Robust Motion Vector Relaxation for X-Ray Fluoroscopy Using Generalized Gauss-Markov Random Fields

Til Aach and Dietmar Kunz

Philips Research Laboratories,
Weisshausstr. 2, D-52066 Aachen, Germany
Email: (aach/kunz)@pfa.research.philips.com

Abstract. We describe a Bayesian motion estimation algorithm which is part of a temporally recursive noise reduction filter for X-ray fluoroscopy images. Our algorithm draws its robustness against high quantum noise levels from a statistical regularization, where a priori expectations about the spatial and temporal smoothness of motion vector fields are modelled by generalized Gauss-Markov random fields. We show that by using generalized Gauss-Markov random fields both smoothness and motion edges can be captured, without the need to specify an often critical edge detection threshold. Instead, our algorithm controls edges by a single parameter by means of which the regularization can be tuned from a median-filter like behaviour to a linear-filter like one.

Keywords: fluoroscopy, image restoration, Bayesian motion estimation, generalized Gauss-Markov random fields, thresholdless edge model.

1 Introduction

We describe a robust X-ray fluoroscopy motion estimator which we use within a motion compensated temporally recursive noise reduction filter. Our estimation algorithm is well able to cope with the high noise levels in fluoroscopy (see Fig. 1), which are caused by the very low X-ray dose rates used in fluoroscopic imaging [1, 2].

Motion compensated temporal filters for usually only moderately noisy real world video sequences [3] are often based on the well known block matching algorithm for motion estimation, which, however, is very noise-sensitive. To make motion estimation robust against noise, we regularize the estimation problem by suitably formulated prior knowledge, namely by exploiting the expected smoothness in both space and time of 'true' motion vector fields. We express these prior expectations by generalized Gauss-Markov random fields applied within a Bayesian framework. The advantage of using generalized Gauss-Markov random fields is that, apart from smoothness, they can also describe motion edges without requiring explicit specification of edge detection thresholds. Edges are controlled by a single 'soft' parameter, by means of which the influence of regularization can be tuned from a median-filter like behaviour to a linear-filter like one.

2 Bayesian Motion Estimation

2.1 MAP-Approach

Let Y_n denote the n-th frame of the observed fluoroscopy sequence, with $y_n(k)$, $k = 1, \ldots, N$ being the grey level at pixel k, and N being the number of pixels in each frame. (For simplicity, we use a one-dimensional notation here). We seek motion vectors $v_n(k)$ such that for each pixel k, $k + v_n(k)$ is the corresponding location in the previous frame Y_{n-1}, with the motion vector matrix $v_n(k)$, $k = 1, \ldots, N$ forming the motion field V_n. Given the images Y_n and Y_{n-1} *and* the previous motion field estimate V_{n-1}, we intend to determine the motion field estimate $\hat{V}_n$ so that it obeys

$$\hat{V}_n = \arg\max_V p\left(V | Y_n, Y_{n-1}, \hat{V}_{n-1}\right) \quad , \tag{1}$$

with $p(V|Y_n, Y_{n-1}, \hat{V}_{n-1})$ being the a posteriori probability density function (pdf) for a motion field V given the observations. With conditional statistical independence between (Y_n, Y_{n-1}) and $\hat{V}_{n-1}$, this can be rewritten to

$$\hat{V}_n = \arg\max_V \left[p\left(Y_n, Y_{n-1}|V\right) \cdot p\left(\hat{V}_{n-1}|V\right) \cdot p\left(V\right)\right] \quad . \tag{2}$$

2.2 The Observation Model

To specify the likelihood term $p(Y_n, Y_{n-1}|V)$, we assume that temporal grey level changes are caused only by either motion or (quantum) noise (cf. [4]). As the influence of motion can be compensated by the *displaced* frame difference (DFD) $D(V_n) = \{d[k, v_n(k)]\}$, where $d[k, v_n(k)] = y_n(k) - y_{n-1}(k + v_n(k))$, what remains is modelling quantum noise. Approximating the Poissonian distribution of quantum noise by a Gaussian one with signal-dependent variance, we obtain

$$p\left(Y_n, Y_{n-1}|V\right) \propto \prod_{k=1}^{N} \frac{1}{\sqrt{4\pi\sigma^2(k)}} \cdot \exp\left\{-\frac{d^2(k, v(k))}{4\sigma^2(k)}\right\} \quad , \tag{3}$$

where we also made the approximating assumption of spatially uncorrelated noise. The variable $\sigma^2(k)$ is the quantum noise variance at each pixel k. In the remainder of this paper, we assume that the quantum noise variance σ^2 as a function of intensity — and hence also as a function of location k — is known [2], either by being predicted from acquisition parameter settings of the system, or by estimation from the acquired images.

2.3 The Motion Field Model

We model $p(V)$ by a Markov random field, where statistical dependencies are described locally [5]. Invoking the Hammersly-Clifford theorem, this permits to write $p(V)$ as a Gibbs density, i.e. $p(V) \propto \exp\{-E(V)\}$, where $E(V)$ is an *energy* ascribed to a motion field. This density clearly favours low energies, hence $E(V)$

should be specified in such a way that it is low for smooth motion vector fields V. As a consequence of the Markovian assumption, this energy can be decomposed into a sum of local energy terms according to $E(V) = \sum_{k=1}^{N} E_L(k, V)$. Each local energy $E_L(k, V)$ depends only on the eight vectors adjacent to pixel k (2nd order Markov random field). To obtain a local energy which favours smooth vector fields, we penalize differences between neighbouring motion vectors by

$$E_L(k, V) = \lambda \cdot \sum_{l \in N_k} g_l \cdot \Delta_l^{\nu}(k) \quad , \tag{4}$$

with $\Delta_l(k) = ||v(k) - v(l)||_2$. Our model views differences between $v(k)$ and its four diagonally adjacent neighbours as less severe than differences between $v(k)$ and its four nearest neighbours. This is expressed through the weights g_l, with $g_l = 1$ if $v(l)$ is horizontally or vertically adjacent to $v(k)$, and $g_l = 0.5$ for the diagonal neighbours of $v(k)$. The factor $\lambda > 0$ is a weighting parameter.

For exponent values $\nu > 1$, this model clearly discourages large differences between adjacent vectors, and hence encourages (locally) smooth motion fields. For $\nu = 2$, (4) is referred to as a Gauss-Markov random field ("quadratic regularization") [6]. While this model is quite common to describe smoothness, it inherently blurs abrupt transitions of motion occurring between neighbouring and differently moving objects. The reason for this is that motion edges correspond to relatively large values of $\Delta_l^2(k)$, which are governed by the quickly decaying tails of the (multidimensional) Gaussian density. To prevent such blurring, (4) can be replaced by

$$E_L(k, V) = \lambda \cdot \sum_{l \in N_k} g_l \cdot \min\left\{\Delta_l^2(k), T\right\} \quad , \tag{5}$$

which is quadratic if $\Delta_l^2(k)$ is smaller than the *edge detection threshold* T, and flat otherwise [7]. The necessity of having to specify this threshold, however, is of twofold disadvantage: firstly, it is often hard to find an appropriate threshold value to reliably detect edges of usually unknown magnitude. Furthermore, it is by no means certain that all edges can be captured by single threshold parameter. Also, (5) is not convex, which is disadvantageous for the maximization of (2).

These problems can be avoided by employing a density the tails of which do not fall off as rapidly as those of a quadratic Gaussian density. To achieve this, it suffices to *reduce* the exponent ν, what results in a *generalized* Gauss-Markov random field [6]. Reduction of ν corresponds to a gradual increase of the magnitude of the tails of the density function, what in turn makes edges more likely to occur. Eventually, for $\nu = 1$, a (multidimensional) Laplacian density is reached.

In temporal direction, we use a similar model. Assuming that each vector $\hat{v}_{n-1}(k)$ depends only on its successor $v(k)$, we have $p(\hat{v}_{n-1}(k)|V) = p(\hat{v}_{n-1}(k)|v(k))$. Assuming furthermore conditional statistical independence results in $p(\hat{V}_{n-1}|V) = \prod_{k=1}^{N} p[\hat{v}_{n-1}(k)|v(k)]$. Inserting the generalized Gaussian model yields

$$p[\hat{v}_{n-1}(k)|v(k)] \propto \exp\left\{-\lambda' \cdot ||v(k) - \hat{v}_{n-1}(k)||_2^{\nu}\right\} \quad . \tag{6}$$

3 The Estimation Algorithm

Inserting the derived expressions for $p(Y_n, Y_{n-1}|V)$, $p(\hat{V}_{n-1}|V)$ and $p(V)$ into (2), and taking the negative logarithm results in the following optimization criterion:

$$C(V|Y_n, Y_{n-1}, \hat{V}_{n-1}) = \sum_{k=1}^{N} \Big[\frac{d^2(k, v(k))}{\sigma^2(k)} + \lambda' \cdot ||v(k) - \hat{v}_{n-1}(k)||_2^{\nu} + \lambda \cdot \sum_{l \in N_k} g_l \cdot \Delta_l^{\nu}(k) \Big] \ , \tag{7}$$

where a factor of four was absorbed into λ and λ'. We minimize this criterion by a deterministic ICM-relaxation [5] on a blockwise basis, where each block comprises e.g. 16×16 pixels and counts as one macro-pixel. Initialization of the relaxation can for instance be done by full search block matching or by simply using the motion vector field for the previous frame. The initial estimate is then optimized by repeatedly scanning the image grid. For each block, the motion vector is tentatively replaced by the following candidates: each of its eight spatial neighbours, its temporal predecessor, the four vectors differing from the current one by half a pixel, and the average of all vectors in N_k weighted by g_l. For each of these candidates and the current vector, the optimization criterion is *locally* recomputed, and the candidate minimizing the criterion retained. In practice, this relaxation converges in only three iterations.

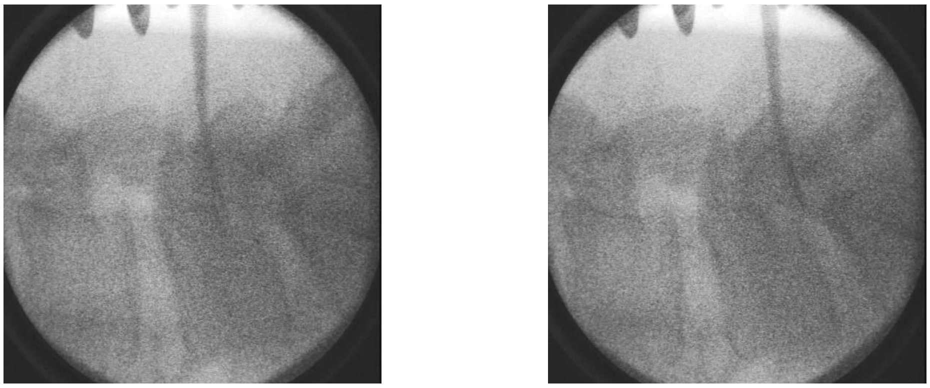

Fig. 1. Two subsequent fluoroscopy frames depicting vertebrae which remain almost still, whereas the surgical instrument in the upper image parts starts moving abruptly in horizontal direction.

Fig. 1 shows two subsequent frames of a noisy fluoroscopy sequence, between which the depicted surgical tool starts to move abruptly in horizontal direction. The motion field estimated by the Gaussian model ($\nu = 2$) is shown left in Fig. 2. Clearly, the motion of the tool is not captured due to edge blurring. This is quite different in the right hand side of Fig. 2, where the tool motion vectors are clearly delineated from the stationary background. Here, estimation was based on the generalized Gaussian model with $\nu = 1.3$.

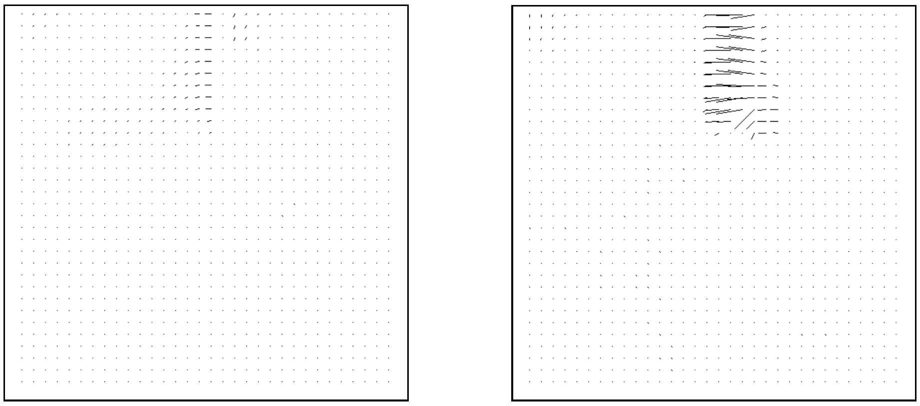

Fig. 2. *Left:* Motion vector field with $\nu = 2$, oversmoothing the motion of the tool. *Right:* Motion vector field with $\nu = 1.3$, where the motion of the tool is clearly reconstructed.

4 Conclusions

We have described a regularized motion estimator which is on the one hand able to cope with strong quantum noise levels, while on the other hand well reconstructing motion edges. No edge threshold is necessary, rather, edges are controlled by a single 'soft' parameter. In combination with a recursive noise reduction filter [3], significant noise reduction can be achieved without noticeably blurring objects.

An extended paper of this algorithm is to appear in a forthcoming issue of the Philips Journal of Research.

References

1. T. Aach, U. Schiebel, G. Spekowius, "Digital image acquisition and processing for medical x-ray imaging applications," Proc. ISEP-96, Cologne,Sept. 21–22, 82–90.
2. T. Aach, D. Kunz, "Spectral estimation filters for noise reduction in x-ray fluoroscopy imaging," Proc. EUSIPCO-96, Trieste, Sept. 10–13, 571–574.
3. E. Dubois, S. Sabri, "Noise reduction in image sequences using motion compensated temporal filtering," IEEE Trans. Comm. 32(7), 826–831, 1984.
4. T. Aach, A. Kaup, "Disparity-based segmentation of stereoscopic foreground / background image sequences," IEEE Trans. Comm. 42(2), 673–679, 1994.
5. J. Besag, "On the statistical analysis of dirty pictures," J. Roy. Stat. Soc. B 48(3), 259–302, 1986.
6. C. Bouman, K. Sauer, "A generalized gaussian image model for edge-preserving MAP-estimation," IEEE Trans. Im. Proc.2(3)296–310, 1993.
7. A. Blake, A. Zisserman, Visual Reconstruction. Cambridge, MIT Press, 1987.