AM-FM Demodulation Methods for Reconstruction, Analysis and Motion Estimation in Video signals

Víctor Murray and Marios S. Pattichis
The University of New Mexico, Department of Electrical and Computer Engineering
Albuquerque, N.M. 87131, U.S.A.
Emails: vmurray@ieee.org, pattichis@ece.unm.edu

Abstract

In this paper, we present new 3D Amplitude-Modulation Frequency Modulation (AM-FM) methods for video image analysis. We develop a framework that allows us to reconstruct the input video using AM-FM representations and also provide motion estimates that correspond to these AM-FM representations. The proposed motion estimation method provides three motion estimation equations per channel filter (AM, FM motion equations and a continuity equation). The methods are implemented using efficient, separable 1D filterbanks. We demonstrate the method in motion tracking, trajectory estimation and video signal reconstruction.

1 Introduction

The AM-FM representation allows us to model non-stationary 3D signal content in terms of amplitude and phase functions. The method has been recently extended to cover video applications in [1] using:

$$I(x,\, y,\, t) = \sum_{n=1}^{n=M} a_n(x,\, y,\, t) \cos \varphi_n(x,\, y,\, t). \tag{1}$$

In (1), a continuous signal $I(.)$ is a function of a vector of spatiotemporal coordinates $(x,\, y,\, t)$. A collection of M AM-FM component images $a_n(x,\, y,\, t) \cos \varphi_n(x,\, y,\, t)$, $n = 1, 2, \ldots, M$, are used to model essential signal modulation structure. The amplitude functions $a_n(x,\, y,\, t)$ are always assumed to be positive.

For each phase function $\varphi_n(x,\, y,\, t)$ we define the instantaneous frequency (IF) as the gradient of the phase $\nabla\varphi_n(x,\, y,\, t)$. The instantaneous frequency vector $\nabla\varphi_n(x,\, y,\, t)$ can vary continuously over the spatiotemporal domain of the image. The term AM-FM demodulation is used to refer to the computation of the amplitude function $a(x,\, y,\, t)$, the phase function $\varphi(x,\, y,\, t)$, and the instantaneous frequency vector function $\nabla\varphi(x,\, y,\, t)$ from the given input video $I(x,\, y,\, t)$.

The proposed method is an extension of our prior work in [1, 2] and it is also closely related to phase-based methods for motion estimation [3]. In [2], we introduced robust methods for 2D AM-FM demodulation using 2D, separable, multiscale filterbanks. In [1], we extended the multiscale method to 3D.

We first use 3D dominant component analysis (see [4] for the 2D case) to determine the dominant channel for demodulation. At every pixel, for the dominant channel, we then proceed to estimate the 3D instantaneous frequency and amplitude components. We then proceed to formulate AM, FM and continuity constraint equations as discussed in [1]. The important contribution of [1] over the initial work of [3] is that we consider both AM and FM constraint equations. In [3] and other phase-based research, the AM components are not used in the estimation. Yet, in [1], we demonstrated that the AM and the FM constraint equations are equally important.

In this paper, we consider three extensions of our prior work in [1]. First, we consider video reconstructions using 3D AM-FM decompositions. Second, we consider motion estimation over multiple scales (an extension of our 2D work in [2]). Third, we provide motion trajectories over the tracked videos to demonstrate the utility of our approach over general videos.

We describe our methods in section 2. Results are shown in section 3 and finally, the conclusions are given in section 4.

2 Methods

2.1 Video Reconstruction

In this section, we discuss how to reconstruct a 3D signal using its AM-FM components. We consider two different procedures: using AM-FM harmonics (Sec. 2.1.1) or using AM-FM components extracted from different scales (Sec. 2.1.2).

2.1.1 Least-Squares Reconstructions using AM-FM harmonics

We consider reconstructing a 3D signal using AM-FM harmonics (see [2] for a related 2D model):

$$\hat{I}(n_1, n_2, n_3) = \sum_{n=1}^{n=h} c_n a(n_1, n_2, n_3) \cos\left(n\varphi(n_1, n_2, n_3)\right), \tag{2}$$

where h represents the number of AM-FM harmonics, and n_1, n_2 and n_3 represent the discrete versions of x, y and t. In (2), we assume that the instantaneous amplitude $a(n_1, n_2, n_3)$ and the instantaneous phase $\cos\varphi(n_1, n_2, n_3)$ have been estimated using dominant component analysis (DCA).

We then want to compute the AM-FM harmonic coefficients c_n, $n = 1, 2, \ldots, h$, so that $\hat{I}(n_1, n_2, n_3)$ is a least-squares estimate of $I(n_1, n_2, n_3)$ over the space of the AM-FM harmonics. This is accomplished by first computing an orthonormal basis over the space of the AM-FM harmonics using the Modified Gram-Schmidt (MGS) Algorithm (see [5]). We then obtain the least square estimates directly by simply projecting the input video on the orthonormal AM-FM basis functions and then map the coefficients back to the original harmonics.

2.1.2 Multiscale least-squares reconstructions

In this paper, we consider AM-FM methods based on the use of multiple scales (see [2] for the 2D case). In analogy to [2], we only consider one, two and three decomposition levels. We then proceed to estimate dominant dominant AM and FM components over each scale, as well as the low-pass filter.

It is important to recognize that adding decomposition levels also reduces the total amount of video signal energy that is captured by the decomposition. As we shall explain next, this is a direct consequence of our desire to localize the spatiotemporal content at each pixel. First, let us note that for a single scale decomposition, video signal energy is captured by the low-pass filter component and the dominant high-frequency components, selected from the high-frequency channels. Then, in two-scale decompositions, the 3D spectrum captured by the low-pass filter is further decomposed into two new scales. We again find the dominant components in this second scale while the lowest frequency components are captured by the new low-pass filters. Similarly, for three-scales, we decompose the frequency spectrum of the 3D low-pass filter.

The extracted dominant components from each scale allow us to provide decompositions using an independent AM-FM component per scale. Furthermore, the corresponding dominant channel filters allow us to extract local spatiotemporal content over each pixel. This approach allows us to re-formulate the classical motion estimation problem with several independent equations over each scale. It is also important to note that the AM-FM decomposition also allows us to track both continuous and discontinuous motions since at every pixel we can associate three different dominant channels from three different scales.

2.2 Motion Estimation

We now review the method developed in [1, 3], detailing how to estimate motion parameters over each spectral channel. In what follows, we assume an AM-FM component model for the output of each channel:

$$I(x, y, t) = a(x, y, t) \exp(j\varphi(x, y, t)). \tag{3}$$

In [1], we derived the smoothness, AM and FM constraints:

$$E_s = \iint \left[u_x^2 + u_y^2 + v_x^2 + v_y^2\right] dxdy. \tag{4}$$

$$E_{AM} = \iint \left[a_x u + a_y v + a_t\right]^2 dxdy \tag{5}$$

$$E_{FM} = \iint \left[\varphi_x u + \varphi_y v + \varphi_t\right]^2 dxdy. \tag{6}$$

We combine all constraints together to get the final expression

$$E = E_s + \lambda E_{FM} + \beta E_{AM}. \tag{7}$$

To solve (7), we iteratively compute velocity estimates at each iteration p using:

$$u^{(p+1)} = u_{ave}^{(p)} - \lambda \frac{u_{N1}}{D} \varphi_x - \beta \frac{u_{N2}}{D} a_x \tag{8}$$

$$v^{(p+1)} = v_{ave}^{(p)} - \lambda \frac{v_{N1}}{D} \varphi_y - \beta \frac{v_{N2}}{D} a_y, \tag{9}$$

where:

$$u_{ave}(m,n) = \frac{1}{4}[u(m+1,n) + u(m-1,n) + u(m,n+1) + u(m,n-1)],$$

$$v_{ave}(m,n) = \frac{1}{4}[v(m+1,n) + v(m-1,n) + v(m,n+1) + v(m,n-1)],$$

$$u_{N1} = \left(\varphi_x + \beta a_y^2 \varphi_x - \beta a_y a_x \varphi_y\right) u_{ave} + \varphi_y v_{ave} + \left(\beta a_y^2 + 1\right) \varphi_t - \beta a_y \varphi_y a_t,$$

$$u_{N2} = \left[\lambda \left(a_x \varphi_y - a_y \varphi_x\right) \varphi_y + a_x\right] u_{ave} + a_y v_{ave} - \lambda a_y \varphi_y \varphi_t + \left(\lambda \varphi_y^2 + 1\right) a_t,$$

$$D = 1 + \lambda \left(\varphi_y^2 + \varphi_x^2\right) + \lambda\beta \left(\varphi_x a_y - \varphi_y a_x\right)^2 + \beta \left(a_x^2 + a_y^2\right),$$

and similarly for v_{N1}, v_{N2}, after exchanging x with y derivatives. We initialize the estimation by using zero-velocity estimates. Using the estimated motion vectors, we apply a Kalman filter to track the trajectories throughout the video.

We also note that we can generate pure AM and pure FM estimates by manipulating the constraint coefficients (λ, β) in (7). For generating FM estimates, we simply use $\beta = 0$. Similarly, for AM estimates, we set $\lambda = 0$. For AM-FM estimates, we simply use λ, $\beta \neq 0$. We typically set λ, β to about 10 so the solution will more closely track the optical flow constraint equation(s) (see (5) and (6)).

3 Results

In this section, we present video reconstruction and motion tracking results from the *taxi video* sequence (see Hamburg taxi video in [6]) and for a *two people fighting video* (see [7]).

We summarize video reconstruction results in Tables 1, 2, 3 and 4, and Fig. 1. We provide the mean squared error (MSE) in Tables 1 and 3. Here, we note that the zeroth-scale refers to least-squares reconstruction using the low-pass filter output from each scale. This low-pass filter varies with scale as we detailed in the methods section. It is interesting to note the relative importance of the different scales as summarized in the optimal coefficients in Tables 2 and 4. In Table 2, we note that in three-scale reconstructions, the second scale AM-FM component coefficient is equal to the low-pass filter component. As we shall see, the motion estimates from this level will prove very useful. On the other hand, the use of AM-FM harmonics did not reduce the MSE by any significant amount. In Table 4, for three-scale reconstruction, we see significant values for AM-FM coefficients for all scales.

We show two different targets to demonstrate the independence among the AM, AM-FM and the FM estimates.

Table 1. MSE in the taxi video.

Multiscale least-squares reconstructions			
levels	single-scale	two-scale	three-scale
0	34.7781	115.1141	186.2611
0, 1	34.5966	110.9505	185.2432
0, 1, 2	-	110.7431	180.1949
0, 1, 2, 3	-	-	179.9799
Reconstructions using AM-FM harmonics			
harmonics	single-scale	two-scale	three-scale
1	34.7336	115.3749	202.3249
2	34.5546	115.3658	201.5240
3	34.2859	115.3632	197.0237
4	34.2842	115.2712	194.7118
5	34.2179	115.1972	194.6775

Table 2. Coefficients used in the taxi video.

Coefficients for AM-FM reconstructions from multiple scales.			
level	single-scale	two-scale	three-scale
0	0.9967	0.9999	1.0011
1	0.3459	1.0007	0.2205
2	-	0.3698	1.1030
3	-	-	0.3765

We show a person tracking example in Fig. 2 (a). In this example, both the AM and the dominant AM-FM component equations provide good trajectory tracking results. It is interesting to note that in this case, the person image is well-localized and its motion appears to be easier to follow.

It is much more interesting to examine the FM tracking results over the taxi region (see Fig. 2 (b)). Tracking individual pixels over the taxi region proved to be much more challenging. To understand why, we simply note the uniform intensity regions over the surface of the taxi. Yet, the FM motion equation provided nice trajectory tracking results over the edges of the taxi images. We note that these nice tracking results could not have been reproduced with the AM motion equations.

Table 3. MSE in the fight video.

Multiscale least-squares reconstructions			
levels	single-scale	two-scale	three-scale
0	73.6264	156.8473	216.8916
0, 1	73.6263	155.8162	215.9473
0, 1, 2	-	155.7542	214.0385
0, 1, 2, 3	-	-	211.3922
Reconstructions using AM-FM harmonics			
harmonics	single-scale	two-scale	three-scale
1	73.7216	157.2826	225.0483
2	73.7179	154.9638	222.8834
3	73.4665	154.9636	221.4078
4	73.4570	154.5769	220.1254
5	73.4135	154.5308	220.1254

Table 4. Coefficients used in the fight video.

Coefficients for AM-FM reconstructions from multiple scales.			
level	single-scale	two-scale	three-scale
0	0.9972	0.9987	0.9991
1	0.0037	0.7138	0.7012
2	-	0.1070	0.8942
3	-	-	0.6987

4 Conclusions

In this paper, we have developed the theory and application of new methods for video image analysis using AM-FM representations. We have shown video reconstructions using AM-FM harmonics from multiple-scales. Furthermore, we showed how AM-FM demodulation over a single channel filter can contribute three different equations (AM, FM, and continuity equations). Thus, compared to existing standard methods of using two motion equations per pixel, the proposed AM-FM methods can be used to generate three equations per channel. In addition, the AM-FM approach provides reconstruction methods for the input video.

References

[1] V. Murray, S. Murillo, M. Pattichis, C. Loizou, C. Pattichis, E. Kyriacou, and A. Nicolaides, "An AM-FM model for motion estimation in atherosclerotic plaque videos," in *Proc. Fortieth Asilomar Conference on Signals, Systems and Computers*, 2007.

[2] V. Murray, P. Rodriguez V., and M. S. Pattichis, "Robust multiscale AM-FM demodulation of digital images," in *IEEE International Conference on Image Processing*, vol. 1, 2007, pp. 465–468.

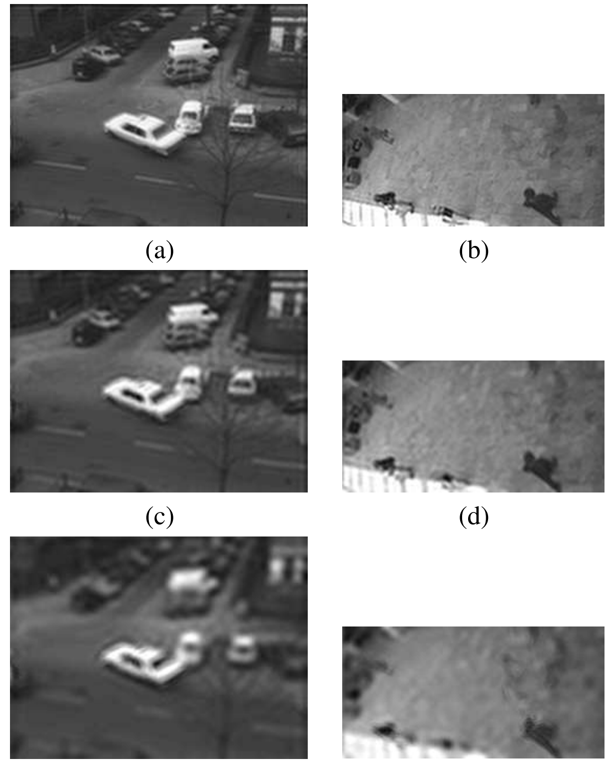
[3] D. J. Fleet and A. D. Jepson, "Computation of component image velocity from local phase information," *International Journal of Computer Vision*, vol. 5, no. 1, pp. 77–104, 1990.

[4] J. P. Havlicek, "AM-FM image models," Ph.D. dissertation, The University of Texas at Austin, 1996.

[5] J. W. Demmel, *Applied Numerical Linear Algebra*. SIAM, 1997, ch. Linear Least Square Problems, pp. 105–117.

[6] J. Barron, D. Fleet, and S. Beauchemin, "Performance of optical flow techniques," *International Journal of Computer Vision*, vol. 12, no. 1, pp. 43–77, 1994.

[7] [Online]. Available: http://www-prima.inrialpes.fr/PETS04/caviar_data.html



(a) (b) (c) (d) (e) (f)

Figure 1. Video reconstructions for the videos (we only show frame 1) using dominant components from each scale. (a)-(b) Single-scale. (c)-(d) Two-scale. (e)-(f) Three-scale. (a), (c) and (e) Taxi video. (b), (d) and (f) Fight video.

(a) (b)

Figure 2. Frame 1 of the taxi video for tracking with the motion estimated during the video. (a) Zoom in the person. (b) Zoom in the taxi.