to the entire scene, the image-vibration signal is prepared. (The four sequences of integrated dominant motion parameters are called the *image-vibration signals*, since these signals correspond to the real image content translation, rotation, and zoom in a movie scene. One signal is shown in Figure 23.)

After the vibration signals are prepared they must be filtered to obtain the desired correction parameters for image-sequence stabilization. Different filter types from linear low-pass filters to nonlinear filters such as the median operator are possible. As can be seen from Figure 23, image vibration is a stochastic process, where image position is different from a mean position for a number of frames, or “peaks”. The filtered signal should follow the main course of the raw signal, but peaks over a few frames should be suppressed. For this task, we are using a nonlinear median filter. The median filter size can be chosen in such a way that peaks of distinct length in the signal can be filtered. A peak of length n -frames can be filtered by a median operator with a width of approximately $2n$. For example, for filtering of a 0.25 sec. peak (six frames) in the image-vibration signal, a median filter of a width of at least 12 is required. The median filter is capable of removing vibration peaks up to a distinct length. In addition to the nonlinear median filter, a linear low-pass filter with a high cutoff frequency is applied to the signal to filter high-frequency components, which remain after median filtering.

The dominant motion includes not only image vibration but also tilt or pan from camera motion. An algorithm that is able to handle image stabilization during fast camera motion is presented in [19]. The median filter and linear filter width is estimated by the general course of the vibration signal. When the general course indicates the beginning or

end of a camera pan or tilt, the filter width is reduced to avoid smooth-filtered vibration signals.

After filtering the four vibration signals, each image in the sequence must be warped to the position given by the filtered vibration signals. The difference between the raw vibration signals and the filtered vibration signals must be compensated for by Equation (3). In this context, the initial estimates are the four vibration-signal parameters, the update estimates are the four desired update parameters, and the better estimates are the four filtered vibration-signal parameters. The task is now to find the four update parameters that, when combined with the vibration-signal parameters, give the filtered vibration-signal parameters. After transforming Equation (3),

$$\begin{aligned}\hat{a}_1 &= \frac{a_1^N a_1^I + a_2^N a_2^I}{(a_1^I)^2 - (a_2^I)^2} & \hat{a}_3 &= a_3^N - \hat{a}_1 a_3^I + \hat{a}_2 a_4^I \\ \hat{a}_2 &= \frac{a_2^N - \hat{a}_1 a_2^I}{a_1^I} & \hat{a}_4 &= a_4^N - \hat{a}_2 a_3^I - \hat{a}_1 a_4^I.\end{aligned}$$

The four update parameters ($\hat{a}_1, \hat{a}_2, \hat{a}_3, \hat{a}_4$) are calculated for each frame in the scene. Warping an entire image with these parameters is done by calculating the displacements of the four corner points (by Equation (1)), which are then used to apply linear interpolation to all image points inside these four corner points. The result is a stabilized image sequence (see MPEG-12), whose dominant motion corresponds to the filtered image-vibration signals.

6 Temporal Brightness and Color-Variation Reduction

As can be seen from the defects overview, there are two main causes for brightness and color variations, namely flicker and mold. Their various effects can be characterized by the following properties:

- **Temporal support** stands for the number of consecutive frames that must be corrected. This property varies from one frame in the case of flicker (Figure 3) up to ten frames in the case of mold (Figure 4).
- **Amount of variation:** We distinguish between *minor* and *major* variations. Minor variations occur when the histogram shift between a “correct” and a brightness-varied frame is in maximum about 10% of the full gray-value range (typical for flicker, see Figure 3). Major varying histograms of brightness-varied frames are compressed either in the range of high gray values (e.g., lightening by mold, see Figure 4) or in the range of low gray values (e.g., darkening by mold).
- **Locality of variation** is subdivided into *global* and *local* variations. Global variation is when the brightness or color variation is constant within one frame (typical for flicker, see Figure 3). Local variation is when the brightness or color variation is not constant within a frame. Within this category, we have to distinguish between a spatially continuous change of brightness or color (e.g., lightening by mold, see Figure 4) and a spatially discontinuous change of brightness or color (e.g., local defects by mold, see Figure 5).

The most common brightness and color variations are flicker and lightening by mold, but there are rare variations of darkening by mold and local defects by mold.

In principle, at least two strategies can be used to overcome the temporal brightness and color variations. The first is to segment, track, and brightness (color)-correct all objects through an entire scene. This method promises good results, but object segmentation enforces motion estimation and thus is a time-consuming task. Also, the object segmentation itself has to be very reliable in heavily disturbed image sequences, or false corrections can occur.

The second strategy, proposed here, can handle all categories mentioned above, except local defects by mold. (These variations with a spatially discontinuous change of brightness or color are partially handled by the one-frame defect detection and removal. See section 4.1 and 4.2).

Section 6.1 explains the base algorithm for correcting spatially global brightness and color variations with a temporal support of more than one frame. Section 6.2 explains an extended algorithm of Section 6.1, which can also handle spatially local brightness and color variations.

6.1 Temporal Global Brightness and Color Correction

For brightness or color correction a two-step approach was chosen. First, good frames (called reference frames) are determined, and then the gray-value distribution (in the case of monochrome images) or the three gray-value distributions (in the case of color images) of all remaining frames in the image sequence are corrected.

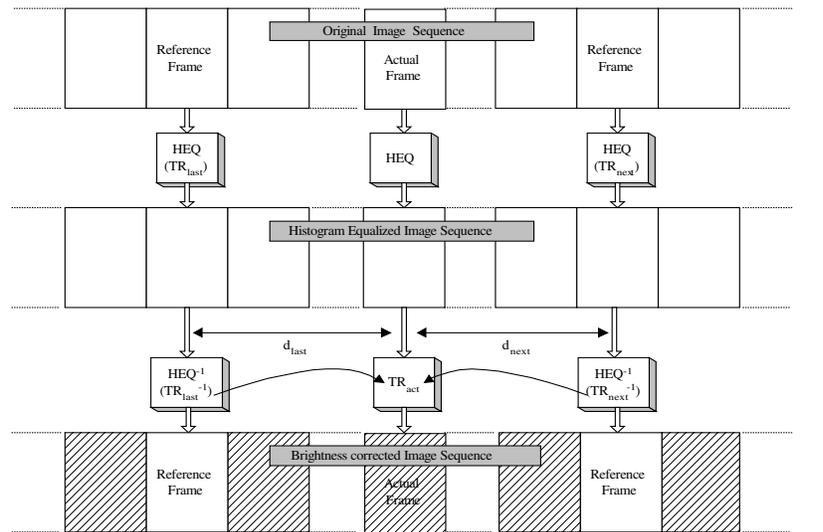
The reference frames can be detected either automatically or semiautomatically, depending on the number of defects and depending on the image content. Experiments have shown that the distance from one reference frame to the next should be in the range of five to twenty frames.

The distance between reference frames is limited by the amount of image content change. The reason for this limitation is the assumption that image content does not vary significantly between reference frames. If this condition holds, then the gray-value distributions of two consecutive reference frames are similar, and, thus, the gray-value distribution of a nonreference frame can be estimated by the gray-value distributions of the closest reference frames in the image sequence. If there is, for example, a camera pan or fast motion in the scene, the distance from one reference frame to the next should be as small as possible, to hold the condition of similar reference frames.

The main idea behind the algorithm is as follows.

- Histogram-equalized images of identical content already appear at equal brightness. All frames of the histogram-equalized image sequence in Figure 14 appear at equal brightness (and also if there are global brightness variations between the frames in the original image sequence in Figure 14).
- To achieve the characteristics of the reference frames, the inverse histogram equalization functions of the nearest reference frames should be applied. TR_{last} is the histogram equalization transformation function of the last reference frame, and TR_{next} is the same one of the next reference frame inside the scene. A linear combination of the inverse TR_{last} and TR_{next} functions is used to avoid

Figure 14. Scheme of temporal brightness and color correction. A two-step histogram transformation procedure is used. First, all frames in the image sequence are histogram equalized. Then a nonreference frame is corrected by the inverse histogram equalization function of the nearest reference frames inside the scene.



HEQ ... Histogram Equalization

HEQ⁻¹ ... Inverse Histogram Equalization

TR ... Histogram Transformation Function

d_{last} ... Distance to Last Reference Frame

d_{next} ... Distance to Next Reference Frame



... Histogram Transformation Step



... Changed Frame Histograms after Processing

$$TR_{act} = d_{last} \cdot TR_{next}^{-1} + d_{next} \cdot TR_{last}^{-1}$$

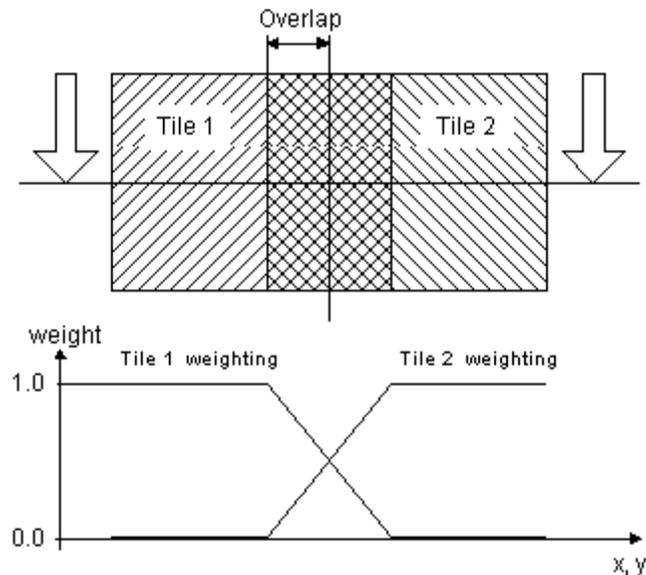
discontinuities over time. The closer the actual frame is to the last reference frame, the higher the influence of the inverse TR_{last} function. This procedure allows one to follow the changes of the gray-value distribution from one reference frame to the next, since the used histogram-transformation functions of the reference frames are weighted according to the distance from the actual processed frame. The inversion of the histogram equalization function is estimated by a standard routine from [23].

This temporal gray-value distribution correction scheme can be applied only to image sequences with global brightness or color variations, since the correction is working on the whole image area. The temporal global brightness and color correction scheme is suitable for the correction of flicker.

6.2 Temporal Local Brightness and Color Correction

The temporal global brightness and color-correction algorithm must be adapted to apply to image sequences with local varying gray-value distributions (e.g., shown by lightening by mold, see Figure 4). To handle such spatially local variations, the two-step histogram transformation procedure is first applied to small regions of the actual processed frame. The resulting gray-value distribution-corrected regions are then merged to give the whole gray-value distribution-corrected image. Temporal local brightness and color correction is accomplished by the following steps:

Figure 15. Weighting function for overlapping tiles.



- Divide actual frame and reference frames into a distinct number of overlapping tiles.
- Apply the temporal global brightness or color correction scheme to each tile.
- Merge the temporal brightness or color-corrected tiles to a resulting image.

In the division step, a small tile size (and thus a high number of tiles) would be necessary to handle spatially local brightness or color shifts, but the minimal size of the tiles depends on the amount of image content motion from the last reference frame to the next. If motion is high and the tile size small, the condition of similar gray-value distributions in the reference frame tiles and the corresponding actual frame tile is not fulfilled, and thus the temporal local brightness and color correction algorithm itself results in brightness or color shifts. To avoid this effect, a solution is to adapt the size of compared tiles automatically by the amount of motion. To do so, the image area of the actual processed frame is first divided into a fixed number of overlapping tiles (e.g., 15 in x -direction and 10 in y -direction), called the “base tiles”. Then, for each base tile, a virtual tile size is estimated. This virtual tile size is used in the temporal global brightness or color correction scheme instead of the base tile size. For regions with low local motion, the virtual tile size is as small as the base tile size, and, thus, the spatial correction of brightness or color shifts is local. Otherwise, for regions with high local motion from one reference frame to the next, the virtual tile size must be increased, and, thus, the spatial correction of brightness or color shifts can be done only in a more global style. The adaptation of the virtual tile size is controlled automatically by local motion accumulation between reference frames. The local motion estimation can be done very fast, since the necessary spatial resolution is low.

In the merging step, the overlapping base tiles must be combined. An overlap of approximately 30% of the base tile size is necessary. When the gray-value distribution variation is local in the base tile, the gray-value distribution correction is an estimation for the whole base tile size. Without overlapping base tiles, a gray-value discontinuity between

spatially neighboring tiles would occur. To avoid this, overlapping gray-value distribution-corrected base tiles are weighted according to Figure 15 when the result image is built up. Tiles are overlapping in x - and in y -directions.

This temporal gray-value distribution-correction scheme can be applied to image sequences with local brightness or color variations like lightening by mold (see Figure 4) or darkening by mold, but global variations like flicker (see Figure 3) can also be handled.

7 Results

In order to test the algorithms and gain experience in handling the process flow, several scenes were selected and restored. The restored and reprinted scenes were reviewed by film experts in order to judge the quality of the digital restoration. The test sequences were taken from the film *Le mort qui tue* (1913/14, director Louis Feuillade, from the series *Fantomas*) and from *Opernball* (1956, director E. Marischka).

For testing and viewing in video resolution, a tele-cine copy of the 35 mm film was used. A single-frame recorder (Pronto Video from DIVI) was used as the interface (buffer for a 3 min. clip) between a digital Betacam VCR and the 64 GByte hard disk system. For high-resolution images, DLT tapes were taken as transfer and backup media for 10 min. of B/W film. The reprint was done by an AGFA Forte camera together with an Oxberry cine module. Five DEC Alpha workstations (450 MHz) computed the defect analysis and restoration process for the whole 110 min. of *Opernball*, with an average throughput of 20 sec. per frame.

The close cooperation with film experts shows that a film-restoration system must be able to restore distinct defect classes. Because of art-historical reasons, the user (a film expert) decides which defect classes should be restored. To fulfill this requirement, the algorithms used for a digital film-restoration system are built up in a modular manner. Basic algorithmic modules are, for example, local and global motion estimation. Additional algorithmic modules are required for each defect class shown in the defects overview.

A typical restoration task for a scene of *Opernball* includes the following algorithmic modules: global color variation reduction, local motion analysis, global motion analysis, local color-variation reduction, scratch removal, one-frame defect detection and removal, noise suppression, and image-sequence stabilization. To get reliable motion estimates for this moldy film, the global color variation reduction is required before motion estimation. After motion information is available, the local color-variation reduction algorithm can be applied. The resulting image sequence is used for all further restoration steps.

The quality of the digital restoration is tested in two ways. First, film experts certify a good restoration quality regarding the visual appearance of restored image sequences. Second, the quality of distinct algorithms is studied in an objective way. For this reason, representative test scenes were taken for each defect class.

For the detection of one-frame defects, two test scenes were used: the so-called "cellar scene" from the B/W film *Le mort qui tue* and the "piano scene" from the color film *Opernball*. The cellar scene (see MPEG-7) from the *Le mort qui tue* is low defected (low flicker, medium number of contrast-rich dust and dirt areas, no scratches, and low noise). Because of flicker, the global color (brightness) variation reduction algorithm is applied before one-frame defect detection. The piano scene

Figure 16. Total number of detected one-frame defects.

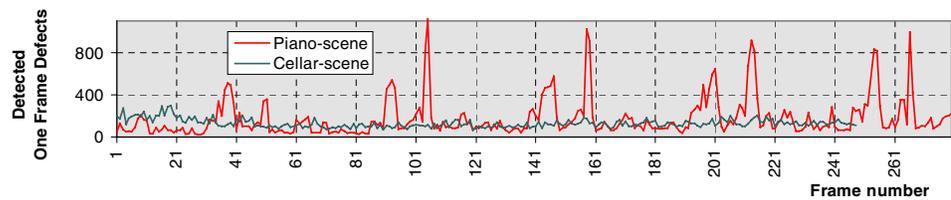
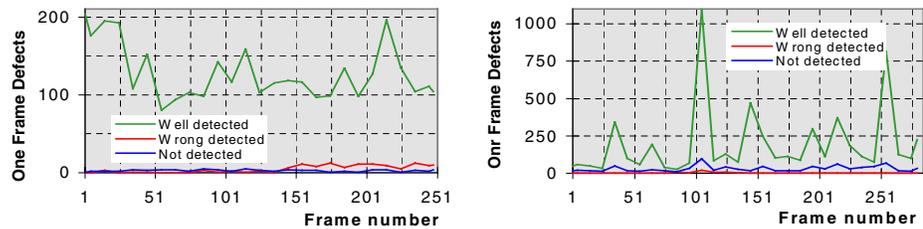


Figure 17. Well-detected, wrongly-detected, and not-detected one-frame defects.



a) Cellar scene: 88 % of not detected defects are caused by very low contrast, 12 % are more frame defects. 44 % of wrong detected defects are caused by lighting changes, 56 % are caused by occlusion of fast moving objects.
 b) Piano scene: 95 % of not detected defects are caused by very low contrast, 5 % are more frame defects. 32 % of wrong detected defects are caused by lighting changes, 68 % are caused by occlusion of fast moving objects.

(see MPEG-4) from the *Opernball* is heavily defected (much mold, high number of low-contrast dust and dirt areas, scratches, and high noise). Because of mold, the local color-variation reduction algorithm is applied before one-frame defect detection. A split screen of the original and the one-frame defect replaced image sequence is given for the cellar scene (see MPEG-8) and for the piano scene (see MPEG-9).

The total number of detected one-frame defects for both scenes is shown in Figure 16. Noticeable are the peaks in the curve for the piano scene. This high number of defects for certain frames originates in the moldy material (see Figure 18). Moldy areas have a compressed gray-value distribution. By applying the required local color-variation reduction algorithm, this gray-value distribution is widened, and dust, dirt, and noise are increased in moldy areas.

Figure 17 shows detailed results for the cellar and the piano scenes. For every tenth frame of both image sequences, the well-detected, the wrongly-detected, and the not-detected defects were classified manually. The not-detected defects originate mainly in very low contrast of defects, and are removed well by the noise-suppression module. The small remainder of not-detected defects have their origin in the fact that they occur over several frames and thus are not detected with the one-frame defect detection scheme. The wrongly-detected defects can be split into two groups: defects detected due to massive lighting changes from frame to frame, and defects detected in occluding areas of fast-moving objects. In the cellar scene (see Figure 17a), fast-moving objects appear from frame number 130 to 250 and in the piano scene, from frame number 90 to 120. The presented one-defect detection scheme decreases the amount and the size of wrongly-detected occlusion areas. Because of the fast motion of objects, the replaced wrong-defect area produces low disturbances in the optical flow; nevertheless, additional work is necessary for reliable detection of occlusion areas in heavily defected image sequences.

For testing the temporal local or global color correction scheme, the piano scene from the color film *Opernball* is used. Two versions of this

Figure 18. Temporal color variations. The red and the blue channel are especially disturbed by mold.

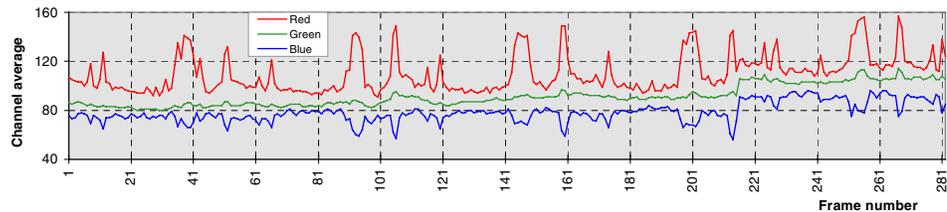


Figure 19. Comparison of different color correction schemes. Temporal local color correction eliminates mold in the best way.

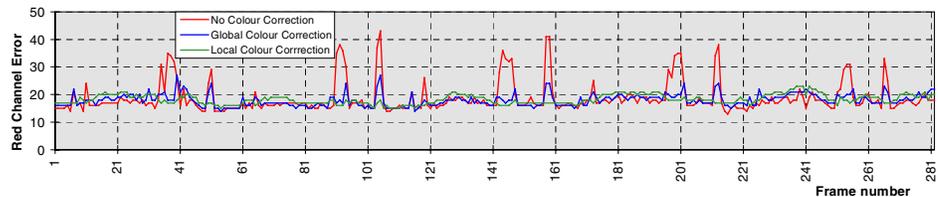


Figure 20. Local color variation caused by mold.



scene are available, a very defected one and a well-stored reference version. Figure 18 shows the pixel value average of each image channel for the defected-scene version.

After manual reference frame determination, the temporal global color-correction scheme and the temporal local color correction-scheme are applied separately to the defected-scene version. An error value is calculated for both correction schemes by taking the channel average of the pixel differences from the reference image sequence and the respective restored image sequence. The restored-image sequence and the reference-image sequence are spatially registered before error-value calculation by a global motion-compensation scheme to avoid false differences caused by different image-sequence sampling. Figure 19 shows the error value of the red channel, calculated for the original defected-image sequence (red curve), for the temporal global color-corrected image sequence (blue curve), and for the temporal local color-corrected image sequence (green curve). Because of different quantization of the reference-image sequence and the defected-image sequence all curves show a bias of about fifteen quantization steps (eight-bit quantization is used for one image channel).

The temporal global color-correction scheme is unable to handle spatially local color variations caused by mold. (See Figure 19, the blue curve for frame number 155 to 160.) The temporal local color-correction scheme can handle this situation well, indicated by the constant course of the green curve in Figure 19. A small part of the defected piano scene is shown in Figure 20; results of the temporal local color-correction are given in Figure 21; and a split screen of both image sequences is shown in MPEG-10.

For testing the image-vibration detection scheme, a test pair was prepared by shifting an image in the x and y directions. The translation in x (a_3) and in y (a_4) was chosen with 2.0 pixels at the raw image size. (In this case, the image was not scaled or rotated.) The dominant motion parameter estimation algorithm is then applied on this image

Figure 21. Image sequence from Figure 20 after temporal local color correction.

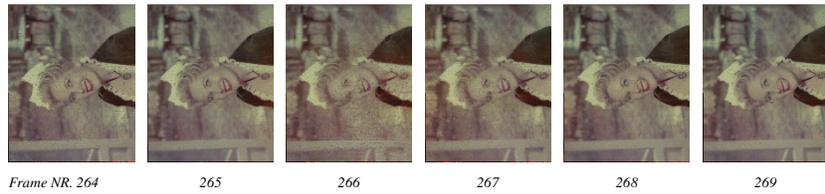
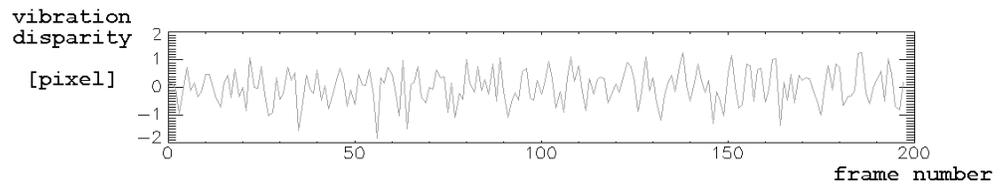


Table 1. Dominant motion estimation results of an x - and y -shifted test image pair.

	a_1	a_2	a_3	a_4
a) Ground truth	1.0	0.0	2.0	2.0
b) Low pass filter				
Level / Type / Width				
1 / Box / 9	1.00002	0.00001	1.96	1.95
1 / Box / 7	1.00005	0.00001	1.94	1.94
1 / Box / 5	1.00005	0.00004	1.93	1.91
1 / Box / 3	1.00004	0.00012	1.89	1.82
c) Multi-resolution iteration with low pass filter				
Level / Type / Width				
4 / Box / 5	0.99884	0.00027	2.46	2.35
3 / Box / 7	0.99987	0.00008	2.05	2.11
2 / Box / 9	1.00003	0.00002	1.96	1.96
1 / Box / 9	1.00002	0.00002	2.01	1.99

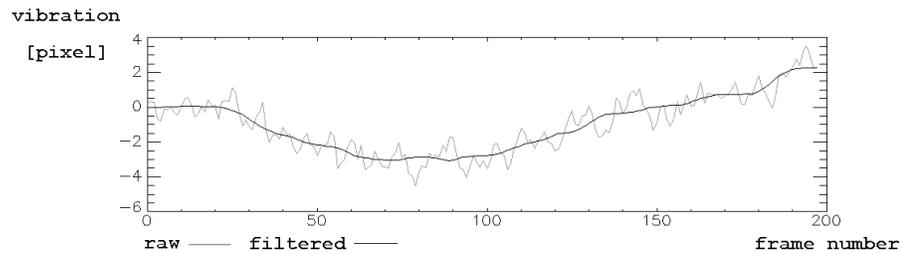
Figure 22. Vibration-disparity signal in y direction.



pair at the half resolution (pyramid level one), first only with varying box filter sizes from width three to nine. No multiresolution iteration scheme was used in this first test. As can be seen in Table 1(b), the accuracy increases when the filter size increases, but it is relatively far away from the ground truth parameters (Table 1(a)). In a second test, the multiresolution iteration was running from level 4 down to level 1 with experimentally determined box filter sizes. The result is much closer to the optimal ground truth parameter than when using low-pass filtering only.

For demonstrating the image-vibration filtering algorithm, the “basket” scene (see MPEG-11) from the film *Le mort qui tue* is used. This image sequence contains image vibrations as well as camera motions. The image vibration detection scheme results in the four image-vibration disparity signals. One of them is shown in Figure 22. Dominant object motion is estimated by the proposed image-vibration stabilization scheme. The corresponding dominant object-vibration signal is shown in Figure 23, and a split screen of the original basket scene and the vibration-stabilized image sequence is shown in MPEG-12.

Figure 23. Image vibration of the dominant object in y direction, before (raw) and after filtering.



8 Conclusions

Digital image-sequence restoration can handle a variety of defect classes that cannot be processed by classical film-restoration equipment. This paper presents algorithms for

- dust and dirt detection and replacement,
- temporal local color correction, and
- image-vibration detection and image-vibration stabilization.

Each of these algorithms constitutes a module in our restoration system that can be activated on demand (e.g., if image vibration is present in a scene).

Algorithms for handling dust and image vibration are designed according to a two-step approach, which allows expert user interaction with intermediate analysis data. Intermediate data are the one-frame defects masks for dust and dirt processing and the four vibration signals for image vibrations. Intermediate data can give hints for an expert user. The final decision whether analysis results are good or not remains with the skilled operator.

One-frame defects can be detected reliably with the presented algorithm in a fully automatic way. Further work must be done in determining occlusion areas to avoid false detections.

During the fully automatic removal phase of one-frame defects, it is necessary to correct the image characteristics like brightness or color of those image areas, which are then pasted into a defect area. Otherwise, natural changing image characteristics like flicker produce visible replacement defects in the pasted area. Image structure inside the replaced defect area is estimated by motion-compensated image structure of neighbor frames and thus provides full spatial resolution. One-frame defect detection can fail in two cases: 1) small, fast-moving objects can be removed erroneously, (more-sophisticated individual motion-tracking algorithms would be required), and, 2) the rather unlikely case that one-frame defects (dust) occur at identical positions in consecutive frames (very high amount of dust). These more-frame defects would require an extension of our algorithm that also considers frames I_{k-i} and I_{k+i} ($i > 1$) to restore frame I_k .

The proposed semiautomatic temporal global brightness and color-correction algorithm is able to correct flicker in a very robust way with minimum time and effort. The temporal local brightness and color-correction algorithm can correct color and brightness variations caused by mold. A compromise between time consumption and possible locality of mold correction is found. Problems can occur if there is a large change in content between two reference frames (high amount of fast motion in the scene), which can lead to wrong changes in color caused by our correction algorithm. One possible remedy is to increase the number of

reference frames (if a sufficient amount of good frames is available). While our algorithm implements a motion-adaptive behavior, improved algorithms would require motion segmentation and tracking of object color, which is computationally far more complex.

Image-vibration detection is accomplished by dominant object tracking through an entire scene. Motion of the dominant object between consecutive frames is modeled with only four parameters for the entire image, resulting in a very robust behavior against local image disturbances and noise. In addition to image vibrations, camera motions like pan, tilt, zoom, and rotation are estimated directly with the same algorithm. Image stabilization detects stationary or moving dominant objects by median filter width adaptation during the vibration filter process. The proposed algorithm is very robust and fails only in rare cases where image content prevents the detection of a dominant object.

Future work will be in the development of new image-processing algorithms for the detection of additional defect classes like scratches or subtitles. Increasing the speed of existing algorithms is now in progress with the ESPRIT project FRAME, in cooperation with the Vienna Centre for Parallel Computing (VCPC).

Today, at least 10% of the 2.2 billion meters of film that are conserved in the world film archives need to be restored. In addition to classical film restoration, a digital film-restoration system will be the most powerful tool for this restoration work. Besides the film market, spinoffs in the area of HDTV (format conversion, compression), video restoration, and multimedia application can also be seen.

9 Acknowledgments

The authors would like to thank Paul Reichl from Taurus Film, the Filmarchiv Austria, and Gaumont (France) for the film material; their colleagues Heimo Müller-Seelich, Walter Plaschzug, and Gerhard Paar from Joanneum Research; and Bruno Despas from Laboratoires Neyrac Films (LNF) for their help and many useful discussions.

References

- [1] M. I. Sezan and R. L. Lagendijk. *Motion Analysis and Image Sequence Processing*. Kluwer Academic Publishers. Boston, 1992.
- [2] A. M. Tekalp. *Digital Video Processing*. Prentice Hall PTR, 1995.
- [3] M. Irani, B. Rousso, and S. Peleg. Computing occluding and transparent motions. *International Journal of Computer Vision*, 12(1), 5–15, 1994.
- [4] S. F. Wu and J. Kittler. A differential method for simultaneous estimation of rotation, change of scale and translation. *Signal Processing: Image Communications*, 2, 69, 1990.
- [5] C. Y. R. Chen, D. S. Meliksetian, M. Cheng-Sheng Chang, and L. J. Liu. Design of a multimedia object-oriented DBMS. *Multimedia Systems 1995–3*, 217–227. Springer, 1995.
- [6] J. R. Bergen, P. J. Burt, R. Hingorani, and S. Peleg. A three-frame algorithm for estimating two-component image motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(9), 886–896, September 1992.
- [7] T. L. Kunii, Y. Shinagawa, R. M. Paul, M. F. Kahn, and A. A. Khokhar. Issues in storage and retrieval of multimedia data. *Multimedia Systems 1995–3*, 217–227. Springer, 1995.

- [8] F. Frey and R. Gschwind. Mathematical bleaching models for photographic three-color materials. *Journal of Imaging Science and Technology*, 38 (6), November/December 1994.
- [9] R. Gschwind and F. Frey. Electronic imaging, a tool for the reconstruction of faded color photographs. *Journal of Imaging Science and Technology*, 38 (6), November/December 1994.
- [10] L. Rosenthaler, A. Wittmann, A. Günzl, and R. Gschwind. Restoration of old movie films by digital image processing. Paper presented at IMAGECOM 96 in Bordeaux/Arcachon, 1996.
- [11] G. Paar and W. Pölzleitner. Robust disparity estimation in terrain modeling for spacecraft navigation. In *Proc. 11th ICPR*. International Association for Pattern Recognition, 1992.
- [14] W. Rucklidge. *Efficient computation of the minimum Hausdorff distance for visual recognition*. PhD Thesis, Technical Report TR94-1454. Cornell University, Department of Computer Science, September 1994.
- [15] R. Zabih, J. Miller, and K. Mai. A feature-based algorithm for detecting and classifying scene breaks. ACM Multimedia Conference, 1995.
- [16] J. R. Bergen, P. J. Burt, R. Hingorani, S. Peleg. A three-frame algorithm for estimating two-component image motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(9): 886–896, September 1992.
- [17] A. C. Kokaram, R. D. Morris, W. J. Fitzgerald, and R. J. W. Rayner. Detection of missing data in image sequences. *IEEE Transactions on Image Processing*, 4(11): 1496–1508, November 1995.
- [18] A. C. Kokaram and P. J. W. Rayner. Removal of replacement noise in motion picture sequences using 3D autoregressive modelling. *Signal Processing*, 7, 1760–1763, 1994.
- [19] P. Schallauer. *Digital Image Sequence Restoration*. Diploma Thesis, Technical University of Graz, June 1996.
- [20] G. Paar and W. Pölzleitner. Descent and landing phase: vision based spacecraft motion estimation and elevation modeling. *Proc. Int. Symp. on Missions, Technologies and Design of Planetary Mobile Vehicles*, Toulouse, September 1992.
- [21] W. Pölzleitner, G. Jakob, and G. Paar. A statistical framework for stereo. In *Proc. Photonics East*, SPIE: The International Society for Optical Engineering, 1995.
- [22] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, 1992.
- [23] XITE, *X-based Image Processing Tools and Environment*. Image Processing Laboratory, Department of Informatics, University of Oslo, Norway, 1995.
- [24] A. C. Kokaram, R. D. Morris, W. J. Fitzgerald, and R. J. W. Rayner. Interpolation of missing data in image sequences. *IEEE Transactions on Image Processing*, 4(11): 1509–1519, November 1995.
- [25] A. Hamapur, R. Jain, and T. E. Weymouth. Production model based digital video segmentation. *Multimedia Tools and Applications*, 1, 1–38, 1995.

- [26] H. Zhang, A. Kankanhalli, and S. Smolinar. Automatic partitioning of video. *Multimedia systems* 1(1):10–28, 1993.
- [27] A. Nagasaka and Y. Tanaka. Automatic video indexing and full video search for object appearances. In W. E. Knuth (ed.), *IFIP Trans., Visual Database Systems II*, 113–128, 1992.

Editors in Chief

Christopher Brown, *University of Rochester*

Giulio Sandini, *Università di Genova, Italy*

Editorial Board

Yiannis Aloimonos, *University of Maryland*

Nicholas Ayache, *INRIA, France*

Ruzena Bajcsy, *University of Pennsylvania*

Dana H. Ballard, *University of Rochester*

Andrew Blake, *University of Oxford, United Kingdom*

Jan-Olof Eklundh, *The Royal Institute of Technology (KTH), Sweden*

Olivier Faugeras, *INRIA Sophia-Antipolis, France*

Avi Kak, *Purdue University*

Takeo Kanade, *Carnegie Mellon University*

Joe Mundy, *General Electric Research Labs*

Tomaso Poggio, *Massachusetts Institute of Technology*

Steven A. Shafer, *Microsoft Corporation*

Demetri Terzopoulos, *University of Toronto, Canada*

Saburo Tsuji, *Osaka University, Japan*

Andrew Zisserman, *University of Oxford, United Kingdom*

Action Editors

Minoru Asada, *Osaka University, Japan*

Terry Caelli, *Ohio State University*

Adrian F. Clark, *University of Essex, United Kingdom*

Patrick Courtney, *Z.I.R.S.T., France*

James L. Crowley, *LIFIA—IMAG, INPG, France*

Daniel P. Huttenlocher, *Cornell University*

Yasuo Kuniyoshi, *Electrotechnical Laboratory, Japan*

Shree K. Nayar, *Columbia University*

Alex P. Pentland, *Massachusetts Institute of Technology*

Ehud Rivlin, *Technion—Israel Institute of Technology*

Lawrence B. Wolff, *Johns Hopkins University*

Zhengyou Zhang, *Microsoft Research, Microsoft Corporation*

Steven W. Zucker, *Yale University*