



Alignment of velocity fields for video surveillance

Jorge S. Marques^{a,*}, Miguel Barão^b, João M. Lemos^c

^aISR, Instituto Superior Técnico, Av. Rovisco Pais, 1049-001 Lisboa, Portugal

^bINESC-ID, Universidade de Évora, R. Alves Redol, 9, 1000-029 Lisboa, Portugal

^cNESC-ID, Instituto Superior Técnico, R. Alves Redol, 9, 1000-029 Lisboa, Portugal

ARTICLE INFO

Article history:

Received 31 October 2010

Available online 18 May 2012

Communicated by S. Sarkar

Keywords:

Image analysis

Image alignment

Motion analysis

Video surveillance

ABSTRACT

Velocity fields play an important role in surveillance since they describe typical motion behaviors of video objects (e.g., pedestrians) in the scene. This paper presents an algorithm for the alignment of velocity fields acquired by different cameras, at different time intervals, from different viewpoints. Velocity fields are aligned using a warping function which maps corresponding points and vectors in both fields. The warping parameters are estimated by minimizing a non-linear least squares energy. Experimental tests show that the proposed model is able to compensate significant misalignments, including translation, rotation and scaling.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Video surveillance systems use multiple cameras as a way to enlarge their field of views and to obtain richer information about the scene. This strategy requires the ability to align images from multiple sources and to compensate for global geometric transformations (Mittal and Huttenlocher, 2000; Foresti et al., 2005; Jwa et al., 2008; Reilly et al., 2010). This is known as the image registration problem.

Image registration has been extensively studied in the last decades (Szeliski, 2006). Most methods align pairs of images by estimating a geometric transformation between their domains. Typically they adopt one of the following strategies (Szeliski, 2006): either they align sparse sets of points (fiducial points) detected in the images or they consider all image pixels (dense information) and align them based on a color constancy hypothesis *i.e.*, they assume that the corresponding points in two images have the same intensity or color (Baker and Matthews, 2004).

While the above algorithms are mostly developed for dealing with static imagery, this paper considers a different problem in which we describe the object motion in the scene by velocity fields and wish to align velocity fields extracted by different cameras from different viewpoints. Thus, we wish to align geometric information describing the motion of objects in the scene, a problem that has received much less attention than image registration in the literature.

A related problem concerns the use of the image gradient in image alignment (e.g., see Pluim et al. (2000), Ruiz-Alzola et al. (2000), Butz and Thiran (2001), Mei and Porikli (2006), Haber

and Modersitzki (2007) and Shams et al., 2007). The gradient direction and orientation carries important information about rotation and scaling and has been used to improve alignment algorithms. In the available literature, several directions have been followed. Some works try to align edge maps using the gradient magnitude (Butz and Thiran, 2001), while others try to reduce the influence of local minima by modifying the cost function (Pluim et al., 2000) or to provide an initial estimate for the rotation angle using an histogram of gradient direction (Shams et al., 2007). These works do not attempt to align vector (gradient) fields but only to improve the accuracy and robustness of intensity alignment algorithms by using additional information (gradient magnitude or orientation). The alignment of the gradient fields is addressed in (Ruiz-Alzola et al., 2000; Haber and Modersitzki, 2007) following two interesting different approaches. The first work extracts a set of local estimates of motion based on a translation model and it is therefore tailored to problems in which the rotation of the images is small. The second work addresses global motion estimation using mutual information applied to the gradient orientation without using the magnitude.

This paper describes a novel algorithm for the alignment of velocity fields based on the minimization of the residue energy. The residue is defined as the difference between transformed velocities at corresponding points. In this case, the color constancy hypothesis does not hold, since the points and velocities are modified by the warping operation. Alignment is performed by solving a non-linear least squares problem using the recursive Gauss–Newton algorithm. Although the method proposed in this paper shares some common aspects with the Lucas–Kanade (LK) algorithm for image alignment, there are three main differences: (i) the proposed method aligns dynamic information (motion fields)

* Corresponding author.

E-mail address: jsm@isr.ist.utl.pt (J.S. Marques).

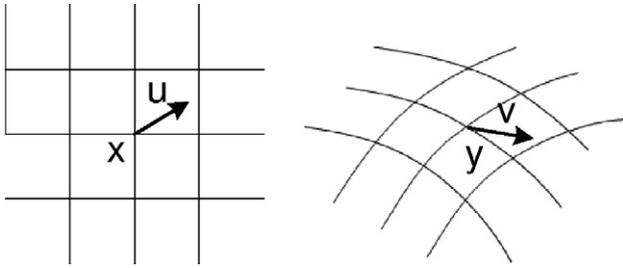


Fig. 1. Warping of points and vectors.

instead of static one (image intensities); (ii) the cost function is different and (iii) the color constancy hypothesis adopted in the LK algorithm is not required.

The paper is organized as follows. Section 2 formulates the problem. Section 3 derives the alignment algorithm. Section 4 presents experimental results and Section 5 draws some conclusions.

2. Problem formulation

2.1. Alignment model

Let us consider a pair of velocity fields $u, v : D \rightarrow \mathbb{R}^2$, with $D \subset \mathbb{R}^2$, obtained by a pair of cameras with different viewpoints. These fields represent the velocity of a point target observed by the two cameras. Since we know the velocity fields, the target trajectories, $x(t), y(t) \in \mathbb{R}^2$, in the two image planes, are described by the differential equations

$$\dot{x}(t) = u(x(t)), \quad \dot{y}(t) = v(y(t)). \tag{1}$$

We will assume that both trajectories are linked by a warping function

$$y = \phi(x, \theta), \tag{2}$$

where $\theta \in \mathbb{R}^m$ is a vector of parameters to be estimated and $\phi : \mathbb{R}^{2+m} \rightarrow \mathbb{R}^2$ is a differentiable function.

We wish to find the best warping function (the best θ) that maps trajectories obtained by the first camera onto trajectories of the second. Since x, y are related by (2) the velocities of the point target in both images are also related. Computing the derivative of y with respect to t we obtain, using (2)

$$v(y) = \frac{\partial \phi}{\partial x}(x, \theta)u(x), \tag{3}$$

where

$$\frac{\partial \phi}{\partial x} = \begin{bmatrix} \frac{\partial \phi_1}{\partial x_1} & \frac{\partial \phi_1}{\partial x_2} \\ \frac{\partial \phi_2}{\partial x_1} & \frac{\partial \phi_2}{\partial x_2} \end{bmatrix} \tag{4}$$

is the Jacobian matrix of the warping function and $\phi = [\phi_1 \ \phi_2]^T$. This is shown in Fig. 1. Points and velocity vectors are transformed in different ways and must obey conditions (2) and (3). This shows that corresponding velocities u, v are different since u is multiplied by the Jacobian matrix.

2.2. Cost functional

Given a warping function $\phi(x, \theta)$, we need to define an alignment measure, a problem similar to the choice of an image alignment measure. There are, however, two important differences. First, we wish to align velocity fields instead of image intensity. Second, the color constancy hypothesis is not assumed to hold. Vectors do not remain invariant under a geometric transformation, since they are modified according to (3) i.e., they are multiplied by the Jacobian matrix.

In this paper, the alignment between two velocity vectors fields at a point $x \in D$ is assessed by computing the *residue vector*

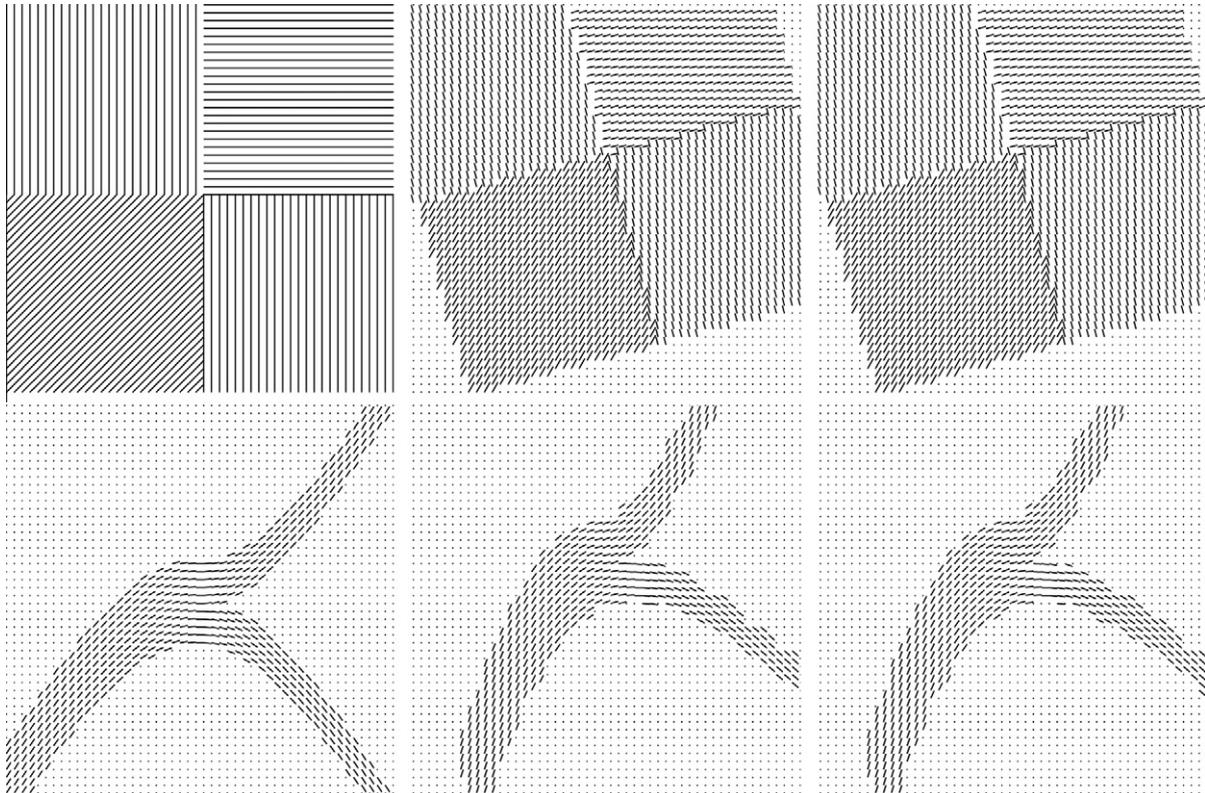


Fig. 2. Alignment results: pairs of misaligned vector fields (1st and 2nd columns) and 1st fields after alignment (3rd column).

$$r(x) = v(y) - \frac{\partial \phi}{\partial x}(x, \theta)u(x), \quad (5)$$

where $y = \phi(x, \theta)$. The alignment measure is the l^2 norm of the residue, considered as a function of θ and given by

$$E(\theta) = \sum_x \left\| v(y) - \frac{\partial \phi}{\partial x}(x, \theta)u(x) \right\|^2, \quad (6)$$

where $\|\cdot\|$ stands for the Euclidean norm and E is called *residue energy*. The sum is performed for all $x \in D$ and such that the transformed position $y = \phi(x, \theta) \in D$. This guarantees that both $u(x)$ and $v(y)$ are defined.

3. Parameter estimation

We wish to find the parameter θ that minimizes E . This is a difficult task since the cost function E is a non-convex function and may have multiple local minima. To overcome this difficulty, we will adopt the Gauss–Newton method for non-linear least squares problems. For that sake, assume that we have an initial estimate θ and allow a small displacement

$$\theta \rightarrow \theta + \Delta\theta. \quad (7)$$

The energy becomes,

$$E(\theta + \Delta\theta) = \sum_x \left\| v(y + \Delta y) - \frac{\partial \phi}{\partial x}(x, \theta + \Delta\theta)u(x) \right\|^2, \quad (8)$$

with $y + \Delta y = \phi(x, \theta + \Delta\theta)$. Then, we approximate the non-convex cost function by a quadratic one replacing the functions by the first order terms of the Taylor series, assuming that $\Delta\theta$ is small. This leads to the system of linear equations (see Appendices A and B)

$$R\Delta\theta = r, \quad (9)$$

$$R = \sum_x \sum_{i=1}^2 R_i(x)^T R_i(x), \quad (10)$$

$$r = \sum_x \sum_{i=1}^2 R_i(x)^T r_i(x), \quad (11)$$

where

$$R_i(x) = \frac{dv_i}{dy}(y) \frac{\partial \phi}{\partial \theta}(x, \theta) - u(x)^T \frac{\partial^2 \phi_i}{\partial \theta \partial x}(x, \theta), \quad (12)$$

$$r_i(x) = \frac{\partial \phi_i}{\partial x}(x, \theta)u(x) - v_i(y). \quad (13)$$

Eqs. (7) and (9) are recursively applied until convergence is achieved. The parameter estimate computed at the k -th iteration will be denoted by $\theta^{(k)}$. Although this algorithm resembles the Lucas–Kanade method for image alignment, it is remarked that it uses dynamic information (velocity fields) instead of image intensity and a different transformation model, given by (3).

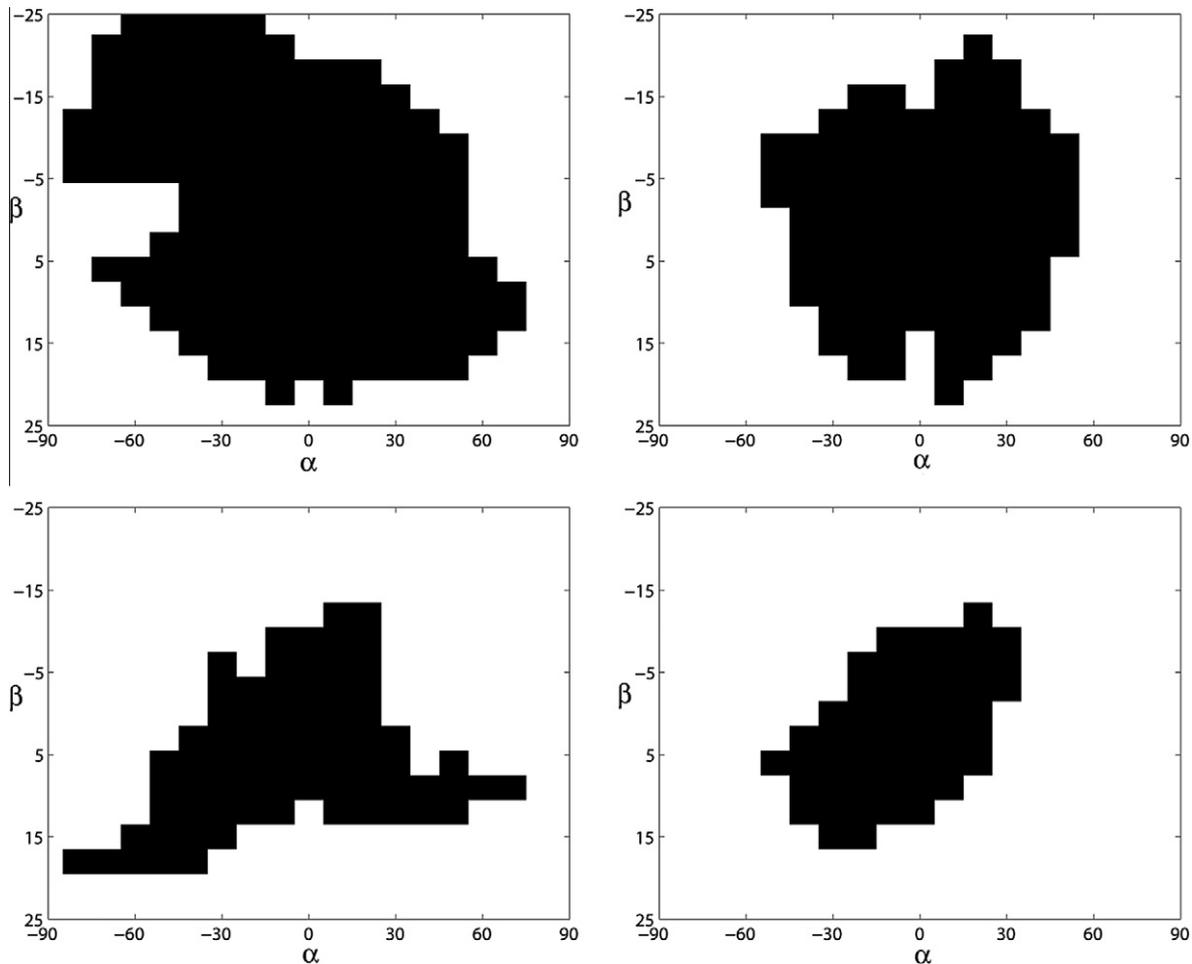


Fig. 3. Attraction region as a function of the rotation angle θ and translation amplitude α for both examples. Translation $t = \alpha \hat{e}$ is performed in two different directions $\hat{e} = [1 \ 0]^T$ (left), $\hat{e} = \frac{1}{\sqrt{2}}[1 \ 1]^T$ (right).

4. Results

This section illustrates the performance of the alignment algorithm with synthetic and real data.

4.1. Synthetic data

We will consider synthetic velocity fields and apply a known geometric transformation. We will then try to compensate the misalignment by using the algorithm proposed in this paper.

Fig. 2 (columns 1 and 2) show two pairs of vector fields to be aligned. Each field is defined on a uniform grid of 50×50 points. The first field (1st column) was analytically defined and the second (2nd column) was obtained by applying a rotation with amplitude $\alpha_0 = 15^\circ$ followed by a translation with displacement $t_0 = [0 \ 5]^T$.

The fields were aligned by the proposed algorithm assuming that the warp function is an affine transformation $\phi(x, \theta) = Ax + t$ with parameters $A \in \mathbb{R}^{2 \times 2}$ and $t \in \mathbb{R}^2$. The alignment algorithm was initialized with the identity transformation ($A^{(0)} = I$, $t^{(0)} = 0$) and the estimates were iteratively updated until convergence is achieved. The third column of Fig. 2 showed the original vector field after the alignment. A comparison between columns 2 and 3 shows

that a very good alignment was achieved by the proposed algorithm in both examples.

The performance of the alignment algorithm can be assessed by objective quality measures. We will adopt the SNR criterion, defined as follows,

$$SNR = 10 \log_{10} \frac{\sum_x \|u(x)\|^2}{\sum_x \|r(x)\|^2} \text{ dB.} \quad (14)$$

The SNR results for the alignment experiments of Fig. 2 are 26.3 dB and 17.1 dB, respectively, stressing the good performance of the proposed algorithm in both examples.

An interesting question is concerned with the basin of attraction of the algorithm and may be phrased as *how much can we misalign the vector fields and still recover?* We assume that the initial warp function is the identity function. Fig. 3 shows the attraction (convergence) regions for both vector fields, considering all rotation angles in the range $[-90^\circ, 90^\circ]$ and translations up to 25 (i.e., half the field support). For the sake of simplicity we assume that $t = \beta \hat{e}$ where \hat{e} is the versor that defines the direction of the translation. Two directions were considered in this experiment and the attraction regions were computed in both cases (see Fig. 3). In the first alignment problem, the attraction regions are large compris-



Fig. 4. Alignment experiment using a pair of surveillance cameras operated at different time intervals: pedestrians trajectories (1st row); motion fields 1, 2 (2nd row) aligned fields 2, 1 (3rd row).

ing rotations in the range $[-45^\circ, 45^\circ]$ and translations in the range $[-15, 15]$. This means that the algorithm recovers large field deformations. In the second example, the attraction regions are smaller since the vector fields are zero in most of the domain and there is less information available. The attraction region can be increased, in both examples, by performing multiple initializations.

4.2. Pedestrian motion

Motion fields have been recently used to characterize pedestrian activities in video sequences (Nascimento et al., 2009). This raises one interesting question. If we have a pair of cameras, is it possible to align the motion fields extracted by each of them? The next example addresses this issue. We consider a pair of cameras observing pedestrians motion from two different viewpoints. We acquired 5186 images with the first camera and 8141 with the second at a frame rate of 25 fps. Then, we extracted the pedestrian trajectories and computed the motion fields associated to both cameras, using the method described in (Nascimento et al., 2009). Then we applied the algorithm proposed in this paper to align both motion fields.

The results are shown in Fig. 4. This first line shows the pedestrian trajectories observed by both cameras. Since the images were acquired at different time instants, is it not possible to establish correspondence between pedestrian trajectories observed by both cameras. The trajectories were lowpass filtered using a Gaussian filter with $\sigma = 2$ pixel. The second line shows the motion fields obtained for each of the cameras and the third line displays the aligned motion fields. The first motion field in line 2 is displayed after the warping in the second column of line 3, and vice versa. Therefore, the comparison should be made between the 2nd and 3rd lines, along the same columns. The proposed algorithm manages to compensate for the flow misalignment in this example. The SNR result achieved by the algorithm in this experiment was 6.0 dB, much less that the values obtained by the same algorithm in the synthetic cases. However, it should be stressed that the field estimates are noisy, since they were computed from a limited number of trajectories, associated to different pedestrians in each image.

5. Conclusion

This paper presents an algorithm for the alignment of velocity fields using parametric warping functions and a non-linear cost functional. This algorithm uses the solution of a non-linear optimization problem which is recursively solved using a Taylor series expansion. Experimental tests with synthetic and real data show the ability of the algorithm to recover from significant misalignments of velocity fields. The algorithm can be easily be extended to include other types of information (e.g., image intensity, color). Future developments should include the alignment of multiple motion fields extracted from video data, representing multiple motion regimes of objects of interest in the scene.

Acknowledgments

We thank the anonymous reviewers for their valuable comments and suggestions. This work was supported by FCT and project PTDC/EEA-CRO/098550/2008.

Appendix A. Minimization of quadratic cost function

First, we will approximate functions $v(\phi(x, \theta + \Delta\theta))$, $\frac{\partial\phi}{\partial x}(x, \theta + \Delta\theta)$ by first order terms of their Taylor expansions. In the first case, this leads to

$$v(\phi(x, \theta + \Delta\theta)) \approx v(\phi(x, \theta)) + \frac{dv}{dy}(\phi(x, \theta)) \frac{\partial\phi}{\partial\theta}(x, \theta)\Delta\theta, \quad (15)$$

where

$$\frac{\partial v}{\partial y} = \begin{bmatrix} \frac{\partial v_1}{\partial y_1} & \frac{\partial v_1}{\partial y_2} \\ \frac{\partial v_2}{\partial y_1} & \frac{\partial v_2}{\partial y_2} \end{bmatrix}, \quad \frac{\partial\phi}{\partial\theta} = \begin{bmatrix} \frac{\partial\phi_1}{\partial\theta_1} & \cdots & \frac{\partial\phi_1}{\partial\theta_m} \\ \frac{\partial\phi_2}{\partial\theta_1} & \cdots & \frac{\partial\phi_2}{\partial\theta_m} \end{bmatrix}. \quad (16)$$

The second function is more difficult to represent since it is a matrix function and the linearization involves a tensor with three indices. To keep the notation simple we will separately consider each component of $\phi(x, \theta)$

$$\phi(x, \theta) = \begin{bmatrix} \phi_1(x, \theta) \\ \phi_2(x, \theta) \end{bmatrix}. \quad (17)$$

In this case

$$\frac{\partial\phi_i}{\partial x}(x, \theta + \Delta\theta) \approx \frac{\partial\phi_i}{\partial x}(x, \theta) + \left(\frac{\partial^2\phi_i}{\partial\theta\partial x}(x, \theta)\Delta\theta \right)^T, \quad (18)$$

where

$$\frac{\partial^2\phi_i}{\partial\theta\partial x} = \begin{bmatrix} \frac{\partial^2\phi_i}{\partial\theta_1\partial x_1} & \cdots & \frac{\partial^2\phi_i}{\partial\theta_m\partial x_1} \\ \frac{\partial^2\phi_i}{\partial\theta_1\partial x_2} & \cdots & \frac{\partial^2\phi_i}{\partial\theta_m\partial x_2} \end{bmatrix}. \quad (19)$$

Let us now replace these expressions in the energy cost functional (8). We obtain

$$E(\theta + \Delta\theta) = \sum_x \left\| v(y + \Delta y) - \frac{\partial\phi}{\partial x}(x, \theta + \Delta\theta)u(x) \right\|^2, \quad (20)$$

$$E(\theta + \Delta\theta) = \sum_x \sum_{i=1}^2 \left(v_i(y + \Delta y) - \frac{\partial\phi_i}{\partial x}(x, \theta + \Delta\theta)u(x) \right)^2, \quad (21)$$

$$E(\theta + \Delta\theta) = \sum_x \sum_{i=1}^2 \left(v_i(\phi(x, \theta)) + \frac{dv_i}{dy}(\phi(x, \theta)) \frac{\partial\phi}{\partial\theta}(x, \theta)\Delta\theta - \frac{\partial\phi_i}{\partial x}(x, \theta)u(x) - u(x)^T \frac{\partial^2\phi_i}{\partial\theta\partial x}(x, \theta)\Delta\theta \right)^2. \quad (22)$$

The energy can now be written as

$$E(\theta + \Delta\theta) = \sum_x \sum_{i=1}^2 (R_i(x)\Delta\theta - r_i(x))^2, \quad (23)$$

where

$$R_i(x) = \frac{dv_i}{dy}(y) \frac{\partial\phi}{\partial\theta}(x, \theta) - u(x)^T \frac{\partial^2\phi_i}{\partial\theta\partial x}(x, \theta), \quad (24)$$

$$r_i(x) = \frac{\partial\phi_i}{\partial x}(x, \theta)u(x) - v_i(y). \quad (25)$$

A necessary condition for minimizing E is

$$\frac{dE}{d\Delta\theta} = 0. \quad (26)$$

Therefore,

$$2 \sum_x \sum_{i=1}^2 R_i(x)^T (R_i(x)\Delta\theta - r_i(x)) = 0, \quad (27)$$

$$\sum_x \sum_{i=1}^2 R_i(x)^T R_i(x)\Delta\theta = \sum_x \sum_{i=1}^2 R_i(x)^T r_i(x). \quad (28)$$

This is a linear system of equations

$$R\Delta\theta = r, \quad (29)$$

where,

$$R = \sum_x \sum_{i=1}^2 R_i(x)^T R_i(x), \quad (30)$$

$$r = \sum_x \sum_{i=1}^2 R_i(x)^T r_i(x). \quad (31)$$

Appendix B. Affine transform

The affine transform maps each point $x \in \mathbb{R}^2$ into a point

$$y = Ax + t, \quad (32)$$

where $A \in \mathbb{R}^{2 \times 2}$ is a square matrix and $t \in \mathbb{R}^2$ a translation vector. The vector of parameters is defined as $\theta = [a_{11} \ a_{12} \ a_{21} \ a_{22} \ t_1 \ t_2]^T$ and the required derivatives are easily obtained

$$\frac{\partial \phi}{\partial x} = A \quad \frac{\partial \phi}{\partial \theta} = \begin{bmatrix} x_1 & x_2 & 0 & 0 \\ 0 & 0 & x_1 & x_2 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0_{2 \times 2} \end{bmatrix}, \quad (33)$$

$$\frac{\partial^2 \phi_1}{\partial \theta \partial x} = [I_{2 \times 2} | 0_{2 \times 2} | 0_{2 \times 2}], \quad \frac{\partial^2 \phi_2}{\partial \theta \partial x} = [0_{2 \times 2} | I_{2 \times 2} | 0_{2 \times 2}]. \quad (34)$$

References

Baker, S., Matthews, I., 2004. Lukas–Kanade 20 years on: a unifying framework. *Internat. J. Comput. Vision* 56, 221–255.

- Butz, T., Thiran, J., Affine registration with feature space mutual information. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2001, pp. 549–556
- Foresti, G.L., Micheloni, C., Snidaro, L., Remagnino, P., Ellis, T., 2005. Active video-based surveillance system: the low-level image and video processing techniques needed for implementation. *IEEE Signal Process. Mag.*, 25–37.
- Haber, E., Modersitzki, J., 2007. Intensity gradient based registration and fusion of multi-modal images. *Methods Inform. Med.*, 292–299.
- Jwa, S., Özgüner, U., Tang, Z., 2008. Information-theoretic data registration for UAV-based sensing. *IEEE Trans. Intell. Trans. Syst.* 9, 5–15.
- Mei, X., Porikli, F., Fast image registration via joint gradient maximization: application to multimodal data. In: *SPIE Conference Electro-Optical and Infrared Systems Technology and Applications*, vol. 6395, 2006.
- Mittal, A., Huttenlocher, D., Scene modeling for wide area surveillance and image synthesis. In: *IEEE Internat. Conf. on Computer Vision and Pattern Recognition*, 2000, pp. 2160–2167.
- Nascimento, J., Figueiredo, M., Marques, J., Trajectory analysis in natural images using mixtures of vector fields. *IEEE International Conference on Image Processing*, 2009, pp. 4353–4356.
- Pluim, J., Maintz, J., Viergever, M., 2000. Image registration by maximization of combined mutual information and gradient information. *IEEE Trans. Medical Imaging* 19, 809–814.
- Reilly, V., Idrees, H., Shah, M., Detection and tracking of large number of targets in wide area surveillance. In: *ECCV*, 2010, pp. 186–199.
- Ruiz-Alzola, J., Westin, C.F., Warfield, S.K., Nabavi, A., Kikinis, R., Nonrigid registration of 3D scalar vector and tensor medical data. In: *MICCAI*, 2000, pp. 541–550.
- Shams, R., Sadeghi, P., Kennedy, R.A., Gradient intensity: a new mutual information-based registration method. In: *IEEE Conf. on Computer Vision and Pattern Recognition*, 2007, pp. 17–22.
- Szeliski, R., 2006. Image alignment and stitching: a tutorial. *Foundations and Trends in Computer Graphics and Computer Vision* 2, 1–104.