FAST HIGH QUALITY INTERPOLATION OF MISSING DATA IN IMAGE SEQUENCES USING A CONTROLLED PASTING SCHEME

P.M.B. van Roosmalen[1], A.C. Kokaram[2] and J. Biemond[1]

Information and Communication Theory Group[1]
Faculty of Information Technology and Systems
Delft University of Technology, The Netherlands
e-mail: p.vanroosmalen@its.tudelft.nl

Electrical Engineering Dept.[2]
Trinity College, Dublin 2
Republic of Ireland
e-mail: anil.kokaram@tcd.ie

ABSTRACT

An important topic in image restoration is interpolation of *missing data* in image sequences. Missing data is a result of dirt on film and of ageing processes where the film contents is replaced by data that bears little relationship with the original scene. We present a method for interpolating missing data with the aim of achieving higher fidelity and more consistency in the interpolated results than can be achieved by existing methods. This by combining *autoregressive* models and *markov-random field* techniques. Experimental results confirm the superior performance of the proposed method over existing methods.

1. INTRODUCTION

Missing data is a result of dirt on film and of ageing processes where the film contents is replaced by data that bears little relationship with the original scene. Methods for detecting and correcting missing data can be found in [1,2,3,4,5]. In this paper we concentrate on interpolating missing data and we assume that the spatial locations of the missing data are already known.

Methods for interpolating missing data are either heuristic or model based. Heuristic interpolators, such as spatio-temporal median filters [2], are fast but can introduce significant errors as no constraints are placed on the relation between interpolated intensities and their neighbors. Model-based interpolators, based on *markov random fields* (MRF) or 2D/3D *autoregressive* (AR) models [2,3,4,5] do take relationships between neighboring data into account and give high quality restorations but are computationally more involved. We focus on the model based interpolators.

Kokaram describes a method for interpolating missing data based on a 3D AR model [2]. For each region containing missing data a set of AR parameters is determined from a selection of regional data from 3 consecutive frames. The missing data is localized in the center of the region selected from the central frame. The missing data are interpolated such that the linear-mean-squared-prediction-error, computed using the estimated AR parameters, is minimized. A necessary assumption for this approach is that the data be statistically stationary. Motion compensation is applied to the data to fulfill this requirement. In [4] Goh points out that the assumption of stationarity is not met for regions that are occluded and that become uncovered (and vice versa). Goh suggests estimating the AR model parameters and interpolating the missing data using 2 frames only. One frame is the current frame which contains the missing data. The other frame is either the preceeding or succeeding frame and this depends on what (motion compensated) frame exhibits the smallest mean squared error with the current frame in the region of the missing data, i.e. on the direction in which the motion compensation gives the best match. Kalra [5] further refines this approach by subdividing regions with missing data in multiple regions and interpolating the missing data for each region. This is motivated by the fact that when the missing data covers a large region, a single set of AR coefficients may not be able to model a block of pixels adequately.

There are a number of drawbacks to the methods mentioned. First, the fidelity of the interpolated data in textured regions and in noisy film sequences is not that of its surroundings. This is because AR prediction can smooth data (the approach taken by Kalra reduces the severity of the problem but doesn't solve it fundamentally). Second, the problem of occlusion can in principle be solved following the approach by Goh. However, the direction of interpolation should be determined pixelwise instead of blockwise because the resolution of the occlusion field is finer than that what a block-based approach achieves. Furthermore, by subdividing missing data in a number of regions, as suggested by Kalra, mismatches may well occur within the interpolated results near the region boundaries. Finally, all the approaches assume that the reference regions in the motion compensated previous/next frames do not contain missing data in the regions of interest. This assumption is not always correct and can lead to incorrect interpolated data, as will be demonstrated.

This paper presents a method for correcting missing data with high quality that is robust to the problems mentioned. Our approach to builds on the methods described. Here, however, the interpolated data do not consist of intensities predicted by AR models. Instead we paste pixels from either the previous or from the next motion compensated reference frames. Using explicit pasting instead of full spatio-temporal data regeneration (as imployed in previous work) is motivated by the observation that in many cases the image sequence is redundant enough enough for simple copy operations (i.e., 1 tap temporal AR process) to suffice for data interpolation.

The strategy for determining the direction of interpolation (i.e., paste from previous frame or paste from next frame) is based on the assumption that the corrected image follows a 2D AR process. This process reports on the consistency of the interpolated result generated by the pasting process: the pixel intensity from that reference frame that gives the smallest prediction error is pasted into the current frame. The approach described so far can lead to smoothing in the sense those values will be pasted from the reference frames that are as close as possible to the predicted values which, inherently to AR prediction, can be smooth. To maintain high fidelity we constrain the direction of interpolation to be a consistent one locally.

Note that our method is robust to occlusion and missing data in either the previous frame or in the next frame. Data from occluded regions and from regions covered by dirt tend to be very different from the original data and lead to large prediction errors and will thus not be selected for pasting.

The outline of the remainder of this paper is as follows. In Section 2 we describe our approach in more detail. Then, in Section 3, we describe the experiments and the results. We conclude this paper with a discussion in Section 4.

2. CONTROLLED PASTING FOR INTERPOLATING MISSING DATA

We denote the observed (motion compensated) previous, current and next frames by $Y_-(\boldsymbol{r})$, $Y(\boldsymbol{r})$, $Y_+(\boldsymbol{r})$ respectively, where $\boldsymbol{r}$ indicates the spatial coordinates. $M(\boldsymbol{r})$ is the binary *blotch detection mask* that indicates for each pixel in the current frame whether or not it is missing. It is assumed that $M(\boldsymbol{r})$ has been determined, e.g. by following [1].

Let $D(\boldsymbol{r})$ be a binary *occlusion mask*. Usually occlusion masks indicate whether or not data in a reference frame is occluded. In our case, however, it indicates for each spatial location which of the frames $Y_-(\boldsymbol{r})$ or $Y_+(\boldsymbol{r})$ is most appropriate to serve as a reference, e.g. "0" for $Y_-(\boldsymbol{r})$ and "1" for $Y_+(\boldsymbol{r})$. We now model the likelihood of an unimpaired frame I to be proportional to:

$$\begin{aligned} &P(I = i | Y_-, Y, Y_+, M, D) \propto \\ &\exp\Big(-\frac{1}{T}\sum_{r \in S}[(1 - M(\boldsymbol{r})) \cdot \delta(Y(\boldsymbol{r}) - i(\boldsymbol{r})) + \\ &M(\boldsymbol{r}) \cdot (d(\boldsymbol{r}) \cdot Y_-(\boldsymbol{r}) + (1 - d(\boldsymbol{r})) \cdot Y_+(\boldsymbol{r}) - AR(i, \boldsymbol{r}))^2]\Big) \end{aligned}, \quad (1)$$

where S is the set of all spatial locations and $AR(i, \boldsymbol{r})$ indicates a linear prediction of $i(\boldsymbol{r})$. The first term in (1), which given $M(\boldsymbol{r})$ is a constant, states that the intensities at uncorrupted sites are likely to be equal to the observed intensities $Y(\boldsymbol{r})$. Using the 2D AR model, which is our underlying model of an unimpaired image, the second term relates the original, clean data to either $Y_-(\boldsymbol{r})$ or $Y_+(\boldsymbol{r})$ depending on the binary occlusion mask.

To achieve some local consistency in the occlusion mask, the following prior is assumed:

$$P(D = d) \propto \exp\Big(-\frac{1}{T}\sum_{\boldsymbol{r} \in S} \beta \cdot f(d(\boldsymbol{r}))\Big), \quad (2)$$

where β is a constant that defines the strength of the self-organization, and $f(d(\boldsymbol{r}))$ is the number of *eight-connected* neighbors of $d(\boldsymbol{r})$ that have different values from $d(\boldsymbol{r})$.

Combining (1) and (2), and ignoring the constant term in (1), gives the *a posteriori* distribution of the corrected frame and the *direction mask*:

$$\begin{aligned} &P(I = i, D = d | Y_-, Y, Y_+, M) \propto \\ &\exp\Big(-\frac{1}{T}\sum_{\boldsymbol{r} \in S} M(\boldsymbol{r})[d(\boldsymbol{r})(Y_+(\boldsymbol{r}) - AR(i, \boldsymbol{r}))^2 + \\ &\qquad (1 - d(\boldsymbol{r}))(Y_-(\boldsymbol{r}) - AR(i, \boldsymbol{r}))^2 + \beta \cdot f(d(\boldsymbol{r}))]\Big) \end{aligned} \quad (3)$$

Maximizing (3) *w.r.t.* $i(\boldsymbol{r})$ and $D(\boldsymbol{r})$ gives the corrected image $\hat{i}(\boldsymbol{r})$. However, as mentioned before, we wish to avoid smoothing by restricting the correction to consist of values pasted from the reference frames. It is appropriate to have the occlusion mask $D(\boldsymbol{r})$ control the pasting because it indicates which reference frame is likely to be closest to the original unimpaired data. We assume that at least one of the reference frames is uncorrupted. Therefore let:

$$\hat{i}(\boldsymbol{r}) = \begin{cases} Y_-(\boldsymbol{r}) & if \quad D(\boldsymbol{r}) = 0 \text{ and } M(\boldsymbol{r}) = 1 \\ Y_+(\boldsymbol{r}) & if \quad D(\boldsymbol{r}) = 1 \text{ and } M(\boldsymbol{r}) = 1 \\ Y(\boldsymbol{r}) & else \end{cases}. \quad (4)$$

Because of (4) $\hat{i}(\boldsymbol{r})$ depends directly on $D(\boldsymbol{r})$. Therefore, searching the maximum of the a posteriori distribution (3), given the observed data and the blocth detection mask, requires finding the most likely configuration of $D(\boldsymbol{r})$. We apply *simulated annealing* (SA) for this purpose.

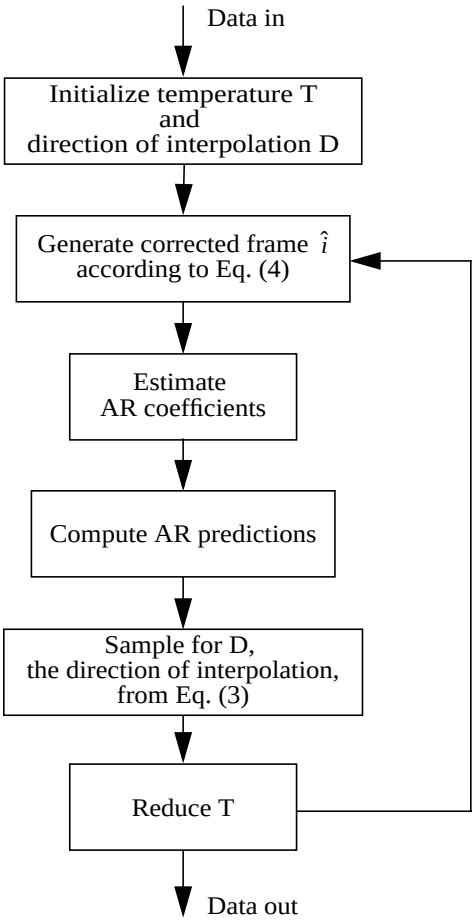

Figure 1. Overview of the CP scheme.

Figure 1 summarizes the *controlled pasting* (CP) approach. The data put into the system consist of the current frame and the motion compensated previous and next frames. The blotch detection mask, which indicates for each pixel whether or not it is considered to be part of a blotch, also belongs to the input data. Initially the occlusion mask is assigned binary values at random.

The main loop is as follows. First a corrected frame $\hat{i}$ is interpolated. Next the AR coefficients need to be estimated (their true values are not known). Strictly speaking, in an SA scheme, all unknowns should be sampled and this includes the AR coefficients (e.g., see [3]). However, good results are obtained by using the *least squares* estimate for the AR coefficients at this step using the current state of the image data. This helps increase the speed of the process.

We used a quarter plane prediction model, see Fig. 2. The image regions are selected such that at most 20% of the area consists of missing data. The predicted values are used for evaluating the cost function determined by the summation in (3). Based on the computed costs the direction of interpolation is updated using simulated annealing. The main loop is repeated for the duration of the cooling schedule. The system output is the corrected frame $\hat{i}$.

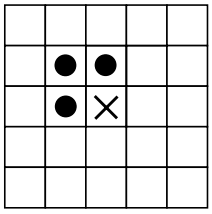

Figure 2. Support (dots) used for AR prediction (cross).

3. EXPERIMENTS AND RESULTS

For our test sequence we use the *Western* sequence which was also used by Kokaram in [*2*]. This sequence has been artificially corrupted with blotches that can have any gray level. Figure *3b* shows frame 35 from this sequence, the blotches are clearly visible. Figures *3 a,c* show the motion compensated previous and next frames (an hierarchical blockmatcher with some smoothness constraints was used for determining the required motion vectors) and it is clearly visible that these frames corrupted by blotches as well. The blotches were detected using the method described in [1].

Although our method has been described using SA for optimisation, we find that good results are obtained just by using 15 iterations of the Gibbs Sampler alone, i.e. by letting $T = 1$ fixed. The results shown are generated as such.

Figures *3 d,e,f* show three different corrections of frame 35, respectively using the *3DAR* method described by Kokaram, the *B3DAR* method described by Goh, and the *CP* method proposed in this paper. All the corrected frames show a great improvement over the corrupted frame. However, the *3DAR* and the *B3DAR* methods fail where the motion compensated frames are corrupted (see the highlighted boxes in the figures). The former method fails because it always incorporates both motion compensated frames, even when data is corrupt in these frames. The latter method fails because a block based approach determines direction of interpolation regardless on the validity of the data within the block. Figures *3 g,h,i* zoom in on the boxed regions in Figures *3 d,e,f*. Clearly, the proposed method gives results with high fidelity and outperforms the other methods in terms of visual quality.

4. DISCUSSION

We introduced a new method for interpolating missing data in image sequences. It achieves high fidelity by pasting data in a consistent manner instead of applying linear filtering. In effect it is a controlled pasting process. The limited number of iterations required for the simulated annealing and the fact that only corrupted image regions have to be

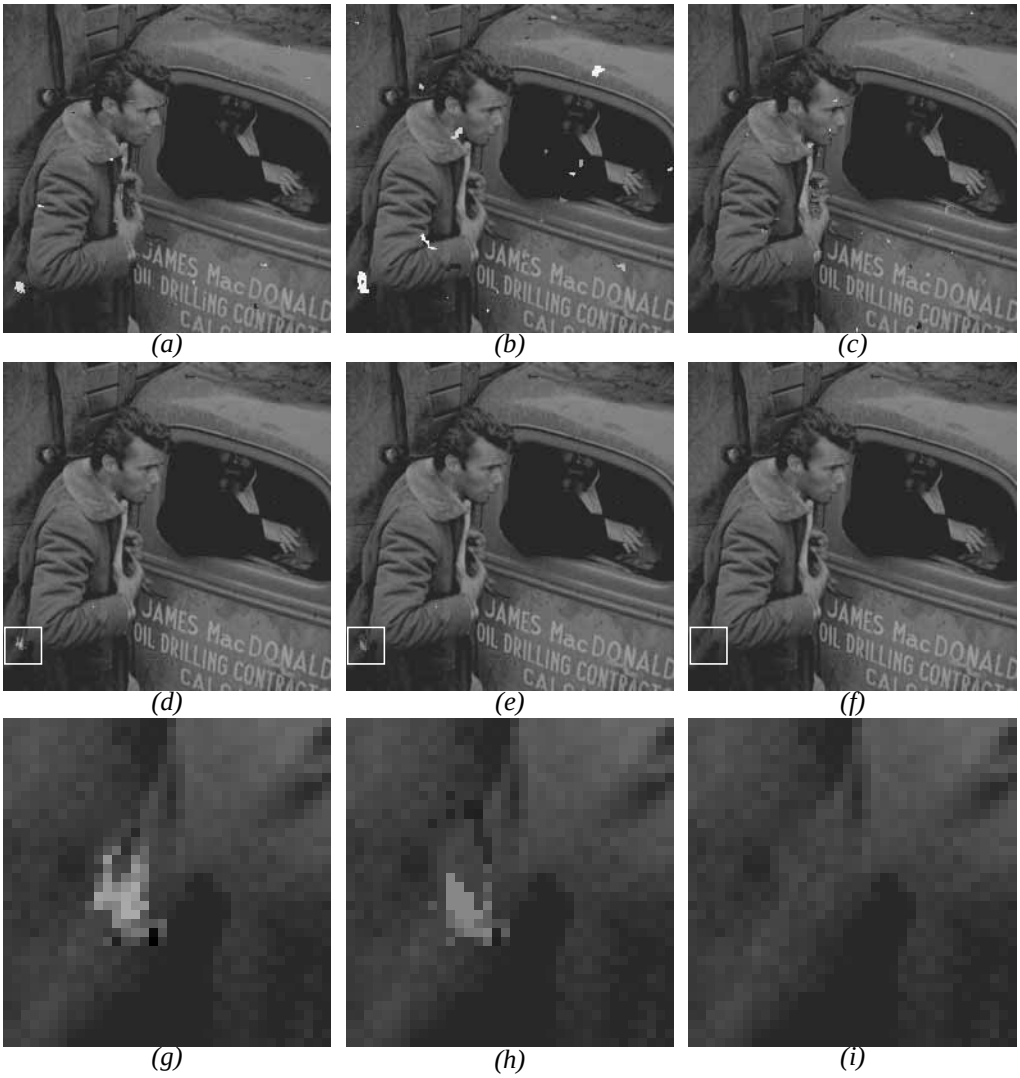

Figure 3. (a) Motion compensated previous frame. (b) Current frame. (c) Motion compensated next frame. (d), (e), (f) Restored frames using 3DAR, B3DAR and CP schemes respectively. Note the differences within the boxed regions. (g), (h), (i) Zoom into boxed regions.

processed makes the algorithm quite fast in practice. Our method does not require explicit knowledge about missing data in the reference frames. This is a useful property as this reduces complexity for an overall detection and correction system, and, more important, it makes the interpolator robust to *misses* of the blotch detector (i.e. when the detector does not detect all the blotches in the reference frames).

Interpolating missing data is not only relevant in case of old film sequences. It can also be applied to image sequences that have been transmitted digitally and that have been corrupted due to bit errors in the received (compressed) data.

REFERENCES

1 *J. Biemond, P.M.B. van Roosmalen, R.L. Lagendijk, "Improved Blotch Detection by Postprocessing", Submitted to ICASSP 99.*

2 *A.C. Kokaram, R.D. Morris, W.J. Fitzgerald, and P.J.W. Rayner, "Interpolation of Missing Data in Image Sequences", IEEE Trans. on Image Processing, pp. 1509-1519, Vol. 4, No. 11, 1995.*

3 *A.C. Kokaram, "Motion Picture Restoration", Springer Verlag, 1998.*

4 *W.B. Goh, M.N. Chong, S. Kalra, and D. Krishnan, "Bi-Directional 3D Auto-Regressive Model Approach to Motion Picture Restoration", Proc. of ICASSP 96, pp. 2277-2280, 1996.*

5 *S. Kalra, M.N. Chong, and D. Krishnan, "A new Auto Regressive (AR) model-based algorithm for Motion Picture Restoration", Proc. of ICASSP 97, Vol. 4, pp. 2557-2560, Munich, Germany, 1997.*