Motion estimation based on the energy flow*

ZHANG Zhongwei, LIU Guizhong**, LI Hongliang and LI Yongli

(School of Electronics and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

Received June 12, 2002; revised October 28, 2002

Abstract By taking pixels of image sequences as moving gas molecules, a novel concept, image temperature, is proposed to describe the natural property of the image motion. The idea comes from the revelation of the Maxwell-Boltzmann Distribution Law in gas dynamic theories. Furthermore, another concept of energy flow corresponding to the optical flow is developed, and the method of the energy flow equation (EFE) is established to estimate image motion. The experiment indicates a better performance of the proposed EFE scheme with significantly reduced false motion estimates when compared to the traditional optical flow equation (OFE).

Keywords: OFE, gas dynamic theories, image temperature, EFE, motion estimation.

The optical flow has been defined as a convenient representation for image motion or the displacement field for any pixel in an image sequence. A common assumption in measuring image motion is that the intensity structures of local time-varying image regions are approximately constant under motion for at least a short duration. Let $S(x, y, t)$ denote the image intensity function, and an optical flow equation can be derived based on this hypothesis.

$$\frac{\mathrm{d}S(x,y,t)}{\mathrm{d}t} = 0, \tag{1}$$

where x and y vary by t according to the motion trajectory. Eq. (1) is a total derivative expression and denotes the rate of change of intensity along the motion trajectory. Using the chain rule of differentiation, Eq. (1) can be expressed as

$$\frac{\partial S(x,y,t)}{\partial x}u + \frac{\partial S(x,y,t)}{\partial y}v + \frac{\partial S(x,y,t)}{\partial t} = 0, \tag{2}$$

where $\frac{\partial S(x,y,t)}{\partial x}$ and $\frac{\partial S(x,y,t)}{\partial y}$ denote the horizontal and vertical spatial gradients of the image intensity respectively, $\frac{\partial S(x,y,t)}{\partial t}$ denotes the temporal gradient of the image intensity, $u = \frac{\mathrm{d}x}{\mathrm{d}t}$ the horizontal image velocity or displacement, and $v = \frac{\mathrm{d}y}{\mathrm{d}t}$ the vertical one correspondingly.

However, the optical flow equation (OFE) has been directly derived from the liquid flow equation which describes the liquid density variation during its flowing[1]. OFE would produce many false motion estimates, which cannot be detected by naked eyes, and reduces the precision of motion estimates. The purpose of this paper is to develop a new physical model to describe the image motion, which substitutes for the traditional optical flow model. In this paper, we first put forward a novel concept, image temperature, then based on the concept, an equation which is defined as the energy flow equation (EFE) is developed to compute the image motion field. The idea comes from the gas molecule distribution law in the gas dynamic theories. By taking pixels of image sequences as moving gas molecules, we deduce a formula to compute the image temperature. Based on the formula, we can change the luminance image data, which is our experimental input data. Based on the analytic method of the optical flow, we develop the concept of energy flow and deduce the EFE. By using the proposed EFE based on OFE scheme, real motion vectors are obtained and false motions are reduced significantly.

1 Gas dynamic theories and the Maxwell-Botzmann distribution law

Maxwell studied gas molecule distributing characters in 1859, and deduced the gas molecule velocity distributing function in the equilibrium state[2],

$$f(v) = 4\pi\left(\frac{m}{2\pi kT}\right)^{\frac{3}{2}} v^2 \mathrm{e}^{\frac{-mv^2}{2kT}}, \tag{3}$$

where m denotes the molecule mass, v the average

* Supported by the Education Ministry of China (Grant No. 200069828), the National Natural Science Foundation of China (Grant No. 60272072)
** E-mail: liugz@mail.xjtu.edu.cn

molecule velocity, T the gas temperature, and k the Boltzmann constant. This equation is the Maxwell distribution law that describes the essential property of gas molecule movement.

Based on the Maxwell distribution law, Boltzmann studied the distribution of micro-particles with discrete energy. He deduced the formula, called Maxwell-Boltzmann distribution law, which can compute micro-particle number with different discrete energies. Suppose micro-particles can only own a series of discrete energies and their values are orderly ε_1, $\varepsilon_2, \cdots, \varepsilon_i, \cdots$, so the number of micro-particles with energy ε_i is

$$N_i = C \cdot e^{-\varepsilon_i/kT}, \tag{4}$$

where C is a constant not related to ε_i.

The average kinetic energy of an ideal gas molecule is

$$\varepsilon_i = \frac{1}{2} m v_i^2. \tag{5}$$

From Eqs. (3), (4) and (5), we know that Eqs. (3) and (4) are all bell-shaped functions of velocity v_i.

The Maxwell-Boltzmann distribution law quantitatively represents the essential property of gas molecule movement, and it is applied very widely to solid physics, laser and contemporary physics.

2 Image temperature

Based on the gas dynamic theories, the gas temperature is the result of molecules moving randomly and is determined by the number of gas molecules and their average kinetic energy. The latter depends on their motion velocity. Supposing that pixels of an image sequence and gas molecules have similar property on random movement, and these pixels can move ceaselessly like gas molecules. The Maxwell-Boltzmann distributing law can also describe the statistic property of pixels motion of image sequences.

Pixels of image sequences would be seen as the moving gas molecules, and we believe that any position of an image in video would own a certain "temperature", and the temperature will be defined as "image temperature". The image temperature is determined by the kinetic energy of the pixel and the number of the pixels with the same kinetic energy corresponding to the definition of the gas temperature. The kinetic energy of the pixel depends on the motion velocity of the pixel. We now develop the equation for computing the image temperature. Based on Eqs. (4) and (5), we obtain

$$N_i = C \cdot e^{-\frac{mv_i^2}{2kT}}, \tag{6}$$

to proceed, we have

$$T = \frac{\frac{1}{2} m v_i^2}{k(\ln C - \ln N_i)}, \tag{7}$$

where m denotes the pixel "mass". Suppose that all pixels mass are identical, and setting m as one), v_i denotes the pixel motion velocity equals to the pixel gray value, N_i the number of the pixels whose gray value equals v_i within a certain time interval t[3]. In this paper, the time interval t refers to the duration of a GOP, k and C are constants.

From Eq. (7), it can be seen that, the temperature of any position in an image is determined by the motion velocity of the pixel at that position and the number of the pixels with the same motion velocity v_i within the time interval t. So the temperature image sequence can be obtained from Eq. (7).

3 Energy flow equation

As we stated above, any position of an image corresponds to a certain temperature, so the following hypothesis of the image motion estimation is reasonable. Suppose that the image temperature structures of local time-varying image regions are approximately constant under motion for at least a short period. Let $T(x, y, t)$ denote the continuous spatial-temporal temperature distribution. If the temperature remains constant along a motion trajectory, we have

$$\frac{\mathrm{d}T(x, y, t)}{\mathrm{d}t} = 0, \tag{8}$$

where x and y vary by t according to the motion trajectory. Eq. (8) is a total derivative expression and denotes the rate of change of image temperature along the motion trajectory. The remainder of this paper will deduce the equation for computing energy flow fields.

At time t, the image temperature at the point (x, y) of an image is $T(x, y, t)$. At time $t + \Delta t$, the point will move to a new position $(x + \Delta x, y + \Delta y)$, its image temperature is $T(x + \Delta x, y + \Delta y, t + \Delta t)$. Supposing that $T(x + \Delta x, y + \Delta y, t + \Delta t)$ is equal to $T(x, y, t)$, we have

$$T(x, y, t) = T(x + \Delta x, y + \Delta y, t + \Delta t). \tag{9}$$

Expanding the right part of Eq. (9) by Taylor expansion, and ignoring the terms exceeding the square, we have

$$\frac{\partial T(x,y,t)}{\partial x}\cdot u+\frac{\partial T(x,y,t)}{\partial y}\cdot v+\frac{\partial T(x,y,t)}{\partial t}=0, \tag{10}$$

where $\frac{\partial T(x,y,t)}{\partial x}$ and $\frac{\partial T(x,y,t)}{\partial y}$ denote the horizontal and vertical spatial gradients of the image temperature respectively, $\frac{\partial T(x,y,t)}{\partial t}$ is the temporal gradient of the image temperature, $u=\frac{\mathrm{d}x}{\mathrm{d}t}$ and $v=\frac{\mathrm{d}y}{\mathrm{d}t}$ are the horizontal and vertical temperature image velocity displacements, respectively, and $T(x,y,t)$ is the image temperature at the position (x,y) obtained from Eq. (7). Because the temperature is the measure of gas molecule kinetic energy in gas dynamic theories, Eq. (8) can be called the energy flow equation (EFE) or the energy flow constraint.

4 Motion estimation by EFE

Estimating the optical flow is a fundamental issue in the image sequence processing, and there are many different methods to estimate the optical flow. Because there are two variables of the optical flow field in an expression, solving Eq. (2) is an "ill-posed" problem, and we can only compute the optical flow field along the gradient direction.

Regularization (a smoothness constraint) by requiring a slowing varying optical flow field was first introduced by Horn and Schunck[4] to solve the OFE of Eq. (2). In the implementation of Horn and Schunck method, the Laplacians of the velocity components have been approximated by FIR high-pass filters to arrive at a Gauss-Seidel iteration. In order to compare the energy flow field with the optical flow field impartially, solving the EFE and the OFE simultaneously uses the Horn and Schunck's method.

The Gauss-Seidel iterative method, Eq. (11), is used in implementing the Horn and Schunck method[4] to solve the energy flow based on the data from two consecutive frames.

$$u^{(n+1)}=\bar{u}^{(n)}-\frac{\partial T}{\partial x}\cdot\frac{\frac{\partial T(x,y,t)}{\partial x}\bar{u}^{(n)}+\frac{\partial T(x,y,t)}{\partial y}\bar{v}^{(n)}+\frac{\partial T(x,y,t)}{\partial t}}{\alpha^2+\left(\frac{\partial T(x,y,t)}{\partial x}\right)^2+\left(\frac{\partial T(x,y,t)}{\partial y}\right)^2},$$

$$v^{(n+1)}=\bar{v}^{(n)}-\frac{\partial T}{\partial y}\cdot\frac{\frac{\partial T(x,y,t)}{\partial x}\bar{u}^{(n)}+\frac{\partial T(x,y,t)}{\partial y}\bar{v}^{(n)}+\frac{\partial T(x,y,t)}{\partial t}}{\alpha^2+\left(\frac{\partial T(x,y,t)}{\partial x}\right)^2+\left(\frac{\partial T(x,y,t)}{\partial y}\right)^2}, \tag{11}$$

where n is the iteration counter, the overbar denotes weighted local averaging, and all partials are evaluated at the point (x,y,t). The initial estimates of velocities $u^{(0)}$ and $v^{(0)}$ are usually taken as zero.

Eq. (11) assumes a continuous spatial-temporal temperature distribution. In computer implementation, all spatial and temporal gradients need to be estimated from the discrete temperature image data obtained from Eq. (7). Both average finite differences and polynomial fitting have approximated the spatial and temporal gradients. We make use of the averaging four finite differences to obtain the estimation of spatial and temporal gradients in Eq. (11). For two consecutive frames k and $k+1$, the image temperature gradients are estimated as

$$\begin{aligned}\frac{\partial T(x,y,t)}{\partial x}\approx\frac{1}{4}\{&T(i+1,j,k)-T(i,j,k)\\&+T(i+1,j+1,k)-T(i,j+1,k)\\&+T(i+1,j,k)-T(i,j,k+1)\\&+T(i+1,j+1,k+1)-T(i,j+1,k+1)\},\end{aligned}$$

$$\begin{aligned}\frac{\partial T(x,y,t)}{\partial y}\approx\frac{1}{4}\{&T(i,j+1,k)-T(i,j,k)\\&+T(i+1,j+1,k)-T(i+1,j,k)\\&+T(i,j+1,k)-T(i,j,k+1)\\&+T(i+1,j+1,k+1)-T(i+1,j,k+1)\},\end{aligned}$$

$$\begin{aligned}\frac{\partial T(x,y,t)}{\partial t}\approx\frac{1}{4}\{&T(i,j,k+1)-T((i,j,k)\\&+T(i+1,j,k+1)-T(i+1,j,k)\\&+T(i,j+1,k+1)-T(i,j+1,k)\\&+T(i+1,j+1,k+1)-T(i+1,j,k)),\end{aligned} \tag{12}$$

and the local averages $\bar{u}$ and $\bar{v}$ are estimated as

$$\begin{aligned}\bar{u}(i,j,t)=\frac{1}{6}\{&u(i-1,j,k)-u(i,j+1,k)\\&+u(i+1,j,k)-u(i,j-1,k)\}\\&+1/12\{u(i-1,j-1,k)\\&-u(i-1,j+1,k)\\&+u(i+1,j-1,k+1)\\&-u(i+1,j+1,k)\},\end{aligned}$$

$$\begin{aligned}\bar{v}(i,j,t)=\frac{1}{6}\{&v(i-1,j,k)-v(i,j+1,k)\\&+v(i+1,j,k)-v(i,j-1,k)\}\end{aligned}$$

$$
\begin{aligned}
&+ 1/12\{v(i-1,j-1,k) \\
&- v(i-1,j+1,k) \\
&+ v(i+1,j-1,k+1) \\
&- v(i+1,j+1,k)\}. \qquad (13)
\end{aligned}
$$

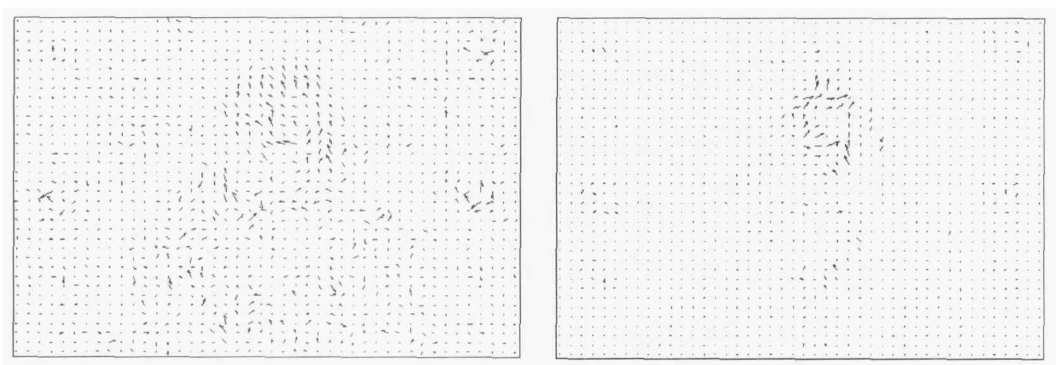
This iterative process stops when the energy flow counted in the current iteration does not change much from that of a previous iteration, e.g. the Maximum Square Error (MSE) or Maximum Absolute Difference (MAD) less than some threshold value or the number of iterations reaches a certain value, i.e. the maximum iteration count.

5 Experiment

EFE and OFE have been used to estimate the motion field between the 488th and the 489th frames of a progressive video, known as the "News" sequence and shown in Fig. 1 (a) and (b), respectively. Fig. 1(c) shows the absolute value of the frame difference (multiplied by 3) without any motion compensation, which is to indicate the amount of motion present. The lighter pixels are those whose intensity has changed with respect to the pervious frame due to the motion. Indeed, the scene contains multiple motions: the speaker is raising her head (from down to up), her mouth is opening, and the background monitors at up-left and down-right are changing. The motion fields estimated by the OFE and the EFE are depicted in Fig. 2 (a) and (b), respectively.

(a) (b) (c)

Fig. 1. Experimental image sequence and frame difference (a) The 488th intensity image; (b) the 489th intensity image; (c) absolute value of difference image of image (a) and image (b)



(a) (b)

Fig. 2. Motion fields estimated from Fig. 1 (a) Motion field estimated by OFE; (b) motion field estimated by EFE.

Fig. 2(b) shows that the motion field estimated by EFE fits the actual motion more accurately, and avoids much of the false motion estimated by OFE in Fig. 2(a), which cannot be detected by the naked eyes.

We evaluate the goodness of the motion estimates on the basis of the peak signal-to-noise ratio (PSNR) of the resulting displaced frame difference (DFD) between the 488th and 489th frames, defined by

$$\mathrm{PSNR} = 10\log_{10}\frac{255 \times 255}{\sum(|S_{489}(i,j) - S_{488}(i + d_1(i,j), j + d_2(i,j))|)}, \qquad (14)$$

where d_1 and d_2 denote components of the motion estimates at every pixel. We also compute the entropy of the estimated 2-D motion field, given by

$$H = -\sum P(d_1)\log_2 P(d_1) - \sum P(d_2)\log_2 P(d_2), \qquad (15)$$

where $P(d_1)$ and $P(d_2)$ denote the relative frequency of occurrence of the horizontal and vertical components of motion vector $\boldsymbol{d}$. The PSNR and entropy of

EFE and OFE are listed in Table 1 (after 20 iterations).

Table 1. Comparison of OEF and EFE methods

Method	PSNR(dB)	Entropy(bits)
OFE	18.691	3.207
EFE	19.178	3.168

6 Discussions and conclusions

Based on the gas dynamic theories and the optical flow analytic method, a new concept, the image temperature, is introduced to estimate the image motion field. The motion field estimated by EFE is clearly improved, compared to that estimated by the traditional OFE. The experiment testified the rationality and availability of the EFE.

Computation of the optical flow field is a hot research area. In order to improve computing results, many scholars have imposed solutions. Nagel[5] has imposed the directional-smoothness method in 1986, which improves significantly the performance of the motion estimates. Tretiak[6] believed that the optical field computing was a differential problem and imposed an accessorial constraint based on the second-order differential operator. Terzopoulus[7] has presented a better directional-smoothness method. Those methods improved significantly the performance of the motion estimates by OFE. We believe that those methods maybe also give us a better performance of the motion estimates by EFE, and it is our future work.

EFE represents the essential property of a pixel movement in an image sequence. Therefore, EFE can be applied to object tracking, shot detection, scene analysis and video indexing, etc.

Acknowledgments The authors would like to thank Prof. Carl Minzner and Mr. Wu Chenggui for their encouragement and help.

References

1 Cafforio, C. et al. Methods for measuring small displacement of television images. IEEE Trans. Info. Theory, 1976, 573.
2 Jones, E. R. Contemporary College Physics. 3rd ed. Boston: Mc-Graw Hill, 2001.
3 Swain, M. J. et al. Color indexing. Int. Journal of Computer Vision, 1991, 7: 11.
4 Horn, B. K. P. et al. Determining optical flow. Artificial Intelligence, 1981, 17: 185.
5 Nagel, H. H. Constraints for the estimation of displacement vector fields from image sequences. In: Proceedings of IJCAI'-83, Karlsruhe, Germany, 1983, 945~951.
6 Tretiak, O. et al. Velocity estimation from image sequences with second order differential operators. In: Proceedings of International Conference on Pattern Recognition. Montreal, Canada, 1984, 16~19.
7 Terzopoulos, D. Regularization of inverse visual problem involving discontinuities. IEEE Trans. Pattern Analysis and Machine Intelligence, 1986, 8(4): 413.