DIRECT MOTION ESTIMATION IN THE RADON TRANSFORM DOMAIN USING MATCH-PROFILE BACKPROJECTIONS

C. Bartels †

†Eindhoven University of Technology
Den Dolech 2 5600 MB Eindhoven
The Netherlands

G. de Haan †‡

‡Philips Research Laboratories
High Tech Campus 36 5656 AA Eindhoven
The Netherlands

ABSTRACT

Integral projections have been proposed as an efficient method to reduce the dimensionality of the search space in motion estimation (ME) algorithms. A number of papers describe methods extending direct (correlation based) block matching algorithms with projections, others derive optical flow methods based on the Radon transform. With a single exception all of the above methods use a single horizontal and vertical projection to keep the ME algorithm computationally simple. In this work we derive a generalization on the direct projection methods: a ME method based on backprojections, and investigate the added value of using multiple projection angles.

Index Terms— Motion Analysis, Radon Transforms

1. INTRODUCTION

Motion estimation is a key technique in modern video transmission chains and display systems. It has a wide variety of applications, e.g. segmentation, compression, noise-filtering, deinterlacing, super-resolution and frame-rate up conversion. This paper focuses on the real-time use of motion estimation, a domain where for example HDTV and mobile devices pose new challenges in speed, quality and cost.

Direct (correlation based) block matching algorithms (BMA) are popular in industry due to their relative efficiency and possible parallel implementations. The trade-off between quality and cost is usually determined by (I) search range / search candidates, (II) matching criterion / cost function and (III) spatio-temporal smoothness constraints.

Projections reduce the dimensionality of the search space and match criterion; as such they have the potential to improve BMA. Their use is not new: integral projections for ME were first introduced in [1, 2], computational efficiency and reduced sensitivity to noise were named advantages of the technique. Since then, a large number of optimizations and derivative methods have been proposed aiming mainly at I and/or II [3–6]. More recently, projection and Radon optical flow based ME methods for global and local use were introduced in [7, 8].

A common feature in the direct correlation based works [1–6] is the use of two projections at 0 and 90 degrees as a basis for motion estimation.

To create projections, we will make use of the Radon transform (Section 2). Correlation of projections, using e.g. the sum of absolute differences (SAD), yields 1-D matchprofiles whereas correlation of blocks yields 2-D matchprofiles[1]. In this work we look at the reconstruction of a 2-D matchprofile from multiple 1-D matchprofiles for which we use an inverse Radon transform method called backprojection. We first show the approximate equality of the backprojected 1-D matchprofiles and the 2-D matchprofile, and then explore the use of the former in a block matching algorithm for local ME. More specifically we look at the performance related to the number of projections, using multiple projection angles.

The paper is therefore layout as follows: Section 2 introduces definitions of the forward and inverse Radon transform, these are required for the formulation of the method in Section 3. Section 4 discusses the results of a local ME implementation and comparison with phase plane correlation ME. In Section 5 we draw conclusions based on this work.

2. DEFINITIONS

2.1. Radon Transform

A line l in $\mathbb{R}^2$ can be parametrized as indicated in Figure 1:

$$p = x\cos\phi + y\sin\phi \tag{1}$$

For any arbitrary function $f(x,y) : \mathbb{R}^2 \rightarrow \mathbb{R}$ the Radon transform [9–11] can be expressed as

$$\begin{aligned}\mathcal{R}f(p,\phi) &= \int\limits_{\underline{k}\in l(p,\phi)} f(\underline{k})d\underline{k} \\ &= \int\limits_{-\infty}^{\infty} f(p\cos\phi - s\sin\phi, p\sin\phi + s\cos\phi)ds\end{aligned} \tag{2}$$

or the line integral of f over all lines l defined by p and ϕ, ds representing an increment along line l. We will use the terms *projection indices* and *projection angles* for p and ϕ respectively. The Radon transform can also be expressed as an area integral over dx, dy using the dirac function δ:

$$\begin{aligned}\mathcal{R}f = \widehat{f}(p,\phi) &= \iint\limits_{(x,y)\in\mathbb{R}^2} f(x,y)\delta(p - x\cos\phi - y\sin\phi)dxdy \\ &= \iint\limits_{(x,y)\in\mathbb{R}^2} f(x,y)\delta(p - \underline{\xi}\cdot\underline{x})dxdy\end{aligned} \tag{3}$$

[1] Shifting a projection / image with one / two degrees of freedom respectively.

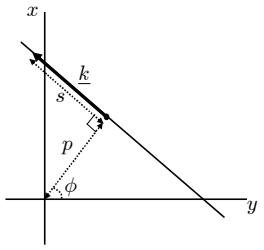

Fig. 1. Parameter definitions for a line in 2-D. Underlined symbols denote vectors.

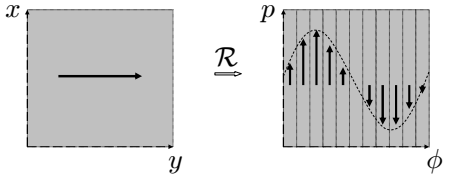

Fig. 2. Translational motion in image and Radon domain.

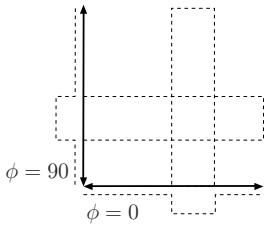

Fig. 3. Backprojection of 2 orthogonal projections.

The last notation consists of the inner product of the unit direction vector $\underline{\xi} = (\cos\phi, \sin\phi)^T$ and the coordinate vector $\underline{x} = (x, y)^T$.

The Radon transform is a linear transform, it further has the following shift property:

$$\mathcal{R}f(\underline{x} - \underline{a}) = \widehat{f}(p - \underline{\xi} \cdot \underline{a}, \phi) \tag{4}$$

A shift in the original domain with $\underline{a}$ results in a shift of the projection with the inner product $\underline{\xi} \cdot \underline{a}$. This property is exploited for motion estimation: translational motion in an image sequence can be reconstructed from projections as the motion is projected in the direction of the projection angles ϕ. Figure 2 shows an image block with an arrow depicting translational motion (left) and the corresponding motion in a number of projections with varying angle ϕ (right).

2.2. Inverse Radon Transform

In this section a method to approximate the inverse Radon transform is introduced. We use a *summation*, or *backprojection*, type of algorithm [9–11] to derive a 2-D matchprofile. Traditionally these types of algorithms are used in (tomographic) image reconstruction from projections. The backprojection operator on a function $h : \mathbb{Y} \rightarrow \mathbb{R}$ is defined as[2]:

$$\mathcal{B}h(x, y) = \int_0^{\pi} h(x\cos\phi + y\sin\phi, \phi)d\phi \tag{5}$$

The backprojection method can be thought of as smearing out the projections in their respective orthogonal directions. As such, each point in the image plane is determined by summation of a single point in all projections, the projection indices p are located along a sinusoid in the Radon domain. A backprojection example is shown in Figure 3.

The $\mathcal{B}$-operator is not exactly equal to the inverse of the Radon transform: distortion occurs when simply backprojecting Radon transformed functions. In [10] the following property of the backprojected image is described:

$$\check{f} = \mathcal{B}\mathcal{R}f \tag{6}$$

The true image $f(x, y)$ is related to $\check{f}(x, y)$ as follows:

$$\check{f}(x, y) = f(x, y) ** \frac{1}{\sqrt{x^2 + y^2}} = \iint \frac{f(x', y')dx'dy'}{\sqrt{(x - x')^2 + (y - y')^2}} \tag{7}$$

[2] (p, ϕ) span up the extended cylinder space $\mathbb{Y}$, readers can approximately envision this as a polar space. We refer to Rowland [9] for exact definitions.

where $**$ represents 2-D convolution. Backprojection affects the frequency spectrum in the sense that the contribution of a particular frequency linearly decreases with the distance from the origin (the zero frequency). Corrected backprojection methods therefore apply a high-pass filtering, ideally a ramp in the frequency domain, on the projections to compensate for this effect. In [11] the filtering operator $\mathcal{C}$ is defined such that $\mathcal{B}\mathcal{C}\mathcal{R}f = f$. That is,

$$\mathcal{C}h(p, \phi) = \int_{-\infty}^{\infty} |r|\Big(\int_{-\infty}^{\infty} h(\tilde{p}, \phi)e^{-j2\pi r\tilde{p}}d\tilde{p}\Big)e^{j2\pi rp}dp \tag{8}$$

We define the filtered backprojection operator:

$$\mathcal{B}_f = \mathcal{B}\mathcal{C} \tag{9}$$

3. MATCH-PROFILE BACKPROJECTION

In this section we will derive the new correlation measure $\mathcal{B}_f\mathcal{S}_Y$ and show its similarity to the standard 2-D correlation measure, i.e., the sum of absolute differences (SAD) denoted as $\mathcal{S}_{2D}$ in Equation 10. Luminance functions L_n and L_{n+1} denote the luminance values in frame n and $n + 1$ at pixel positions (x, y) and $\underline{d} = (d_x, d_y)^T$ denotes the translational displacement. Note that we simplify to an infinite continuous image domain, whereas the real implementation has to be on discrete image blocks of finite size. Thus we let

$$(\mathcal{S}_{2D}[L_n, L_{n+1}])(d_x, d_y) = \iint_{(x,y)\in\mathbb{R}^2} |(L_n(x - d_x, y - d_y) - L_{n+1}(x, y)|dxdy \tag{10}$$

Calculating the SAD on a pair of projections reduces the complexity by an order of magnitude. Each projection has a single displacement d_p. The projection-SAD then follows:

$$(\mathcal{S}_Y[L_n, L_{n+1}])(d_p, \phi) = \int_{t\in\mathbb{R}} |\mathcal{R}L_n(t - d_p, \phi) - \mathcal{R}L_{n+1}(t, \phi)|dt \tag{11}$$

The Radon-SAD which combines individual projection SADs by means of filtered backprojection is defined: $\mathcal{B}_f\mathcal{S}_Y$. The following demonstrates the application of the shift property ($d_p = \underline{d} \cdot \underline{\xi} = (d_x, d_y)^T \cdot (\cos\phi, \sin\phi)^T$). This is done on an unfiltered back-

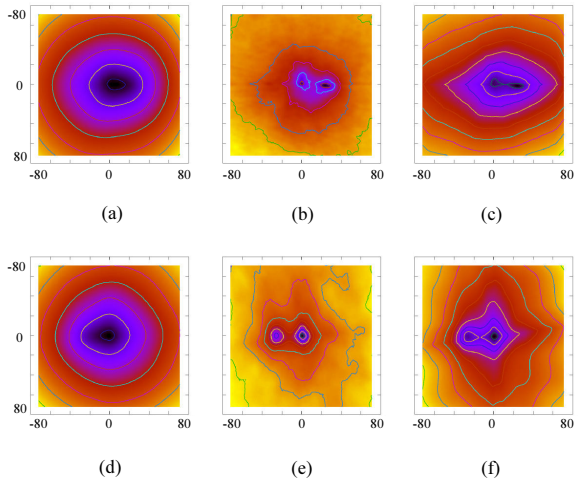

Fig. 4. Global match-profiles for 2 video sequences (top/bottom) with distinct background and foreground motions, showing for a range of shifts values v_x and v_y the SAD value. Backprojections $\mathcal{BS}_Y$ (a, d), filtered Backprojection $\mathcal{B}_f\mathcal{S}_Y$ (b, e) and standard correlation S_{2D} (c, f). Backprojection examples use 50 projection angles.

projection operator for convenience:

$$
\begin{aligned}
& (\mathcal{BS}_Y[L_n, L_{n+1}])(d_x, d_y) \qquad (12)\\
= & \int_0^{\pi} (\mathcal{S}_Y[L_n, L_{n+1}])(d_p[d_x, d_y, \phi], \phi) d\phi \\
= & \int_0^{\pi} \int_{-\infty}^{\infty} |\mathcal{R}L_n(t - d_p[d_x, d_y, \phi], \phi) - \mathcal{R}L_{n+1}(t, \phi))| dt d\phi \\
= & \int_0^{\pi} \int_{-\infty}^{\infty} | \int_{\underline{k} \in l(p,\phi)} L_n(\underline{k} - (d_x, d_y)) - L_{n+1}(\underline{k}) d\underline{k}| dt d\phi
\end{aligned}
$$

The non-linear absolute operator prevents us from proving an (approximate) similarity $[\mathcal{B}_f\mathcal{S}_Y[L_n, L_{n+1}]](v_x, v_y)$ $\approx S_{2D}[L_n, L_{n+1}](v_x, v_y)$. Without the absolute operator we get by the linearity of the Radon transform, the subtraction of 2 (shifted) functions:

$$\mathcal{B}_f(\mathcal{R}f - \mathcal{R}g) = \mathcal{B}_f\mathcal{R}(f - g) = f - g \qquad (13)$$

In practice, it turns out very often that $\mathcal{B}_f\mathcal{S}_Y \approx \mathcal{S}_{2D}$. Figure 4 displays example correlation profiles $\mathcal{BS}_Y$, $\mathcal{B}_f\mathcal{S}_Y$ and $\mathcal{S}_{2D}$ for 2 image sequences with distinct foreground and background motion, which yields 2 clear minima in $\mathcal{B}_f\mathcal{S}_Y$ and $\mathcal{S}_{2D}$. The similarity between $\mathcal{B}_f\mathcal{S}_Y$ and $\mathcal{S}_{2D}$ can be seen in the position and neighborhood of the minima.

We note that in a practical implementation the filtering of projections can have a twofold purpose: (I) providing for a correct implementation of the backprojection algorithm and (II) weighting the frequencies in the image to correlate. Experimentally we found that, depending on the filter type, we can enhance the depth of the correlation peaks for certain video material. In our experiments we use a fixed high-pass filter kernel (-1,-1,0,1,1). In Figure 5 an example of a match-profile is shown with and without high-pass filtering: backprojection of multiple match-profiles enhances the 'true' peaks.

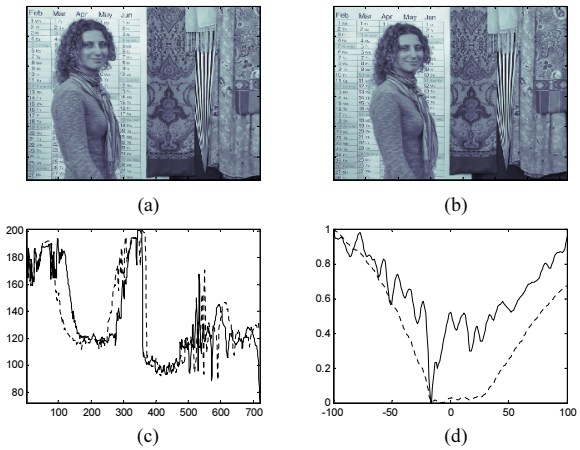

Fig. 5. Dijana sequence, frame $n-1$ (a), frame n (b), unfiltered horizontal projections for both frames (c), match-profiles using $\mathcal{S}_Y$ for unfiltered [dashed] and filtered [solid] projections (d).

4. RESULTS

In this section we evaluate a local backprojection ME implementation with the objective of measuring performance for a varying number of projections. To this end, we replace the frontend part of a phase plane correlation ME implementation (PPC), which we use as our benchmark. Phase plane correlation ME [12] is a 2 step hierarchical block matching method used in professional format converters. It has high sub-pixel accuracy due to normalization of frequencies and an efficient implementation due to the use of FFTs.

PPC can be divided in three parts: (I) For each block a 'correlation surface' is calculated via the frequency domain. (II) In this surface a peak-hunting algorithm finds local maxima corresponding to candidate translational displacements. (III) The candidate motion vectors from II are evaluated for each sub block by means of block matching.

Our backprojection method replaces part I, the 'correlation surface' is now derived by backprojection of matchprofiles.

The method we implemented uses large overlapping square blocks of size 128px with an overlap of 32px at each side and a sub-block size of 8px. We further note that interpolation on the phase plane sequences was implemented using curve fitting, whereas the interpolation in the Radon method was based on interpolation of projections.

The performance criterion that we use, modified mean square error (M2SE), is an extension on MSE which tests the validity of the motion vectors outside the temporal interval in which the motion is estimated [13]. This gives a better performance indication for the 'true' motion. The motion vectors estimated between frame n and $n-1$ are used to calculate motion compensated pixels L_{mc} for frame n using frames $n-1$ and $n+1$. Let $\underline{x} = (x, y)^T$ denote the pixel position, $\underline{D}(\underline{x}, n)$ the local motion vector field and W the measuring window. L_{mc} and M2SE are then defined:

$$L_{mc}(\underline{x}) = \frac{1}{2}(L_{n-1}(\underline{x} - \underline{D}(\underline{x}, n)) - L_{n+1}(\underline{x} + \underline{D}(\underline{x}, n))) \qquad (14)$$

$$M2SE(n) = \frac{1}{|W|} \sum_{\underline{x} \in W} (L_n(\underline{x}) - L_{mc}(\underline{x}))^2 \qquad (15)$$

Table 1 shows the M2SE values for both subpixel and non subpixel accurate methods on a number of synthetic and non-synthetic

Table 1. 5-Frame average M2SE values for both integer and subpixel accurate phase correlation (pp) and projection methods.

	Integer Vector Accuracy						Subpixel Vector Accuracy					
Sequence	pp	2 proj.	4 proj.	8 proj.	16 proj.	32 proj.	pp	2 proj.	4 proj.	8 proj.	16 proj.	32 proj.
camp1	45.7	57.29	54.61	50.91	50.35	49.84	44.35	55.32	52.1	48.78	47.37	47.96
camp2	105.42	145.79	133.34	124.08	123.93	122.28	100.17	138.43	123.68	118.72	118.04	119.26
football	93.62	117.45	107.9	107.08	106.9	107.95	83.46	101.97	98.26	95.39	93.92	93.48
girlsea	91.99	128.74	107.85	105.4	108.51	109.04	89.13	123.85	103.42	102.11	101.99	101.74
lord	139.31	170.77	141.64	137.87	137.38	134.85	137.96	166.46	139.95	137.38	136.58	137.01
lord_d	0	2.5	3.23	0.34	0.55	0.32	0	1.91	2.83	0.41	0.97	0.73
lord_h	0	0.27	1.7	0.57	0.1	0.28	0	0.33	1.77	0.52	0.07	0.05
pan	75.34	94.79	78.05	78.36	77.88	75.75	71.69	91.99	72.21	72.81	73.42	71.52
phantom	51.11	53.91	44.91	35.08	38.34	40.47	49.96	51.88	42.5	33.12	35.94	38.83
pip_fb	46.58	71.94	66.22	58.19	54.51	52.79	46.4	76.95	70.71	59.6	53.8	52.48
porsche	36.69	54.04	49.3	55.02	56.62	54.59	34.6	52.17	44.67	47.72	53.34	52.21
porschefence	49.61	67.55	72.99	75.2	75.81	71.39	47.38	64.28	68.25	71.66	69.82	68.06
renata	91.68	139.23	108.08	87.81	85.51	90.63	84.29	134.67	101.2	82.83	79.76	80.39
startrek	26.63	42.98	43.69	40.65	41.39	41.58	24.99	41.23	39.62	34.91	35.53	35.76
zoom	51.46	56.01	56.21	57.33	55.62	55.6	40.29	45.08	43.91	42.95	42.64	42.56
Average	60.34	80.22	71.31	67.59	67.56	67.16	56.98	76.43	67.01	63.26	62.88	62.80

sequences. Synthetic sequences contain horizontal (lord_h), diagonal (lord_d) and affine motion (zoom). Non-synthetic sequences contain typical film material.

In the comparison we see a steep performance increase related to the number of projection angles which saturates around 8 angles. The PPC method outperforms the Radon based method on average by a small margin (11%), the performance of the latter however much depends on the sequence content. We note that related to the performance there are details which cannot be elaborated within the scope of this paper. Firstly the optimization of the filter used in the Radon method (Section 3). And secondly an effect inherent to the nature of matchprofile backprojection; a single projection orthogonal to the direction of the dominant motion in a block will likely yield the most accurate estimate. Conseqently, the performance could decrease when adding projection angles. We hypothesize that this causes some performance fluctuations in our evaluation. Future research should validate this hypothesis and possibly we can use the effect to our advantage. Finally, we had to leave out a cost analysis of our proposal. We hope to adress this in future publications to improve our benchmark.

5. CONCLUSION

In this paper we present a generalization on direct correlation based projection-methods, which is based on backprojections. We benchmark a local hierarchical block matching algorithm with match-profile backprojections against a, similar hierarchical, phase plane correlation method. The results show that enlarging the number of projection angles increases the performance and brings it close to the performance of phase plane correlation. This establishes the added value of using multiple projection angles. Future work could possibly include extension of this method to global motion and affine models.

6. ACKNOWLEDGEMENTS

The authors would like to thank Guido Janssen for his help with the mathematics of Radon transforms. They also thank Paul Hofman, Nico Cordes and Ralph Braspenning for their ideas, help and support.

7. REFERENCES

[1] J. Kim *et al.*, "Feature-based block matching algorithm using integral projections," *Electronics Letters, IEEE*, vol. 25, no. 1, pp. 29–30, 1989.

[2] J. Kim *et al.*, "A fast feature-based block matching algorithm using integral projections," *IEEE Journal on Selected Areas in Communications*, vol. 10, no. 5, pp. 968–971, 1992.

[3] Y. Fok *et al.*, "An improved fast feature-based block motion estimation," in *Proc. ICIP*, vol. 3, 1994, pp. 741–745.

[4] X. Lu *et al.*, "Improved fast motion estimation using integral projection features for hardware implementation," in *Proc. Int. Symp. on Circuits and Systems*, vol. 2, 1997, pp. 1337–1340.

[5] K. Sauer *et al.*, "Efficient block motion estimation using integral projections," *IEEE Tr. on Circuits and Systems for Video Technology*, vol. 6, no. 5, pp. 513–518, 1996.

[6] J. Lee *et al.*, "Feature-assisted search technique for motion estimation," in *Proc. SPIE*, vol. 5150, 2003, pp. 1503–1512.

[7] A. Crawford, *et al.*, "Gradient based dominant motion estimation with integral projections for real time video stabilisation," in *Proc. ICIP*, vol. 5, 2004, pp. 3371–3374.

[8] D. Robinson *et al.*, "Fast local and global projection-based methods for affine motion estimation," *Journal of Mathematical Imaging and Vision*, vol. 18, no. 1, pp. 35–54, 2003.

[9] S. W. Rowland, *Computer Implementation of Image Reconstruction Formulas*, ser. Topics in Applied Physics, G. T. Herman, Ed. Springer-Verlag, 1979, vol. 32.

[10] S. R. Deans, *The radon transform and some of its applications*. John Wiley & Sons, Inc., 1983.

[11] P. Toft, "The radon transform, theory and implementation," Ph.D. dissertation, Department of Mathematical Modelling, Technical University of Denmark, 1996.

[12] G. A. Thomas, "Television motion measurement for datv and other applications," BBC Research Department, Tech. Rep. PH-283, 1987.

[13] G. de Haan, *et al.*, "True-motion estimation with 3-d recursive search block matching," *IEEE Tr. on Circuits and Systems for Video Technology*, vol. 3, no. 5, pp. 368–379, 1993.