Multi-Focus Image Fusion- A Technical Review

Riddhi Shukla[1] Pragnesh Patel[2]
[1]Student [2]Assistant Professor
[1,2]Department of Computer Engineering
[1,2]Ipcowala Institute of Engineering & Technology, Dharmaj

Abstract— The imaging equipment usually has difficulty in shooting the target object in which all the objects are effectively in focused. Image fusion plays a vital role in many applications. To overcome it multi-focus image fusion technology has emerged. An image is corrupted by noise blurring or limited focal length or due to different sensors and can have the poor visual quality. Image fusion is used to enhance the quality of a degraded image. It is one of the important task and pre-processing step in digital image processing. Image fusion may be categorized into two broad domains which are Spatial Domain and Transform Domain. There are basically three levels of fusion:

1) Pixel level fusion
2) Feature level fusion and
3) Decision level fusion.

The main techniques for pixel level image fusion are average method, principle component analysis, wavelet transform, Brovey transform. For feature level k-means clustering, Region based segmentation and in decision level artificial neural network methods, for gray-scale and RGB images. Earlier proposed method suffers from the noise, artifacts and spectral degradation. The average method leads to the undesirable side effects such as reduced contrast. The weighted wavelet-based method for fusion of PET and CT images has been proposed a pyramid method used for image fusion suffers from blocking artifacts and creates undesired edges. A neural network based fusion requires training sets to get good output.

Key words: Multi-focus image fusion, Image decomposition, DWT, Laplacian Pyramid

I. INTRODUCTION

Image fusion is an important research topic in many related areas such as computer vision, automatic object detection, remote sensing, image processing, robotics, and medical imaging. Image fusion is a sub-field of image processing in which more than one images are fused to create an image where all the objects are in focus. Image fusion is of significant importance due to its application in medical science, forensic and defence departments. The process of image fusion is performed for multi-sensor and multi-focus images of the same scene.

Multi-sensor images of the same scene are captured by different sensors whereas multi-focus images are captured by the same sensor. In multi-focus images, the objects in the scene which are closer to the camera are in focus and the farther objects get blurred. Contrary to it, when the farther objects are focused then closer objects get blurred in the image.

II. LITERATURE REVIEW

[1] Yuji Matsuda, Hajime Hoashi and Keiji Yanai [1] have studied a Deformable Part Model (DPM) feature-fusion method in their paper title “Recognition of Multiple-food images by detecting candidates regions”. In this paper author proposed a two-step method to recognize multiple-food images by detecting candidate regions with several methods and classifying them with various kinds of features. In the first step, they have detect several candidate regions by fusing outputs of several region detectors including Felzenszwalb’s deformable part model (DPM) [2], a circle detector and the JSEG region segmentation[3]. In the second step, they have applied a feature-fusion-based food recognition method for bounding boxes of the candidate regions with various kinds of visual features including bag-of-features of SIFT and CSIFT with spatial pyramid (SP-BoF), histogram of oriented gradient (HoG), and Gabor texture features. In this paper Deformable Part Model (DPM) feature-fusion-based food recognition method used. Advantage of this method is to detect object regions, sliding window approach is adopted in the DPM and to reduce computational cost, linear SVM are used in the DPM method. But disadvantage of this method is Objective is not associating extracted regions with names of food items directly, but listing all the names of the food items.

[2] Cigdem Turan, Kin-Man Lam, Xiangjian [4] have proposed a Discriminant-Analysis of Canonical Correlations (DCC) feature-fusion method in their paper title “Facial Expression Recognition with Emotion-Based Feature Fusion”. In this paper, an emotion-based feature fusion method is using the Discriminant-Analysis of Canonical Correlations (DCC) for facial expression recognition. There have been many image features or descriptors proposed for facial expression recognition. For the different features, they may be more accurate for the recognition of different expressions. In our proposed method, four effective descriptors for facial expression representation, namely Local Binary Pattern (LBP), Local Phase Quantization (LPQ), Weber Local Descriptor (WLD), and Pyramid of Histogram of Oriented Gradients (PHOG), are considered. Supervised Locality Preserving Projection (SLPP) is applied to the respective features for dimensionality reduction and manifold learning. Experiments show that descriptors are also sensitive to the conditions of images, such as race, lighting, pose, etc. Thus, an adaptive descriptor selection algorithm is proposed, which determines the best two features for each expression class on a given training set. These two features are fused, so as to achieve a higher recognition rate for each expression. Advantage of this paper is Fuse the best two feature sets by projecting them into a coherent subspace and achieve higher recognition rate than any of the individual descriptors. But the disadvantage is classifier is learned for each expression which is too time consuming process for classify the expression.

[3] W. C. Olding; J. C. Olivier; B. P. Salmon [5] have studied Markov Random Fields (MRFs) are a popular approach to fusion of multi-source satellite images for the purpose of classification. By modeling pixel labels as random variables and connecting pixels from different images that

occupy the same area, it is possible for the pixels to share information and improve the classification accuracy across all images [6, 7]. Hoberg et al. [8, 9] extends this idea to multi-source images captured at different dates. The simultaneous labeling is performed by constructing a Conditional Random Field (CRF) graph over the pixels with spatial connections to encourage label smoothness and temporal connections to support only probable temporal class transitions. The result is a time series of labeled images that have a higher accuracy compared with classification in isolation and also show the changes of labels through time. Constructing a graph between pixels from multi-resolution images is straightforward as the regular arrangement of the pixels makes it simple to determine which pixels correspond to the same area and should share the same label. The problem is far less simple if we want to simultaneously label segments derived from different images. Unlike pixels which represent points of data, segments are polygons and therefore their geometry should be taken into account when labeling. We propose a method that allows for simultaneous classification of multiple segmented images captured over the same region. The model estimates label assignments for every segment in every image. By enforcing these assignments to be consistent with each other, it is possible for the model to correct segment labelling errors. Several assumptions must hold true for this model to work. Firstly it is assumed that each segment contains only a single land cover class. This assumption applies to all segment based classification methods. Secondly it is assumed that land cover classes do not change between the different image acquisition dates.

[4] Hannandeep Kau, Jyoti Rani [10] have presented the survey of existing fusion schemes and a novel approach of panchromatic and multispectral images fusion using Discrete Wavelet Transform (DWT) in his paper entitled “Image Fusion on Digital Images using Laplacian Pyramid with DWT”. In our proposed system there is discrete wavelet worked on higher level of spatial resolution with frequency coefficient bands, decomposed the source images with the fusion rule in to high and low levels of coefficient bands. By using enhanced laplacian pyramid technique, mapped the local binarized pixels of images within the region which is pixel by pixel fusion. E-Laplacian technique works on gray-scale images, colored or RGB images and medical images. The fmal step performed is an inverse DWT with the new band coefficients to construct the fused image.in this paper Laplacian Pyramid and Discrete wavelet transform methods used. Advantages of this method are Laplacian Pyramid proved to be higher than the standard techniques in terms of edge preservation and is best within the terms of quality. And DWT gave improved results than the conventional strategies like it had good spectral preservation. But disadvantages of method are Laplacian Pyramid needs High Processing Time and discrete Wavelet Transform was poor and it had high shift invariance that reduced the potency.

[5] Yukhe Lavinia, Holly H. Vo, Abhishek Verma [11] have studied on Fusion method that concatenates two and three deep convolutional neural networks (CNN). After examining the classification accuracy of each deep CNN candidates, we inserted a 100 dimensional fully-connected layer and extracted features from the new 100 dimensional and the last fully-connected layers to create a pool of candidate layers. We form our fusion models by concatenating two or three layers from this pool—one from each model—and trained and tested them on the large-scale Stanford 40 Actions still image dataset. We forwarded the concatenated features to Random Forest and SVM for classification. Compared to video, still images cost less memory footprint. The efficacy of still image in capturing the desired action, however, depends on the actions. Some are more suitable to be represented as still image, while others as video. Actions such as sleeping, reading, or running could still be accurately perceived when represented in a still image. On the other hand, performing certain dance moves, making certain hand gestures, or doing other sequential motions would lose their meaning when being represented in a still image. These actions require image sequences to be perceived correctly and thus video would be a better representation medium. Fusion of two high performing deep CNN models achieved better action recognition accuracy than an individual model, and that the fusion of three models increased the performance further. Our experiments demonstrated that the fusion of two deep CNNs generated about 2% increase in accuracy, and that adding another powerful deep CNN model to the fusion duo increased another 2% accuracy. We have also investigated the difficulty level of the action categories. Further investigation could still be done to see if adding another high-performing model to the fusion trio would improve the accuracy. Another idea to improve accuracy is to include object localization method in the fusion methodology.

III. Conclusion of Literature Review

- To detect object regions, sliding window approach is adopted in the DPM and to reduce computational cost, linear SVM are used in the DPM method.
- Objective of our system is not associating extracted regions with names of food items directly, but listing all the names of the food items.
- Fuse the best two feature sets by projecting them into a coherent subspace achieve higher recognition rate than any of the individual descriptors.
- A classifier is learned for each expression which is too time consuming process for classify the expression.
- Popular approach to fusion of multi-source satellite images for the purpose of classification. By modeling pixel labels as random variables and connecting pixels from different images that occupy the same area, it is possible for the pixels to share information and improve the classification accuracy across all images.
- Laplacian Pyramid proved to be higher than the standard techniques in terms of edge preservation and is best within the terms of quality.
- DWT gave improved results than the conventional strategies like it had good spectral preservation.
- Fusion of two high performing deep CNN models achieved better action recognition accuracy than an individual model, and that the fusion of three models increased the performance further.
- Fusion of two deep CNNs generated about only 2% increase in accuracy.
- Laplacian Pyramid needs High Processing Time.

REFERENCES

[1] Matsuda, Yuji, Hajime Hoashi, and Keiji Yanai. "Recognition of multiple-food images by detecting candidate regions." Multimedia and Expo (ICME), 2012 IEEE International Conference on. IEEE, 2015.
[2] P.F. Felzenszwalb, R.B. Girshick, D. McAllester, and D. Ramanan, “Object detection with discriminatively trained part-based models,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.32, no. 9, pp. 1627–1645, 2010.
[3] Y. Deng and B. S. Manjunath, “Unsupervised segmentation of colortexture regions in images and video,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 8, pp. 800–810, 2001.
[4] Turan, Cigdem, Kin-Man Lam, and Xiangjian He. "Facial expression recognition with emotion-based feature fusion." Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2015 Asia-Pacific. IEEE, 2015.
[5] Olding, Willem C., Jan C. Olivier, and Brian P. Salmon. "A Markov random field model for decision level fusion of multi-source image segments." Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International. IEEE, 2015.
[6] R. Nishii, “A markov random field-based approach to decision-level fusion for remote sensing image classification,” Geoscience and Remote Sensing, IEEE Transactions on, vol. 41, no. 10, pp. 2316–2319, Oct 2003.
[7] A. H. Solberg, T. Taxt, and A. K. Jain, “A markov random field model for classification of multisource satellite imagery,” Geoscience and Remote Sensing, IEEE Transactions on, vol. 34, no. 1, pp. 100–113, 1996.
[8] T. Hoberg, F. Rottensteiner, and C. Heipke, “Context models for crf-based classification of multitemporal remote sensing data,” ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 25, pp. 128–134, 2012.
[9] T. Hoberg, F. Rottensteiner, R. Queiroz Feitosa, and C. Heipke, “Conditional random fields for multitemporal and multiscale classification of optical satellite imagery,”Geoscience and Remote Sensing, IEEE Transactions on, vol. 53, no. 2, pp. 659–673, Feb 2015
[10] Kaur, Harmandeep, and Jyoti Rani. "Image fusion on digital images using Laplacian pyramid with DWT." Image Information Processing (ICIIP), 2015 Third International Conference on. IEEE, 2016.
[11] Lavinia, Yukhe, Holly H. Vo, and Abhishek Verma. "Fusion Based Deep CNN for Improved Large-Scale Image Action Recognition." Multimedia (ISM), 2016 IEEE International Symposium on. IEEE, 2016.