GLOBAL MOTION ESTIMATION IN FREQUENCY AND SPATIAL DOMAIN

Sanjeev Kumar, Mainak Biswas and Truong Q Nguyen

Department of Electrical and Computer Engineering, UCSD
La Jolla, California 92093–0407
Email: {sakumar@, mbiswas@, nguyent@ece.}ucsd.edu

ABSTRACT

We propose a fast and robust global motion estimation algorithm based on two-stage coarse-to-fine refinement strategy, which is capable of measuring large motions. Six-parameter affine motion model has been used. Coarse estimation is done in frequency domain using polar, log-polar or log-log sampling of Fourier magnitude spectrum of sub-sampled image. Fourier magnitude spectrum, as translation invariant domain, allows for determination of 4 parameters independent from translation. Sampling scheme is adaptively selected based on past motion pattern. Adaptive selection of sampling scheme insures best trade-off between accuracy and maximum range of motion measurements for current motion pattern. Refinement stage consists of RANSAC based model fitting to motion vectors of randomly selected high-activity blocks, and hence is robust to outliers. Motion vector of blocks is measured using phase correlation, which offers two advantages in this context: sub-pixel accuracy without significant computational overhead, and if a particular block consists of background as well as foreground pixels, both motions are simultaneously measured; as opposed to other methods like block matching which rely on SAD or SSD error metrics and hence fail in such situations. Due to its hardware-friedly nature proposed algorithm holds potential for real-time GME even for television images.

1. INTRODUCTION

In general, the term Global motion is used to describe coherent component of motions of different constituent parts of an object, by a parameterized motion model. The process of estimating these parameters is known as Global Motion Estimation (GME). Commonly, Global motion refers to apparent motion of background, induced by that of camera. Global motion estimation is used in several applications viz. video compression, segmentation, mosaicing, image registration, camera stabilization etc. Possibility of large motions, differently moving foreground objects, and appearing and disappearing image regions make the problem of Global motion estimation very difficult especially under the constraints of limited computational resources.

1.1. Affine Motion Model

The parametric models used include 2-parameter translation model, 4-parameter RST model, 6-parameter affine model, 8-parameter projective model etc. Among these methods affine motion model is very popular as it provides good tradeoff between generality and ease of estimation.It is mathematically expressed as,

$$\tilde{X} = AX + B \tag{1}$$

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} X = \begin{bmatrix} x \\ y \end{bmatrix} \tilde{X} = \begin{bmatrix} \tilde{x} \\ \tilde{y} \end{bmatrix} B = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

X and $\tilde{X}$ are coordinates of corresponding pixels in reference and current image. A is linear part and B is translational part of the affine motion parameters.
Global motion estimation methods can be broadly classified into two categories: spatial domain (Image or pixel domain) and frequency domain (Fourier and wavelet).

1.2. GME in spatial domain

Most common spatial domain methods include those based on minimization of SSD (sum of squared difference) or SAD (sum of absolute difference) error metric. SSD minimization is accomplished by gradient descent methods like ML (Marquardt-Levenburg) [1]. Since, this technique is iterative and each iteration involves image warping and computation of derivatives, it is computationally very intensive and slow. Several speed-up strategies have been suggested in literature, e.g. use of multiresolution framework [1], selective integration and warp free formulation [2]. SAD error metric minimization is accomplished by direct search of parameter space. But, complexity of search increases exponentially with number of parameters. These techniques suffer from the disadvantage that they might get stuck in local minima, although it is less likely in multi resolution framework.

Feature based methods rely on extraction and tracking of feature points [3]. But, extracting reliable features in presence of occlusion junctions, and handling of appearing and disappearing features are very difficult. A closely related class of methods uses block motion vectors instead of coordinates of feature points, which is very suitable for MPEG-2 to MPEG-4 transcoding since motion vectors are already computed [4]. But, range and accuracy of these methods are limited by range and accuracy of block motion vectors. If block motion vector are not available, computational cost of finding them for a reasonable range of motion vectors and sub-pixel accuracy can be prohibitive.

1.3. GME in frequency domain

Phase correlation is very popular and efficient technique for motion estimation [5]. Resampling the Fourier magnitude spectra on log-polar grid has been introduced to also estimate scaling and rotation using phase-correlation and it has been used in image registration [6] and global motion estimation [7].

Estimation of affine parameters in frequency domain is based on Affine Theorem of Fourier Transform [8]. Fourier shift property is exploited to achieve translation invariance by taking magnitude of Fourier spectra of images. By working in this translation invariant (Fourier-Mellin) domain, linear component A of affine transformation can be determined independent of translational component B. Once linear component has been determined, it can be compensated for and translation can be determined using Phase-correlation. For determination of linear component, [9] proposed nonlinear optimization formulation.

2. MOTIVATION

The basic idea behind proposed algorithm is to model affine transformation as a deviation from a family of simpler transformation like similarity (RST) which are subsets of affine transformation and hence, invariably have less than 6 parameters. We call such transformations as low-order or coarse approximation of affine transformation. The method consists of finding the coarse approximation in frequency domain from downsampled image, then refining (upgrading to 6 parameter affine model) this estimate. In order to formalize this idea we reformulate A in (1) as

$$A = \begin{bmatrix} s_x cos\left(\theta_x\right) & -s_x sin\left(\theta_x\right) \\ s_y sin\left(\theta_y\right) & s_y cos\left(\theta_y\right) \end{bmatrix} \tag{2}$$

where, s_x and s_y can be interpreted as non-uniform scaling and θ_x and θ_y can be interpreted as non-uniform rotation. Let us define,

$$\begin{array}{lccl} \text{mean scaling} & s_m & = & \sqrt{s_x s_y} \\ \text{differential scaling} & ds & = & \sqrt{\frac{s_x}{s_y}} \\ \text{mean rotation} & \theta_m & = & \frac{1}{2}\left(\theta_x + \theta_y\right) \\ \text{differential rotation} & d\theta & = & \frac{1}{2}\left(\theta_x - \theta_y\right) \end{array}$$

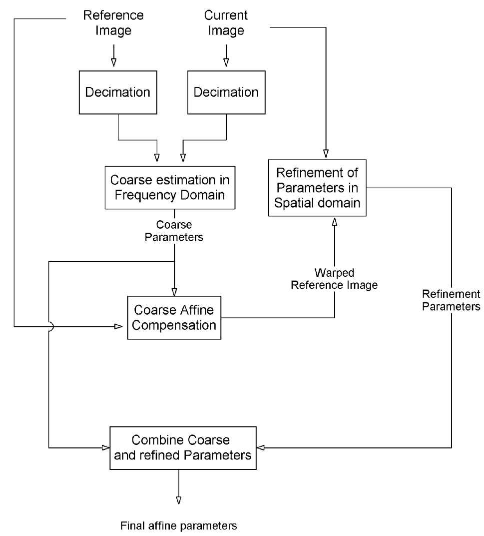

Fig. 1. Top Level Block Diagram

Using these definitions, (2) takes following form.

$$A = \begin{bmatrix} s_m.ds.cos\left(\theta_m + d\theta\right) & s_m.ds.sin\left(\theta_m + d\theta\right) \\ -\frac{s_m}{ds}.sin\left(\theta_m - d\theta\right) & \frac{s_m}{ds}.cos\left(\theta_m - d\theta\right) \end{bmatrix} \tag{3}$$

When $ds = 1$ and $d\theta = 0$, (3) reduces to a similarity transformation, for which log-polar mapping has been widely used in literature. Similarly, when $\theta_x = 0$ and $\theta_y = 0$, (2) reduces to non-uniform scaling transformation, which is recovered using log-log mapping. But when considered separately, these methods are not very attractive since their range of applicability is very limited. Hence our approach is to exploit the benefit of each by means of algorithm described in the next section.

3. PROPOSED ALGORITHM

Fig.1 shows steps of our proposed algorithm. These steps are subsequently described.

3.1. Decimation

Since we are using fourier magnitude spectrum as translation invariant domain, FFT of whole image is needed, which is prohibitively costly. This problem is alleviated by decimating reference and current images (in our implementation, by $\frac{1}{4}$ along each dimension). For better efficiency, integer low-pass filter was used with very few non-zero bits in the coefficients.

3.2. Coarse Estimation in Frequency Domain

Decimated images undergo FFT and magnitude extraction. Since dynamic range of the output of FFT is very high,

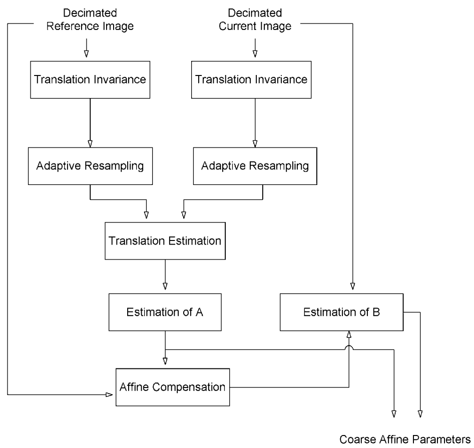

Fig. 2. Coarse Estimation

making interpolation in the frequency domain difficult; this range is compressed by taking $\log$ of magnitude of coefficients.

Fourier magnitude spectrum is resampled on an adaptively selected sampling-grid (Fig.2). For small mean scaling and mean rotation, no remapping is applied to save computations. For small mean scaling, but possibly large rotation, simple polar mapping is used. For small mean rotation, but large and possibly inhomogeneous scaling, log-log mapping is used. For large rotation and scaling, log-polar mapping is used. All the decisions are based on past motion parameters. If past motion parameters are not available e.g. due to scene change, log-polar mapping is used as default. Translation between the resampled images is measured using Phase-correlation method and an estimate of A is inferred from it. For remapping and inferring A, following relations are used,

Log-Log mapping:

$$\begin{array}{l} u = sign\left(x\right)\log_{\alpha}\left(abs\left(x\right)\right) \\ v = sign\left(y\right)\log_{\alpha}\left(abs\left(y\right)\right) \end{array} \qquad A = \begin{bmatrix} \alpha^{-du} & 0 \\ 0 & \alpha^{-dv} \end{bmatrix}$$

Log-Polar mapping:

$$\begin{array}{l} r = \log_{\alpha}\left(\sqrt{x^2+y^2}\right) \\ \theta = \arctan\left(\frac{y}{x}\right) \end{array} \qquad A = \alpha^{-dr}\begin{bmatrix} \cos d\theta & \sin d\theta \\ -\sin d\theta & \cos d\theta \end{bmatrix}$$

Polar mapping:

$$\begin{array}{l} r = \sqrt{x^2+y^2} \\ \theta = \arctan\left(\frac{y}{x}\right) \end{array} \qquad A = \begin{bmatrix} \cos d\theta & \sin d\theta \\ -\sin d\theta & \cos d\theta \end{bmatrix}$$

where, α is a suitably chosen constant, depending on size of image and desired resolution of range space. Exploiting the hermitian-symmetry of Fourier transform of real signals θ is restricted to the range $[0, \pi)$. This estimate of A is used to warp decimated version of reference image in order to compensate for linear component of affine transformation. Again translation estimation is performed, but this time between current image and warped version of reference image(spatial domain), to get an estimate of translational component $B = \begin{bmatrix} dx & dy \end{bmatrix}^T$ of affine parameters .Here, dx and dy are shifts measured between current frame and warped version of reference frame.

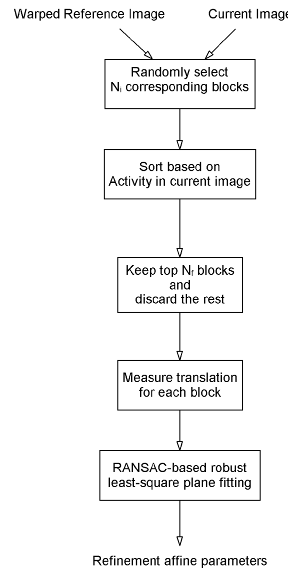

Fig. 3. Refined Estimation

3.3. Refinement in Spatial Domain

Proposed refinement approach is based on following rearrangement of (1) resulting in planar profiles for x and y components of motion vectors.

$$dX = \tilde{X} - X = (A - I_2)\, X + B \tag{4}$$

After the reference image has been warped in order to compensate for coarse affine parameters, following steps are applied to find refinement parameters (Fig.3). N_i random blocks are selected from current image, which also have a corresponding block in the warped reference frame. This strategy alleviates, up to some extent, the problem of appearing and disappearing image regions.

N_i randomly selected blocks are sorted on the basis of activity and only top N_f blocks are considered. The aim of this step is to distinguish between promising and non-promising blocks for further motion estimation, since low activity blocks are likely to give wrong motion vectors due to apperture effect.

Motion vector for each of N_f blocks is computed using phase-correlation as it provides two advantages as compared to other methods, in this regard: subpixel-accuracy

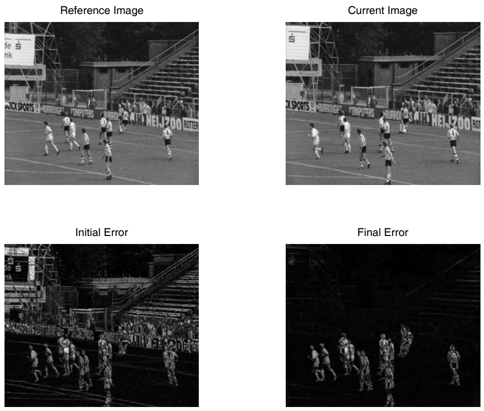

Fig. 4. Simulation Results

and robustness to minority outlier (foreground) pixels. Coordinates of center of blocks (x,y) and motion-vector of blocks (dx,dy) are passed to RANSAC-based robust least-square plane fitting module [10].

Refinement parameters obtained by RANSAC are combined with coarse parameters by multiplying corresponding homogeneous affine matrices. Homogeneous affine matrix is given by $\begin{pmatrix} A & B \\ 0_{1\times 2} & 1 \end{pmatrix}$.

4. SIMULATION RESULTS

In order to test the algorithm for large motion, 1^{st} and 10^{th} frames of the soccer sequence were used as reference and current frames. The robustness of algorithm in presence of differently moving foreground objects is tested by not using any a-priori foreground-background segmentation information. Final error image in Fig.4 shows that motion of background has been well estimated and compensated for. Algorithm was also used for global motion compensated Format Conversion and Slow Motion Playback. Results can be downloaded from `http://videoprocessing.ucsd.edu/global_demo.htm`.

5. CONCLUSION AND FUTURE WORK

In this paper we presented a robust and fast global motion estimation. The algorithm exhibits synergetic combination of frequency domain methods, which can measure very large motions but can not be extented in natural way to higher order motion models, and spatial domain methods which can handle problems due to occluding and other non-reliable regions and can be readily used for affine motion model.

6. REFERENCES

[1] R. Szeliski and J. Coughlan, "Hierarchical spline-based image registration," in *Proceedings CVPR-1994 (IEEE Computer Society Conference on Computer Vision and Pattern Recognition)*, June 1994, vol. 1, pp. 194–201.

[2] Y. Keller and A. Averbuch, "Fast gradient methods based on global motion estimation for video compression," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 4, pp. 300–309, Apr. 2003.

[3] H. Richter, A. Smolic, B. Stabernack, and E. Muller, "Real time global motion estimation for an mpeg-4 video encoder," in *Proceedings PCS-2001 (Picture Coding Symposium)*, Apr. 2001, vol. 1, pp. 401–404.

[4] Yeping Su, Ming Ting Sun, and V Hsu, "Global motion estimation from coarsely sampled motion vector field and the applications," in *Proceedings ISCAS-2003 (IEEE International Symposium on Circuits and Systems)*, May 2003, vol. 2, pp. 628–631.

[5] C. Kuglin and D. Hines, "The phase correlation image allignment method," in *Proceedings ICCS-1975 (IEEE International Conference on Cybernetics and Society)*, 1975, pp. 163–165.

[6] B. S. Reddy and B. N. Chatterji, "An fft-based technique for translation, rotation, and scale-invariant image registration," *IEEE Transactions on Image Processing*, vol. 5, no. 8, pp. 1266–1271, Aug. 1996.

[7] L. Hill and T. Vlachos, "On the estimation of global motion using phase correlation for broadcast applications," in *Proceedings IPA-1999 (IEE International Conference on Image Processing and Its Applications)*, July 1999, vol. 2, pp. 721–725.

[8] R. N. Bracewell, K. Y. Chang, A. K. Jha, and Y. H. Wang, "Affine theorem for two-dimensional fourier transform," *Electronics Letters*, vol. 29, no. 3, pp. 304, Feb. 1993.

[9] L. Lucchese, "Estimating affine transformations in the frequency domain," in *Proceedings ICIP-2001 (IEEE International Conference on Image Processing)*, Oct. 2001, vol. 2, pp. 7–10.

[10] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.