June 2021

FUSING RADAR AND VIDEO CAMERA INFORMATION FOR SEVEN DIMENSIONAL SCENE PERCEPTION

David Frakes

Nicholas Gillian

Ivan Poupyrev

Ricardo Motta

Keith Klumb

FUSING RADAR AND VIDEO CAMERA INFORMATION FOR SEVEN DIMENSIONAL SCENE PERCEPTION

BACKGROUND

Historically, data from a radar sensor and image data from a camera may first be processed separately. Following processing, the radar data and the image data may be fused to create an output that can be used for other purposes.

DETAILED DESCRIPTION

As shown in Figure 1, data collected by a radar sensor can first be fused with data collected by a video camera. Following fusion, the fused data may be further processed then output for use by some other software or hardware component of a system. Such an arrangement can be understood as seven dimensional (7D): three dimensional data present in the data from the camera sensor; three dimensional data present in the data from the radar sensor; and the additional dimension of time (from data from both sensors being gathered over time). The radar sensor can be, for example, as described in the journal article *Soli: Ubiquitous Gesture Sensing With Millimeter Wave Radar*, Lien, J., Gillian, N., Karagozler, M., Amihood, P., Schwesig, C., Olson, E., Raja, H., Poupyrev, I., ACM SIGGRAPH 2016, Article 142.

7D motion flow of a scene can be computed using data from a camera and radar senor where pixel-level flow is established through an intelligent interpolation process. The mathematics of optical flow computer the apparent motion of brightness patterns in the two dimensions within a plane with respect to an image. The motions in the physical world that are being characterized are 3D in nature. As detailed herein, motion detected via an optical flow can be decomposed into in-plane and out-of-plane physical motion components.

In some arrangements, the optical flow can be decomposed into divergent and conservative flows. In the context of fluid mechanics, conservative flows correspond to any fluid flows where the fluid of interest is constant in terms of density. This means that given a control volume, the amount of fluid flowing in is equal to the amount of fluid flowing out. Divergent flows correspond to control volumes wherein the volume flowing into a control volume does not equal the volume flowing out. As an example of this, Figure 2 shows an instance in time showing the conservative nature of the flow: what flows into the volume within the image where the object is moving also flows out.

For example, consider a camera videoing a ball attached to a rope, swinging side-to-side in front of the camera and radar sensor. The ball would be moving in-plane with respect to the imaging plane. The optical flow that results is conservative. In this case, the corresponding complex range Doppler plot would illustrates little or no motion toward or away from the radar sensor.

In contrast, consider a second example in which the ball attached to the rope is swinging toward and away from the video camera and radar sensor. The ball would be moving almost entirely divergent – that is, what flows into the volume within the image where the object is moving does not flow out, or vice versa. In this example, the corresponding complex range Doppler plot, as shown in Figure 3, illustrates considerable motion toward and away from the radar sensor indicative of a divergent flow.

Conservative cases and divergent cases of optical flow correspond directly to cases of in-plane and out-of-plane motion as well as radar signals in the Complex Range Doppler domain that are indicative of the same motion. This fact allows the use of divergence of optical flow as a correlation for peaks in the radar Complex Range Doppler signal toward establishing true

velocity of objects in monocular video that are also sensed with the radar sensor. It is noteworthy that the crosstalk between a camera and a radar sensor need not necessarily be one-way; while the radar sensor can "learn" from the camera, the opposite can also be true. That is, the camera may be able to provide information to the component processing the data from the radar sensor indicative of which object it is that is moving in a certain way. At the same time, the radar sensor may provide information to the component processing the camera data indicative of what the depth of that object is. Such an arrangement allows for quantification of the true physical velocity within the imaging plane (as opposed to a strictly pixel-based metric).

Toward this end, deeper consideration of the structure of optical flow calculation is relevant. A formulation of optical flow is as follows:

$$I(x+\delta x, y+\delta y, t+\delta t) = I(x,y,t) + \frac{\partial I}{\partial x}\delta x + \frac{\partial I}{\partial y}\delta y + \frac{\partial I}{\partial t}\delta t + H.O.T.$$

$$\frac{\partial I}{\partial x}\delta x + \frac{\partial I}{\partial y}\delta y + \frac{\partial I}{\partial t}\delta t = 0$$

$$\frac{\partial I}{\partial x}\frac{\delta x}{\delta t} + \frac{\partial I}{\partial y}\frac{\delta y}{\delta t} + \frac{\partial I}{\partial t}\frac{\delta t}{\delta t} = 0$$

$$\frac{\partial I}{\partial x}V_x + \frac{\partial I}{\partial y}V_y + \frac{\partial I}{\partial t} = 0$$

As indicated above, the partial derivative terms are straight forward with respect to an image I, the spatial coordinates within the image being x and y, and the time variable t indexing images at different times. V_x and V_y represent the "velocities" of brightness patterns in the x and y directions from an image at one time to an image at another.

In the context of correlating optical flow with a radar sensor, an opportunity may be to decompose the V_x and V_y flows into conservative and divergent components as described previously. Such an arrangement would afford a connection of the divergent (i.e., out-of-plane)

components of the optical flow signal with the true velocity signals measured by the radar sensor.

Decomposition of the optical flow field into divergent and otherwise components may be possible via projection onto radial basis functions, which are inherently divergence free. Residuals might then represent the in-plane motion components, with the aforementioned projections accounting for out-of-plane motion.

A more efficient way of evaluating the divergence of optical flow field segments may be adaptive control grid interpolation (ACGI). ACGI uses the optical flow equation as the objective function such that minimizing the associated error drives optimization. Instead of optimizing on the pixel level, however, ACGI minimizes the error associated with the optical flow equation over blocks wherein pixel level flow vectors relate to the flow vectors at the corners of any given block via bilinear interpolation. In addition to the inherent regularization afforded by the interpolation, ACGI makes possible the assessment of motion field divergence at multiple scales given that the algorithm starts at the image level then addresses progressively smaller and smaller blocks as defined by quadtree subdivision, as long as the corresponding error associated with the optical flow equation decreases. The conventional formulation is as follows (note the notation change from x, y, t to n1, n2, k):

$$I[n_1, n_2, k] = I\left(n_1 + d_1[n_1, n_2, k], n_2 + d_2[n_1, n_2, k], k + \delta k\right)$$

$$I[\mathbf{n}, k] \approx I[\mathbf{n}, k + \delta k] + \frac{\partial I[\mathbf{n}, k + \delta k]}{\partial n_1} d_1(\mathbf{n}) + \frac{\partial I[\mathbf{n}, k + \delta k]}{\partial n_2} d_2(\mathbf{n})$$

$$d_1(\mathbf{n}) = \sum_{i=1}^{p} \alpha_i \theta_i(\mathbf{n}) \quad d_2(\mathbf{n}) = \sum_{i=1}^{p} \beta_i \phi_i(\mathbf{n})$$

Here, the basis functions are defined as follows, wherein variables *n* with sub- and superscripts correspond row and column coordinates of the corners of a block region:

$$\theta_1(n_1,n_2)=\phi_1(n_1,n_2)=\left(\frac{n_1^2-n_1}{n_1^2-n_1^1}\right)\left(\frac{n_2^2-n_2}{n_2^2-n_2^1}\right)$$
$$\theta_2(n_1,n_2)=\phi_2(n_1,n_2)=\left(\frac{n_1^2-n_1}{n_1^2-n_1^1}\right)\left(\frac{n_2-n_2^1}{n_2^2-n_2^1}\right)$$
$$\theta_3(n_1,n_2)=\phi_3(n_1,n_2)=\left(\frac{n_1-n_1^1}{n_1^2-n_1^1}\right)\left(\frac{n_2^2-n_2}{n_2^2-n_2^1}\right)$$
$$\theta_4(n_1,n_2)=\phi_4(n_1,n_2)=\left(\frac{n_1-n_1^1}{n_1^2-n_1^1}\right)\left(\frac{n_2-n_2^1}{n_2^2-n_2^1}\right)$$

Expressing the displacement functions using vector notation yields:

$$d_1(\mathbf{n})=\overline{\alpha}^T\overline{\theta}(\mathbf{n}) \quad d_2(\mathbf{n})=\overline{\beta}^T\overline{\phi}(\mathbf{n})$$

The approximate optical flow equation and associated error expression then become:

$$I[\mathbf{n},k]\approx I[\mathbf{n},k+\delta k]+\frac{\partial I[\mathbf{n},k+\delta k]}{\partial n_1}\overline{\alpha}^T\overline{\theta}(\mathbf{n}) + \frac{\partial I[\mathbf{n},k+\delta k]}{\partial n_2}\overline{\beta}^T\overline{\phi}(\mathbf{n})$$

$$\widetilde{E}(\overline{\alpha},\overline{\beta})=\sum_{\mathbf{n}\in R}\left(I[\mathbf{n},k]-I[\mathbf{n},k+\delta k] + \frac{\partial I[\mathbf{n},k+\delta k]}{\partial n_1}\overline{\alpha}^T\overline{\theta}(\mathbf{n})+\frac{\partial I[\mathbf{n},k+\delta k]}{\partial n_2}\overline{\beta}^T\overline{\phi}(\mathbf{n})\right)^2$$

Using a control grid interpolator in the proposed context makes sense because at different scales and within every block segment of the image, decomposition of the motion field segments into conservative and divergent components is efficient. Finding the translational bias or offset that minimizes the vector magnitude associated with the four corner point motion

vectors corresponding to a block leaves behind the divergent component, which may be key to correlating with radar sensor data peaks.

To gather training data, a robotic actuator rig, as shown in Figure 4, can be used to gather radar and optical data using a radar sensor and video camera, respectively. This data can be used for testing, verification, and, potentially, training.

ABSTRACT

Radar sensor data can be fused with video camera data. Following fusion, the fused data may be further processed then output for use by some other software or hardware component of a system. Such an arrangement can be understood as seven dimensional (7D): three dimensional data present in the data from the camera sensor; three dimensional data present in the data from the radar sensor; and the additional dimension of time (from data from both sensors being gathered over time).

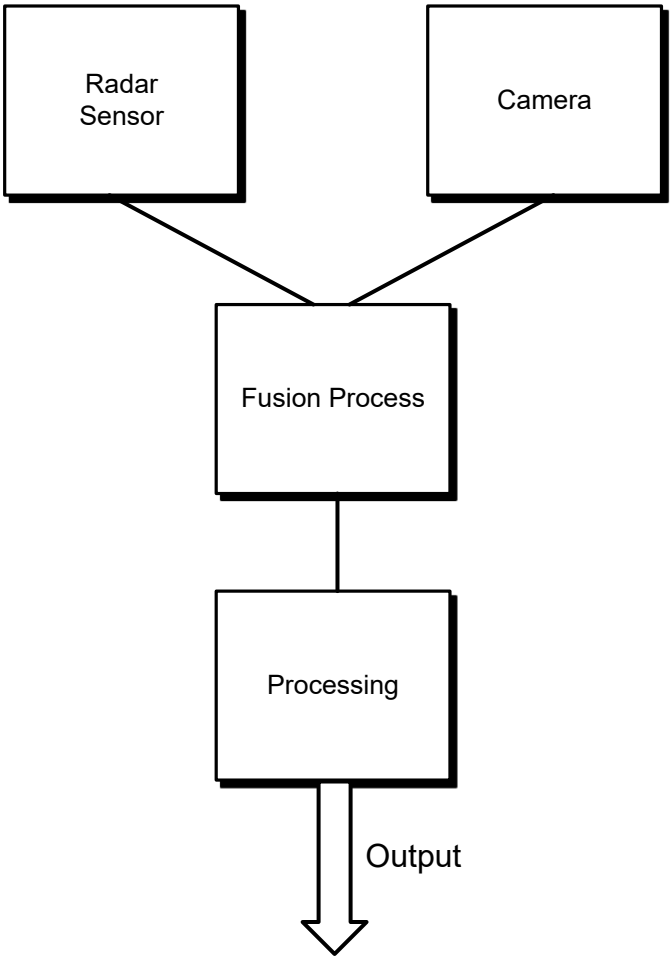

FIG. 1

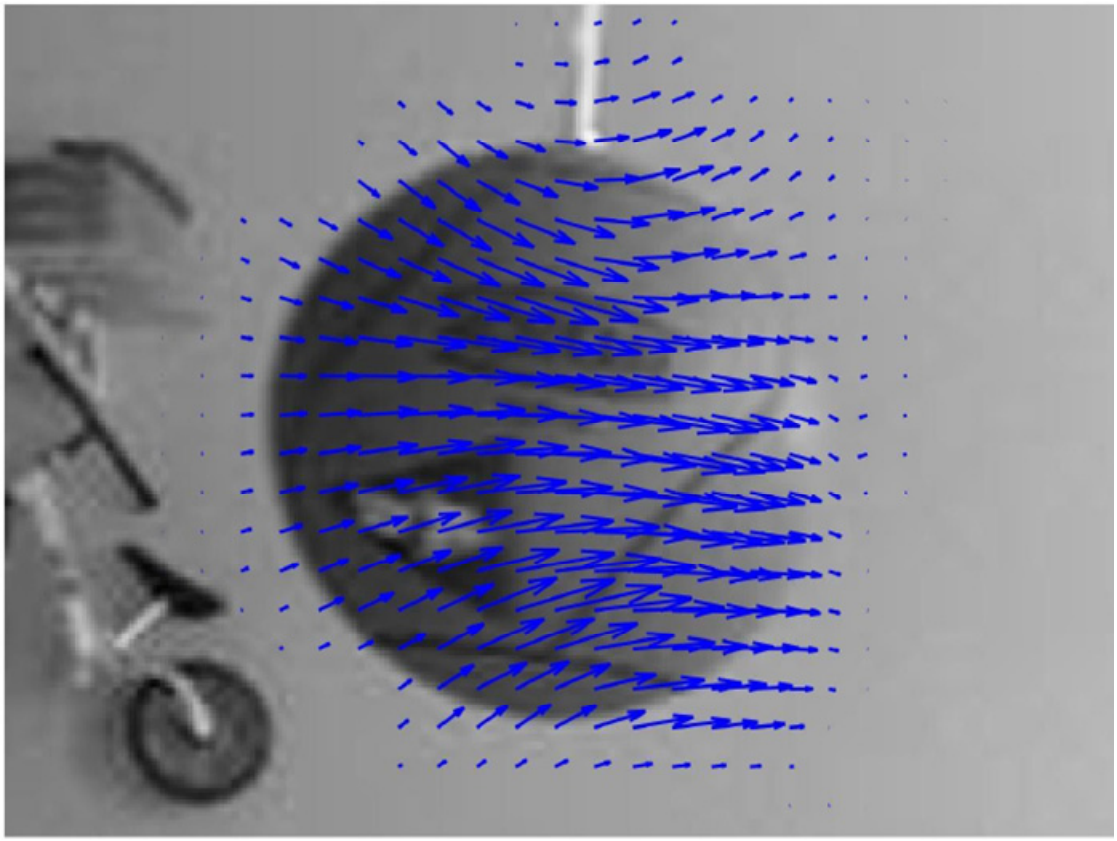

FIG. 2

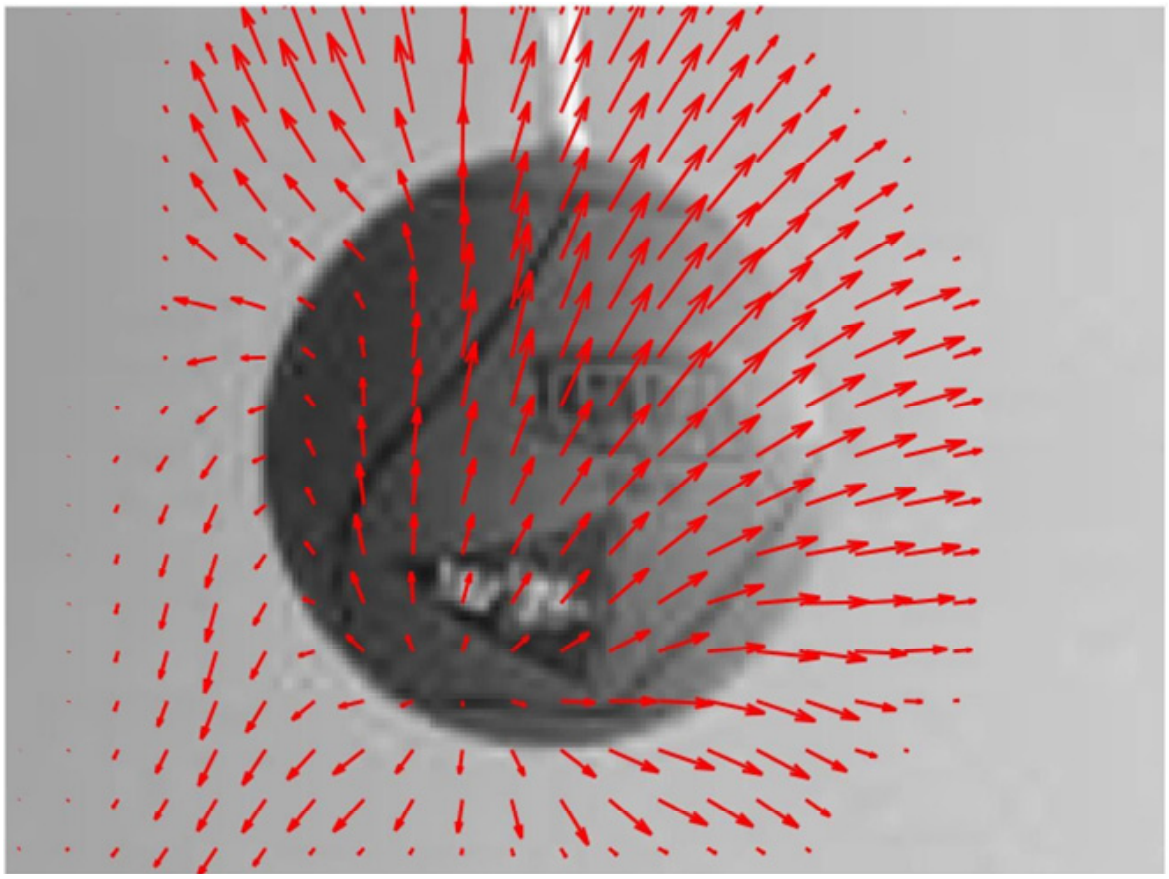

FIG. 3

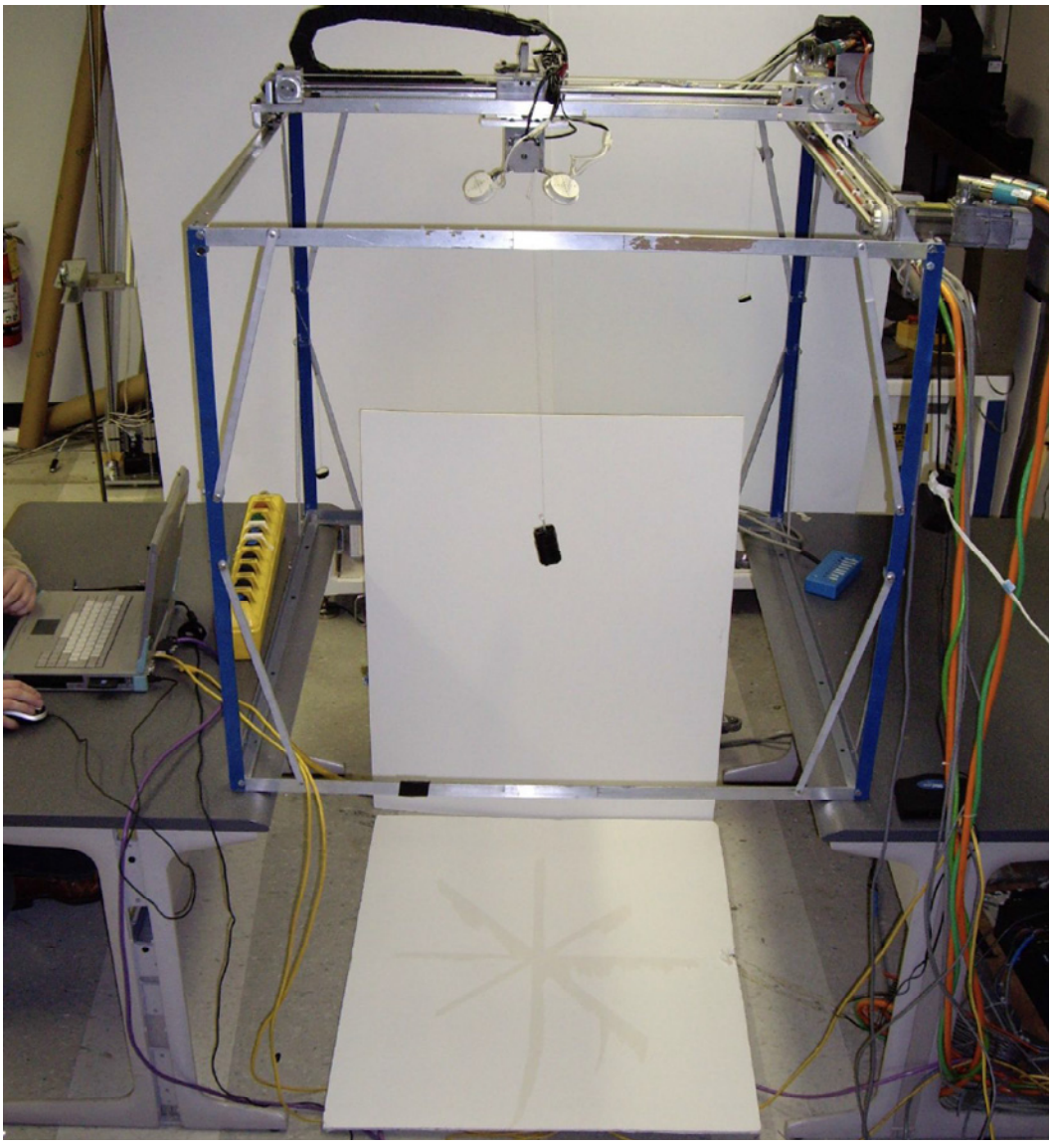

FIG. 4