Two-Dimensional Channel Representation for Multiple Velocities

Hagen Spies[1,2] and Per-Erik Forssén[1]

[1] Computer Vision Laboratory
Dept. of Electrical Engineering
Linköping University, 581 83 Linköping, Sweden
{hspies,perfo}@isy.liu.se
[2] ContextVision AB
582 23 Linköping, Sweden

Abstract. We present a two-dimensional information representation, where small but overlapping Gaussian kernels are used to encode the data in a matrix. Apart from points we apply this to constraints that restrict the solution to a linear subspace. A localised decoding scheme accurately extracts multiple solutions together with an estimate of the covariances. We employ the method in optical flow computations to determine multiple velocities occurring at motion discontinuities.

1   Introduction

The channel representation of scalars has proven very useful for many computer vision tasks such as learning and pattern recognition [1; 2]. In this representation a scalar is encoded as a vector of so called channel values, see Sect. 2. This concept is extended to two-dimensional variables here. As we may encounter linear constraints on the two components we need to handle such information as well. We may consider this approach as the sampling of a 2D likelihood function into a channel matrix, with the encoding of the parameter space done in a way that allows us to give an analytic expression for the local peak positions.

The optical flow constraint equation is a typical example that constrains the two velocity components to a linear subspace. Using the developed channel representation we can encode such constraints and by spatial averaging determine the accumulation points of the constraint lines. Thus we can directly represent and determine multiple motions in the averaged neighbourhood.

There have been a number of attempts to determine multiple motions, e.g. [3; 4; 5; 6]. For a discussion of the Fourier properties of multiple motions see [7]. Often some type of filter bank is used where the filter outputs are either combined into a multiple velocity likelihood function or used as separate constraints in an overdetermined system of equations. In the presented approach we use simple derivative filters to yield the constraint equation that is input into our estimation scheme.
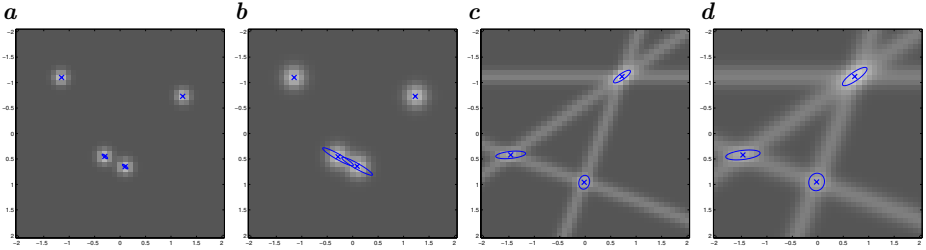
Fig. 1: *Example channel matrices with bin distance 0.1: Encoding 4 points with $\sigma = 0.1$ **a** and $\sigma = 0.15$ **b** . Encoding 4 lines using $\sigma = 0.1$ **c** and $\sigma = 0.15$ **d** .*

2 Two-Dimensional Channel Representation

The channel representation is a monopolar (non-negative) information represen-
tation, where each unit of data is called a *channel*. Each channel value is the
output from a *kernel function*, possibly weighted with some relevance measure.
Here we use Gaussian functions as this results in isotropic 2D kernels. In the 1D
case the scalar value is represented as a vector of channel values, this has been
described in detail in [1; 2].

2.1 Encoding of Vectors

When encoding a two-dimensional vector $[u, v]$ the channel value ϕ at each sam-
pled point $[k, l]$ is obtained by application of the kernel function:

$$\phi(k, l) = \exp\left(-\frac{(u - k)^2 + (v - l)^2}{2\sigma^2}\right) \; . \tag{1}$$

We only discuss the fully isotropic kernel here. Such an encoding of 2D variables
is realised more efficiently as the outer product of the 1D channel vectors for
the two components. For several points the combined channel representation is
simply given by the averaged channel matrix. An example with four thus encoded
points is given in Fig. 1 a and b for different kernel widths σ. Observe that there
is an interference between two points if they are too close to each other. The
number of channels, their distance and the standard deviation of the used kernel
thus limits how many points we can represent simultaneously.

2.2 Encoding of Lines

Often we do not have the exact location of a point but only know that it lies
somewhere on a one-dimensional subspace. We assume that such a linear con-
straint is given either in standard or in normalised form:

$$a\,x + b\,y + c = 0; \quad \cos(\alpha)\,x + \sin(\alpha)\,y - \rho = 0 \; , \tag{2}$$

with $[\cos(\alpha), \sin(\alpha), -\rho] = \frac{1}{\sqrt{a^2+b^2}}[a, b, c]$. The distance of a specific point $[k, l]$ to the line is then given by: $d = \cos(\alpha)\ k + \sin(\alpha)\ l - \rho$. Channel values are obtained by simply applying the basis function:

$$\phi(k, l) = \exp(-\frac{d^2}{2\sigma^2})\ . \tag{3}$$

Each value encodes the likelihood that the line passes through the sampled point. An example with four thus encoded lines is given in Fig. 1 c,d for different σ.

2.3   Decoding

In the decoding step we want to extract the position of local peaks. First we determine local maxima with bin accuracy at $[k, l]$. Then we model the channel values in a small neighbourhood using a 2D Gaussian with centre position $\boldsymbol{s}$, amplitude r and covariance matrix $\boldsymbol{C}$:

$$g(\boldsymbol{x}; r, \boldsymbol{s}, \boldsymbol{C}) = r\ \exp\left(-0.5\ (\boldsymbol{x} - \boldsymbol{s})^T \boldsymbol{C}^{-1}(\boldsymbol{x} - \boldsymbol{s})\right)\ , \tag{4}$$

where $\boldsymbol{x}$ denotes local coordinates. We can express the covariance matrix and its inverse explicitly as:

$$\boldsymbol{C} = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix} \quad \text{and} \quad \boldsymbol{C}^{-1} = \frac{1}{\sigma_x^2\sigma_y^2 - \sigma_{xy}^2} \begin{bmatrix} \sigma_y^2 & -\sigma_{xy} \\ -\sigma_{xy} & \sigma_x^2 \end{bmatrix}\ . \tag{5}$$

For each point in the decoding neighbourhood (e.g. 5×5) we thus obtain one constraint $\phi(\boldsymbol{x}) = g(\boldsymbol{x})$. After taking the logarithm this constraint becomes:

$$\ln g(\boldsymbol{x}) = \ln r - \frac{(x - s_x)^2\sigma_y^2 - 2(x - s_x)(y - s_y)\sigma_{xy} + (y - s_y)^2\sigma_x^2}{2(\sigma_x^2\sigma_y^2 - \sigma_{xy}^2)}\ . \tag{6}$$

This can be written as the scalar product between a known vector $\boldsymbol{a}$ and a unknown parameter vector $\boldsymbol{m}$ with:

$$\boldsymbol{a} = 0.5\left[1\ 2x\ 2y\ -x^2\ -y^2\ -2xy\right]^T\ , \tag{7}$$

$$\boldsymbol{m} = \frac{1}{\sigma_x^2\sigma_y^2 - \sigma_{xy}^2} \begin{bmatrix} 2\ln r(\sigma_x^2\sigma_y^2 - \sigma_{xy}^2) - s_x^2\sigma_y^2 + 2s_x s_y\sigma_{xy} - s_y^2\sigma_x^2 \\ s_x\sigma_y^2 - s_y\sigma_{xy} \\ s_y\sigma_x^2 - s_x\sigma_{xy} \\ \sigma_y^2 \\ \sigma_x^2 \\ -\sigma_{xy} \end{bmatrix}\ . \tag{8}$$

Stacking the constraints for each pixel on top of each other we obtain a least-squares system: $\boldsymbol{Am} = \ln\boldsymbol{\phi}$. The solution is obtained using the pseudo-inverse:

$$\boldsymbol{m} = (\boldsymbol{A}^T\boldsymbol{A})^{-1}\boldsymbol{A}^T\ln\boldsymbol{\phi}\ . \tag{9}$$

We recognise the inverse covariance matrix as:

$$\tilde{\boldsymbol{C}}^{-1} = \begin{bmatrix} m_4 & m_6 \\ m_6 & m_5 \end{bmatrix} \quad \text{thus} \quad \tilde{\boldsymbol{C}} = \frac{1}{m_4 m_5 - m_6^2} \begin{bmatrix} m_5 & -m_6 \\ -m_6 & m_4 \end{bmatrix} . \quad (10)$$

From (8) we find the position and peak amplitude to be given by:

$$[\tilde{s}_x, \tilde{s}_y]^T = \tilde{\boldsymbol{C}}[m_2, m_3]^T ; \; \tilde{r} = \exp\left(0.5(m_1 + m_4 s_x^2 + m_5 s_y^2 - 2m_6 s_x s_y)\right) . \quad (11)$$

The final solution is obtained by adding the bin offset $[u, v] = [k, l] + [s_x, s_y]$. We disregard solutions that fall outside the decoding neighbourhood and set $\tilde{r} = 0$ in such cases. The expectation of the estimated covariance matrix is the sum of the noise covariance $\tilde{\boldsymbol{C}}_n$ and that of the histogram bins $\boldsymbol{C}_b = \text{diag}(\sigma^2, \sigma^2)$. Hence we can compute the covariance matrix of our estimated result to be: $\tilde{\boldsymbol{C}}_n = \tilde{\boldsymbol{C}} - \boldsymbol{C}_b$. Also note that the estimated amplitude $\tilde{r}$ encodes the peak likelihood and thus directly serves as a certainty measure.

The thus estimated peak locations and associated covariance matrices are shown in Fig. 1. For isolated points the covariance vanishes, i.e. we have perfect reconstruction. Increasing the kernel size (σ) leads to stronger interference as can be seen in Fig. 1 b. However the elongated shape of the covariance matrix correctly captures this interference. For linear constraints we find that the intersections are correctly found, see Fig. 1c. Observe that the angle between the lines determines the covariance in the reconstructed point, for 90° we should have an isotropic covariance. However for small values of σ we find a slight anisotropy caused by quantisation effects. For larger σ (Fig. 1 c) this effect disappears. However higher σ values lead to higher uncertainties in the estimates as the peaks (intersections) become less pronounced.

3 Optical Flow

As an example application of the presented framework we now discuss the computation of image velocity. From the assumption of conserved intensity the standard optical flow constraint equation is obtained as:

$$u \, \partial_x \boldsymbol{S} + v \, \partial_y \boldsymbol{S} + \partial_t \boldsymbol{S} = 0 . \quad (12)$$

Here $\boldsymbol{S}$ denotes the signal and $[u, v]^T$ the velocity. As there is only one equation with two unknowns the solution is constrained to lie on a line in the parameter space. This inherent ambiguity is often referred to as the aperture problem. We encode this linear constraint as described in Sect. 2.2.

To obtain a unique solution some form of spatio-temporal smoothness is usually required. Here we simply assume the velocity to be constant in a spatial neighbourhood. The channel matrix for such a neighbourhood is then obtained by averaging the individual matrices. Instead of a standard average it is desirable to give more weight to the central pixel. This is readily achieved by the use of an averaging filter $\boldsymbol{B}$ such as a Gaussian or Binomial. Furthermore we might want

Fig. 2: *Three frames (13,15,17) from the Hamburg taxi sequence.*

to utilise a certainty $\boldsymbol{W}$ at each pixel. The gradient magnitude is a possible choice. In any way this certainty will be zero outside the image thus reducing border effects. The integrated channel matrix is given by a normalised average:

$$\boldsymbol{\Phi'}_{k,l} = \frac{\boldsymbol{B} * (\boldsymbol{W} \cdot \boldsymbol{\Phi}_{k,l})}{\boldsymbol{B} * \boldsymbol{W}} \ , \tag{13}$$

where $*$ denotes convolution. We use the well known Hamburg taxi sequence Fig. 2 to demonstrate the algorithm. The thus computed channel histogram is shown in Fig. 3 b and c for two locations indicated in Fig. 3 a. Note that the channel matrix corresponds to a sampled likelihood function with:

$$p(u, v; \boldsymbol{S}) \simeq \sum_{x,y} b_{x,y} w_{x,y} \exp\left(-\frac{(u\,\partial_x \boldsymbol{S} + v\,\partial_y \boldsymbol{S} + \partial_t \boldsymbol{S})^2}{2\sigma^2((\partial_x \boldsymbol{S})^2 + (\partial_y \boldsymbol{S})^2)} \right) \ . \tag{14}$$

This is in contrast to the standard least squares likelihood function [8]:

$$p(u, v; \boldsymbol{S}) \simeq \exp\left(\frac{-1}{2\sigma^2} \sum_{x,y} b_{x,y} w_{x,y} (u\,\partial_x \boldsymbol{S} + v\,\partial_y \boldsymbol{S} + \partial_t \boldsymbol{S})^2 \right) \ . \tag{15}$$

These two likelihoods are illustrated in Fig. 3 c and d for an area where a moving car is occluded by a tree. We observe that (14) clearly distinguishes the two motions while (15) averages them. The summation in (14) can be thought of as a voting mechanism which makes the approach very robust to outliers, similar to a generalised Hough transform. Note that when there is only one solution and no outliers the expectation value of (14) and (15) coincide.

Points where more than one solution is obtained are indicated in Fig. 3 e. The outlines of two cars are clearly visible. There are no multiple velocities around the bright car as its slow movement can not be separated from that of the background in this case ($\sigma = 0.2$). The certainty of the dominant estimate is shown in Fig. 3 f, this also drops slightly around the brighter car. Finally we show the first and second estimated velocity in Fig. 3 g and h respectively. Around the cars their movements are captured in the less dominant second solution.

It is possible to extract multiple velocities at motion discontinuities. This is illustrated on a synthetic sequence where all four quadrants move in different
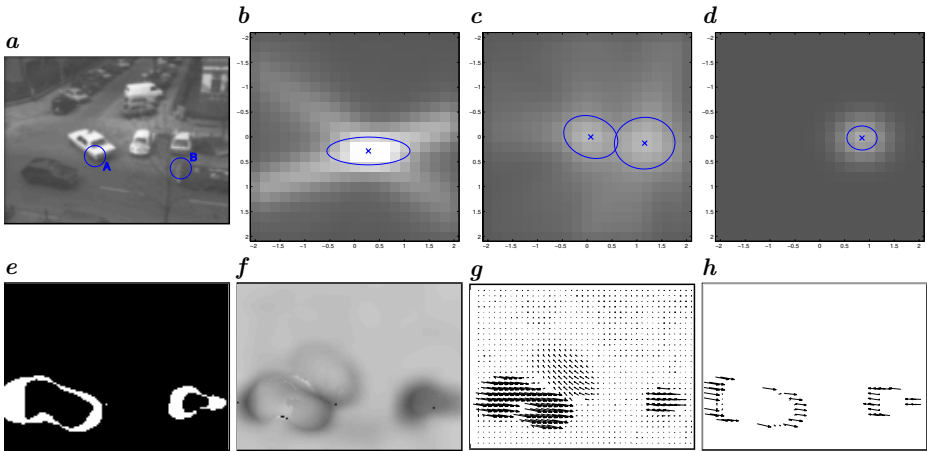
Fig. 3: *Optical flow example:* **a** *frame 15 with marked positions and* **b** *,***c** *channel histogram at positions A and B respectively.* **d** *gives the LS likelihood (15) at point B.* **e** *shows the pixels where more than one solution is obtained,* **f** *gives the confidence in the first estimate and* **g** *and* **h** *show the first and second solution as vector plots.*

directions, an example image is given in Fig. 4 a. The number of solutions is shown in Fig. 4 b; At the centre we get up to four estimates and at the other discontinuities we obtain two solutions. The vector plot (Fig. 4 c) illustrates that the velocities are correctly estimated. The amplitude of the dominant peak drops near the discontinuities as the energy is distributed to several peaks, see Fig. 4 d.

4   Conclusion

We showed how two-dimensional data can be represented in a 2D channel representation. It has been demonstrated how linear constraints can be encoded and that multiple solutions can be extracted. This has been applied to robustly estimate multiple velocities in the presence of occlusions. However, for practical applications we propose to use a simpler optical flow method first and then apply the proposed scheme only at locations where the simple method failed.

Bibliography

[1] Granlund, G.H.: An associative perception-action structure using a localized space variant information representation. In: Proceedings of Algebraic Frames for the Perception-Action Cycle (AFPAC), Kiel, Germany (2000)
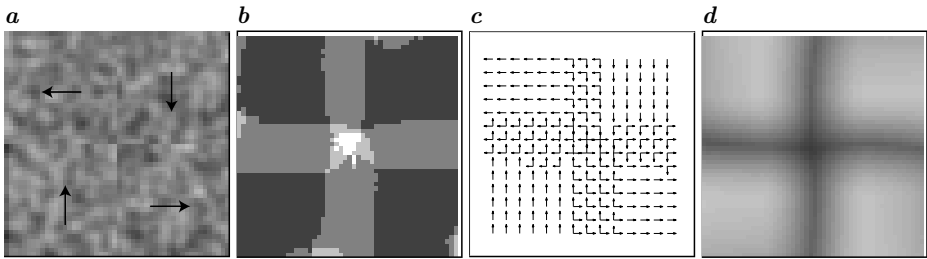
Fig. 4: *Multiple velocities at discontinuities:* **a** *example image,* **b** *number of solutions (range* [0 4]*),* **c** *vector plot and* **d** *confidence (peak amplitude) in the first estimate.*

[2] Forssen, P.E.: Sparse representations for medium level vision. Licentiate Thesis LiU-Tek-Lic-2001:06, Linköping University, Sweden (2001)

[3] Shizawa, M., Mase, K.: Simultaneous multiple optical flow estimation. In: ICPR, Atlantic City, NJ, USA (1990) 274–278

[4] Nestares, O., Navarro, R.: Probabilistic estimation of optical flow in multiple band-pass directional channels. Image and Vision Comp. **19** (2001) 339–351

[5] Mota, C., Stuke, I., Barth, E.: Analytic solutions for multiple motions. In: ICIP. Volume 2., Thessaloniki, Greece (2001) 917–920

[6] Andersson, K., Knutsson, H.: Multiple hierarchical motion estimation. In: SPPRA, Crete, Greece (2002) 80–85

[7] Beauchemin, S.S., Barron, J.L.: On the fourier properties of discontinuous motion. Journal of Mathematical Imaging and Vision **13** (2000) 155–172

[8] Weiss, Y., Fleet, D.: Velocity likelihoods in biological and machine vision. In Rao, R. P. N., Olshausen, B. A., Lewicki, M. S., eds.: Probabilistic Models of the Brain: Perception and Neural Function. MIT Press (2001) 81–100