International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970)
Volume-2 Number-4 Issue-6 December-2012

Video Fusion Using Pixel Averaging, Principal Component Analysis and Laplacian Pyramid – A Comparative Study

P. Surya Prasad[1], P.Poornesh[2], P N R L Chandra Sekhar[3]
Department of ECE M.V.G.R. College of Engineering, Visakhapatnam, India[1,2]
Department of CSE, GITAM University[3]

Abstract

In this paper, performances of various video fusion algorithms are compared by applying them to a set of infrared (IR) and visible band videos. The application of interest is area surveillance and the fusion process aims at integration of complementary information from multi-sensor inputs for enhancing the human perception of the monitored scene and to make the result suitable for further processing. The performance of algorithms viz. pixel averaging, principal component analysis and Laplacian pyramid are compared. A set of measures of effectiveness for comparative performance analysis like Fusion Factor and Fusion Symmetry are defined and applied on the output of the above fusion algorithms.

Keywords

Pixel averaging, principal component analysis, Laplacian pyramid, Fusion Factor, Fusion symmetry.

1. Introduction

Multi-sensor data fusion is a process of combining visual data from different sensors, of different wavelengths to form a single composite video preserving the information of the sources [1]. The fusion of information from sensors with different physical characteristics enhances the understanding of our surroundings and provides the basis for planning, decision-making, and control of autonomous and intelligent machines [2]. The composite image is formed to improve image content and to make it easier for the user to detect, recognize, and identify targets and increase situational awareness. Computational vision systems that provide visual guidance are used in tasks such as detection and recognition needs to be robust with respect to unpredictable environmental conditions. In the past decades it has been applied to different fields such as pattern recognition, visual enhancement, object detection and area surveillance [3]. The objective of video fusion is to reduce uncertainty and minimize redundancy in the output while maximizing relevant information particular to an application or task.

This multi-sensor based video fusion system is a challenging task and fundamental to several modern day image processing applications. It finds many such applications in the fields of security systems, defence applications, intelligent machines, remote sensing, medical imaging and machine vision. For land-use classification, for example, the thematic mapper images of LAND-SAT and SAR images can be fused to obtain a better picture of the area under consideration. In military applications, image fusion is generally applied for object or target recognition. Data can be provided by radar, optical, infrared and other sensors. The requirements for a Video fusion system are a) the fusion must not introduce artefacts that can distract or mislead the human observer; b) the merging operation should be reliable and robust against disturbances and errors [4].

Surprisingly, the idea to couple visible and thermal infrared is not yet seen as a popular research field due to the still high cost of the thermal infrared cameras versus their visible counter parts. Moreover outdoor scenarios are obviously more challenging to visible imagery due to shadows, light reflections, levels of darkness and luminosity. However, on the other hand, moving leaves and grass, cooling winds, moving shadows with clouds, reflecting snow, etc., are challenging for IR imagery too.

Video fusion has the advantage of reducing the computational load and mitigating the rapid brightness variations in the fused video. It is also less sensitive to the presence of noise. The aim of video fusion, apart from reducing the amount of data, is to create new videos that are more suitable for the purposes of human/machine perception, and for further image-processing tasks such as segmentation, object detection or target recognition in applications such as remote sensing and medical imaging. Video fusion is often a vital pre-processing procedure to many computer vision and image processing tasks which are dependent on the acquisition of imaging data via sensors, such as IR and visible. One such task is that of human detection. To detect humans with an artificial system is difficult for a number of reasons. The main challenge for a vision-based pedestrian detector is the high degree of variability with the human appearance due to articulated motion, body size, partial occlusion, inconsistent cloth texture, highly cluttered backgrounds and changing lighting conditions. Moreover, the applications, to protect pedestrians, define hard real-time requirements and rigid performance criteria.

Fusion symmetry measure quantifies the relative distance in terms of mutual information of the fused image with respect to input images. The smaller the FS the more symmetric is the fused image i.e. it captures information from both the input images. Also the traditional criterion of maximizing the joint mutual information is also quantified and a definition called fusion factor is evolved.

2. Fusion Techniques

The most important issue concerning fusion of videos is to determine how to combine the sensor outputs. Pixel based fusion schemes range from simple averaging of the pixel values of registered images to more complex multi-resolution (MR) pyramid [5-7]. Spatial image fusion methods work by combining the pixel values of the two or more images to be fused in a linear or non-linear way. The simplest form is a weighted averaging of the registered input to give the fused video.

Multiscale decomposition based methods combine the multiscale decomposition of the source images. The idea is to perform a multiscale transform on the source images, construct a composite representation of these using some sort of fusion rule, and then construct the fused image by applying the inverse multiscale transform[8-9] The commonly used multiscale decomposition fusion method is pyramid transforms. A pyramid transform fusion consists of a number of images at different scales which together represent the original image. An example for a pyramid transform is the Laplacian Pyramid. Each level of the Laplacian Pyramid is constructed from its lower level using blurring, size reduction, interpolation and differencing in this order.
PCA (Principal Component Analysis) is a general statistical technique that transforms multivariate data with correlated variables into one with uncorrelated variables. These new variables are obtained as linear combination of the original variables.

The majority of applications of a fusion scheme are interested in features within the image, not in the actual pixels. Therefore, it seems reasonable to incorporate feature information into the fusion process. In this paper, pixel averaging method, principal component analysis and Laplacian pyramid methods are applied and compared.

3. Implementation

A. Pixel Averaging

Simple Average mechanism is a simple way of obtaining an output image with all regions in focus. The value of the pixel P (i, j) of each image is taken and added. This sum is then divided by N to obtain the average. The average value is assigned to the corresponding pixel of the output image. This is repeated for all pixel values. The Greatest Pixel Value algorithm chooses the in focus regions from each input image by choosing the greatest value for each pixel, resulting in highly focused output. The value of the pixel P (i, j) of each image is taken and compared to each other. The greatest pixel value is assigned to the corresponding pixel of the output image. This is repeated for all pixel values.

B. Principal Component Analysis

PCA fusion is a technique, which enhances the resolution of several bands of data [15]. The principal component analysis is statistical technique that transforms a multi variety inter-correlated data set into a new un-correlated data set.

It is a way of identifying patterns in data, and express the data in such a way that it highlights the similarities and differences. Since patterns in data are hard to find in high dimension data, in which luxury of graphical representation is not available, PCA is a powerful tool for analysing such data. The other main advantage of PCA is that once we find these patterns in the data, and the data can be compressed. The procedure of PCA is as per the following stepwise procedure.

Step 1: Get some data. Take two 2D images.
Step 2: Subtract the mean.
For PCA to work properly, the mean has to be subtracted from each of the data dimensions. The mean subtracted is the average across each

dimension. So, all the x values have $\bar{x}$ (the mean of the x values of all the data points) subtracted, and all the y values have $\bar{y}$ subtracted from them. This produces a data set whose mean is zero.

Step 3: Calculate the covariance matrix.

Since the data is 2 dimensional, the covariance matrix will be 2x2. So, since the non-diagonal elements in this covariance matrix are positive, both the x and y variable increase together.

Step 4: Calculate the eigenvectors and Eigen values of the covariance matrix. Since the covariance matrix is square, the eigenvectors and Eigen values on this matrix can be calculated.

Step 5: Choose components and form a feature vector.

Here is where the notion of data compression and reduced dimensionality comes into it. Eigen values are quite different values. In fact, it turns out that the eigenvector with the highest Eigen value is the principal component of the data set. The eigenvector with the largest Eigen value and this relation gives the components in order of significance. To be precise, if the data originally have n dimensions, and if n eigenvectors and Eigen values are calculated, and only the first p eigenvectors are selected, then the final data set has only p dimensions.

Feature vector= (eig1 eig 2 eig 3……………eig n)

Step 6: Derive the new data set

This final step in PCA is the easiest. Once the components (eigenvectors) have been chosen that we wish to keep in our data and formed a feature vector, simply take the transpose of the vector and has to be multiplied on the left of the original data set, transposed. Therefore,

Final data= row feature vector x row data vector.

If a night vision RGB and corresponding IR images are taken as inputs, the Fig. 1 shows the fusion of these two images using PCA method.

Fig. 1: RGB night vision image

Fig. 2: IR night vision image

Fig. 3: PCA fused output

C. Laplacian Pyramid

Multiresolution analysis of images provides useful information for computer vision and image processing applications. The multiresolution formulation is designed to represent signals where a single event is decomposed into finer and finer detail. In the context of image analysis, multiresolution decomposition gives a coarse approximation of the image and three detail images viz., horizontal, vertical and diagonal detail images. Thus the features dominant at various resolutions can be studied, which is not possible if conventional Fourier analysis is used. The multiresolution methods most commonly used for image fusion is the Laplacian Pyramid transform.

Several approaches to Laplacian fusion techniques have been documented since Burt and Anderson introduced this transform back [16] as a technique of image encoding. The Laplacian Pyramid implements a “pattern selective” approach to image fusion, so that the composite image is constructed not a pixel at a time, but a feature at a time. The basic idea of this technique is to perform pyramid decomposition on each of the source images, and then integrate all these decompositions to form a composite representation, and finally reconstruct the fused image by performing an inverse pyramid transform. Image pyramids have been initially described for a multi-resolution image analysis and as a model for the binocular fusion in human vision. An image pyramid can be described as collection of low or band pass copies of an original image in which both the band limit and sample density are reduced in regular steps. There are several pyramid-based transform schemes but only the Laplacian pyramid is described here. The pyramid decomposition of an image is shown in the fig.

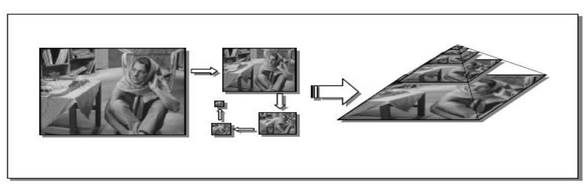


Fig. 4: Pyramid decomposition of Laplacian pyramid

A multi resolution pyramid transformation decomposes an image into multiple resolutions at different scales. A pyramid is a sequence of images in which each level is a filtered and sub sampled copy of its predecessor. The lowest level of the pyramid has the same scale as the original image and contains the highest resolution information. Higher levels of the pyramid are reduced resolution and increased scale versions of the original image.

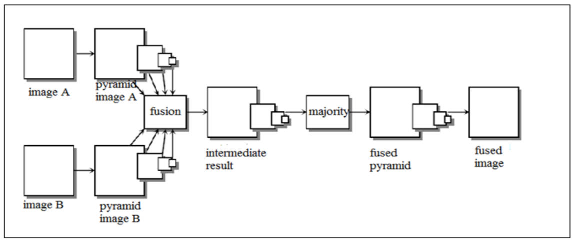


Fig. 5: Schematic diagram of Laplacian pyramid fusion method

If the original image is considered as *g0*, the first step in Laplacian pyramid transform is to low-pass filter the original image *g0* to obtain image *g1*, which is a "reduced" version of *g0*. In similar way *g2* is formed as a reduced version of *g1*, and so on.

The first step is to construct a pyramid for each source image. The fusion is then implemented for each level of the pyramid using a feature selection decision mechanism. It can be used several modes of combination, such as selection or averaging. In the first one, the combination process selects the most salient component pattern from the source and copies it to the composite pyramid, while discarding the less salient pattern. In the second one, the process averages the sources patterns. This averaging reduces noise and provides stability where source images contain the same patter information. The former is used in locations where the source images are distinctly different, and the latter is used in locations where the source images are similar. One other possible approach, chosen in this research, is to select the most Salient component, following next equation

$$F_1(x,y) = \begin{cases} A_1(x,y), if A_1(x,y) > B_1(x,y) \\ B_1(x,y), otherwise \end{cases}$$

Where A_l, B_l and F_l are the two input and fused signals for levels $0 \le l \le N - 1$. Then a consistency filter is applied. The aim of this consistency filter is to eliminate the isolated points. Finally, for level N it is performed an average of both source components.

$$F_n(x,y) = \frac{A_n(x,y) + B_n(x,y)}{2}$$

This function method uses a recursive algorithm to achieve three main tasks. First, it constructs the Laplacian pyramid of the source images. Second, it does the fusion at each level of the decomposition. And finally, it reconstructs the fused image from the fused pyramid.

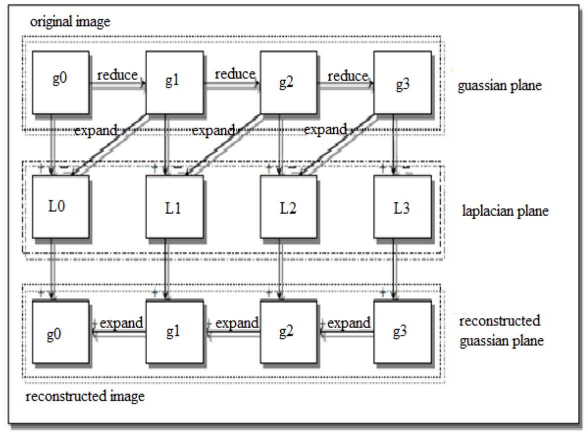


Fig. 6: Block diagram of Laplacian fusion

The level-to-level averaging process is performed for levels $0 < l < N$ and nodes i, j, $0 <= i < C1$, $0 <= j < R$ following the equation:

$$g_1(i,j) = \sum_{\substack{-2 \le m \le 2 \\ -2 < n < 2}} w(m,n) g(2i + m, j + n)$$

The input arguments of this function are the source images (im1, im2): must have the same size, and are suppose to be already registered. Number of scales (ns): an integer that defines the number of pyramid decomposition levels.

The consistency checking is applied if its value is '1'. As shown in the flow chart of Fig. 7, there are five main blocks in the algorithm, Block A: images size checking, Block B: construction of pyramid level n, Block C: pyramid level fusion, Block D: final level analysis and Block E: reconstruction of fused image.

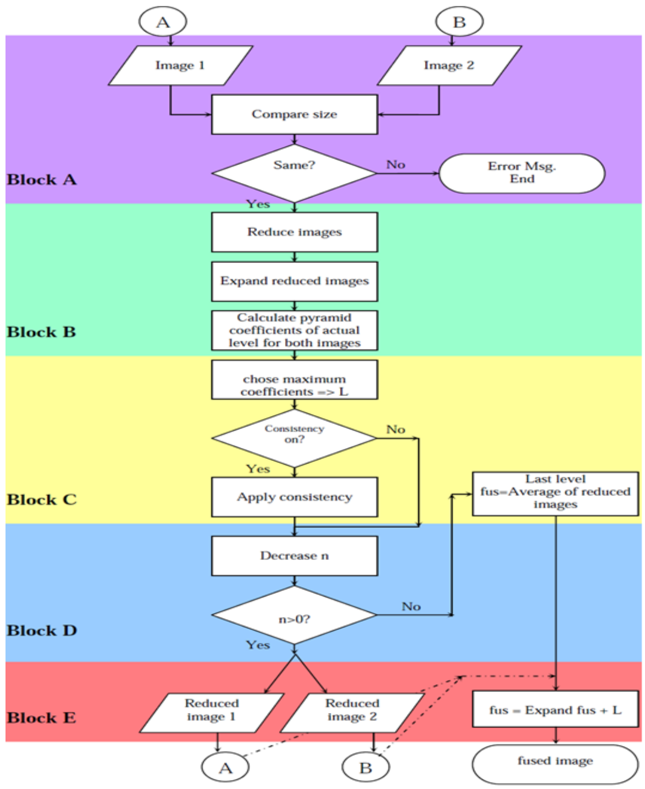


Fig. 7: Flow chart of Laplacian pyramid decomposition

To implement Laplacian pyramid decomposition, two elementary scaling operations are to be defined first, usually referred to as reduce and expand. The reduce operation applies a low-pass filter to the image and down samples it by a factor of two. The expand operation employs a predefined interpolation method and upsamples the image by a factor of two. Given these two operations, the Laplacian pyramid is obtained through the following two-step process:

1. Generate a Laplacian pyramid Li for each of the images I_i.
2. Merge the pyramids $\{L_i\}$ by taking the maximum at each pixel of the pyramid, obtaining the Laplacian pyramid representation L of the fusion result.
3. Reconstruct the fusion result I from its Laplacian pyramid representation.
4. Normalize the dynamic range of the result so that it resides within the range of [0,1], and apply additional post-processing techniques as necessary.

A typical set of night vision RGB and equivalent IR images are taken as inputs. The Figures 8, 9 and 10 shows the fusion of these two images using Laplacian method.

Fig. 8: RGB night vision image

Fig.9: IR night vision image

Fig.10: Laplacian Pyramid fusion image

4. Performance Criteria

Due to the availability of multiple image sensors in many fields such as remote sensing, medical imaging, military applications and area surveillance, sensor fusion has emerged as an interesting area of research. Standard deviation of the difference between the ideal image and the fused image is taken as the performance measure of the fusion scheme in reference. However, in a practical situation, an ideal image is not available. A mutual information criterion is used as the measure for evaluating the performance in references [1-8].

To select the best fusion method two new definitions namely- Fusion Factor and Fusion Symmetry provide useful guidelines in selecting the best Fusion algorithm among the given algorithms.

A. Fusion Factor and Fusion Symmetry

For comparing the different methods, we make use of Mutual information Measure (MIM). Mutual information gives the amount of correlation between two distributions. Given two image $M(i, j)$ and $N(i, j)$, the MIM is defined as,

$$I_{mn} = \sum_{x,y} P_{mn}(x,y)\log(P_{mn}(x,y)/P_m(x)P_n(y))$$

Where, P_m (x) and P_n (y) are the probability density functions in the individual images and P_{mn} (x,y) is the joint probability density function. Estimations for the joint and marginal density functions can be obtained by simple normalizing of the joint and marginal histograms of both the images. The following definitions are the guidelines that have to be followed for selecting the best fusion algorithm.

1) Fusion Factor (FF)

Given two images A and B, and their fused image F, the fusion factor is given by

$$FF = I_{AF} + I_{BF} .$$

A higher value of FF indicates that the fused image contains fairly good amount of information present in both the images. However, a high value of FF doesn't imply that the information from both the images is symmetrically fused.

2) Fusion Symmetry (FS)

Fusion symmetry is an indication of how much symmetric, the fused image is, with respect to input images.

$$FS = abs((I_{AF} / FF) - 0.5)$$

The lower the value of *FS* the better the Fusion algorithm.

I_{AF}: mutual information of image A and fused image F

I_{BF}: mutual information of image B and fused image F

5. Design Cycle

Multi-sensor image fusion can be performed at four different processing levels, according to the stage at which the fusion takes place: signal level, pixel level, feature level, and decision level. Fig. 11 illustrates of the concept of the four different fusion levels.

Table 1: Performance criteria of fused methods

FUSION METHOD	I_{AF}	I_{BF}	FUSION FACTOR	FUSION SYMMETRY
PIXEL AVERAGING	*1.2888*	*0.8847*	*2.1735*	*0.5929*
PCA	*1.4773*	*1.2039*	*2.6812*	*0.5509*
LAPLACIAN PYRAMID	*0.9359*	*2.0913*	*3.0272*	*0.3091*

6. Results and Discussion

Performance of various video fusion algorithms is evaluated and compared by applying the images corresponding to multi-sensor inputs to different video quality subjective metrics. The input images used in all algorithms were pre-registered images, of equal size, taken from corresponding frames of both visual and infrared videos.

I_{AF} is the mutual information of image A and fused image F. I_{BF} is mutual information of image B and fused image F. The higher value of Fusion Factor and lower value of Fusion Symmetry obtained for Laplacian Pyramid method indicates the high performance of this algorithm. The following table shows performance criteria of the above fusion algorithms.

From the above table, it can be clearly seen that higher value of fusion factor obtained for Laplacian fusion method indicates that the fused image contains fairly good amount of information present in both the images. Also the lower value of Fusion symmetry is an indication of how much symmetric, the fused image is, with respect to input images.
The experimental results are shown in the figures (a) to (k) obtained by applying the various video fusion algorithms on the data taken in a typical area surveillance scenario. We have taken 500 frames of the RGB and IR videos of the corresponding scene.
The figures (a) and (b) are 10^{th} frames of corresponding RGB and IR videos.


Fig. 11: (a) (b)

The below figures (c) and (d) are the fused frames of the (a) and (b) by using pixel averaging and PCA fusion algorithms respectively.

(c)

(d)

The below figure (e) is the fused frames of the (a) and (b) by using Laplacian fusion

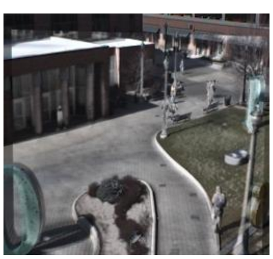
(e)

The below figures (f) and (g) are 203rd frames of corresponding RGB and IR videos.

(f)

(g)

The below figures (h) and (i) are the fused frames of the (f) and (g) by using pixel averaging and PCA fusion algorithms respectively.



(h)

(i)

The below figure (j) is the fused frames of the (g) and (h) by using Laplacian fusion

(j)

References

[1] Anjali Malviya; Bhirud, S. G., Visual Infrared Video Fusion for Night Vision using Background Estimation, Journal of computing, volume 2, issue 4, april 2010, issn 2151-9617.

[2] Y.Chen, C.Han, "Night-time Pedestrian Detection by Visual-Infrared Video Fusion." Proceedings of 7th World congress on Intelligent Control and Automation, June 25-27, 2008, Chongoing, China.

[3] Hall, L.; Llinas, J. An introduction to multisensor data fusion. Proc. IEEE. 1997, 85, 6 23.

[4] Andrea Masini, Francesco Branchitta, Marco Diani, Giovanni Corsini, Sight enhancement through video fusion in a surveillance system, 14th International Conference on Image Analysis and Processing (ICIAP 2007).

[5] Anjali Malviya; Bhirud, S. G. (2009). Image Fusion of Digital Images, *International Journal of Recent Trends in Engineering*, Vol. 2, No. 3, Nov. 2009, pp.146-148, ISSN 1797-9617.

[6] Z. Zhang and R. Blum, A categorization and study of multiscale-decomposition-based image fusion schemes, Proceedings of the IEEE (1999), pp.1315–1328.

[7] Blum, R. S.; Xue, Z.; Zhang, Z. (2006). An Overview of Image Fusion, *Multi-Sensor Image Fusion and Its Applications*, In Blum, R. S., & Liu, Z (1 ed.), pp. 1-36, Boca Raton:Taylor & Francis.

[8] G. Simone, A. Farina, F. C. Morabito, S. B. Serpico and L. Bruzzone, Image Fusion Techniques for Remote Sensing Applications, Information Fusion, Volume 3, Issue 1, Pages 3-15, March 2002.

[9] A Tutorial on Principal Component Analysis Jonathon Shlens_ Center for Neural Science, New York University New York City, NY 10003-6603 and Systems Neurobiology Laboratory, Salk Insitute for Biological Studies La Jolla, CA 92037 (Dated: April 22, 2009; Version 3.01).

[10] Pohl, C.; Van Genderen, J.L. Multisensor image fusion in remote sensing: concepts, methods and applications. Int. J. Remote Sens. 1998, 19, 823–854.

[11] Oliver Rockinger et al., Pixel-level Image Fusion: The case of Image Sequences, Proceedings of the SPIE, 1998.

[12] Dai, X.; Khorram, S. Data fusion using artificial neural networks: a case study on multitemporal change analysis. Comput. Environ. Urban Syst. 1999, 23, 19–31.

Surya Prasad Potnuru, has done his M.Tech. in Communication and Radar Engineering from Electrical Engineering Department, Indian Institute of Technology, Delhi and B.Tech. from the ECE department, JNTU College of Engineering, Kakinada, Andhra Pradesh. He is a life member of IETE. He has made more than twenty international and national publications. He has more than 10 years of teaching experience. Presently he is working as Associate Professor in the department of ECE, M.V.G.R. College of Engineering, Vizianagaram, Andhra Pradesh, India.

Poornesh Patruni born in Visakhapatnam on 04/11/1991. He is currently pursuing B.Tech course in the Department of ECE, MVGR College of engineering, Vizianagaram, Andhra Pradesh. He is a member of IEEE.

Chandra Sekhar P N R L, has done his M.Tech. In CS & T from CSE Department, Andhra University, Visakhapatnam, Andhra Pradesh. He has one international journal publication and three conference papers. He has more than 9 years of academic experience. Presently he is working as Associate Professor in the department of CSE, GITAM Institute of Technology, GITAM University, Visakhapatnam, Andhra Pradesh, India.