International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 11 (2018) pp. 8831-8836


# Thermal and Visible Video Fusion Using Curvelet Transform

**X. Blessy Theresa[1] and K. Madheswari[1]**

[1] *Department of Computer Science and Engineering*
*SSN College of Engineering, Chennai, India.*

**Abstract.**

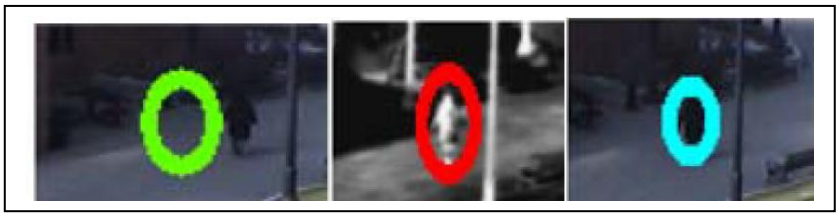
Traditionally, there has been keen interest in human movement from a wide variety of disciplines. Generally, Red Blue Green (RGB) visible cameras are used in surveillance. But the main drawback of Closed Circuit Television (CCTV) cameras arise from their reliance on reflected light. Thermal infrared camera systems are resilient to these effects and are therefore very useful in a security context. However, thermal video cannot detect objects that are at the same temperature as the background and it contains very little texture information. Hence, we overcome these challenges by fusing both thermal and visible camera videos

**Keywords:**Thermal video, Visible video, Fusion, Curvelet transform.

## INTRODUCTİON

Human beings, whether in a standstill or in motion, have an extraordinary visual capability of detecting objects and tracking them in the environment. This powerful property of the human visual system allows people, e.g. to manoeuvre in crowded pavements without bumping into other pedestrians. Implementation of object tracking functionality in computer vision systems is a challenging task and has attracted researcher's interest for decades. In its simplest way, tracking can be defined as a problem of estimating the trajectory of an object in the image plane as it moves around a scene .In other words, a tracker assigns consistent labels to the tracked objects in different frames of a video. Pedestrian detection has been studied widely in the context of video surveillance with fixed cameras and stationary backgrounds. Various approaches have been proposed including background elimination and analysis, periodic motion, symmetry, silhouette shape analysis of the foreground etc., .Pedestrian tracking involves two steps: 1) Identifying the interested pedestrians 2) Following of such pedestrians from frame to frame 3) Determine the object tracks for recognizing their behaviour. All of the mentioned steps are very challenging due to:1) complex backgrounds 2) Poor illumination3) Object occlusion.

Recently, thermal video systems have gained importance in many application areas such as medicine, industry, and military. Thermal infrared image is formed due to emission of heat waves due to heat whereas visual band images that are formed due to light reflection. Surveillance in visible band is highly matured and successful due to the advancement of sensor technology, availability of computing power, and high digital storage capacity. However, the image processing task and applications are challenging if the visible videos are obtained under non-ideal environments. The resulting videos depends on the brightness of external light source which sometimes might be insufficient, particularly in environments with heavy clouds, fog, rain, snow, smoke, darkness and at night-time. In contrast, thermal videos do not depend on any external light source. It is robust against any illumination changes. However, thermal videos are sensitive to temperature changes in the surrounding environment. Currents of cold or warm air could influence the performance of the thermal system. thermal video is also sensitive to variations in the heat patterns of the objects. Thus, thermal and visible imaging acquires highly complementary strengths and limitations. Hence, it is advantages to acquire and process both visual and thermal videos concurrently and jointly in many applications leading to the fusion of information. Video fusion is the process of combining information from two or more videos of a given scene into a single representation. This process is intended for encoding information from source images into a single and more informative one, which could be suitable for further processing or visual perception. RGB camera already has been producing reliable results for scenes with the constant illumination and steady backgrounds. Due to illumination changes, however, it is challenging to handle using inputs from a conventional CCD camera.



a) Visible image b) Thermal image c) Fused image

**Figure 1.** Visible, thermal, and fused images

Hence we propose a method of utilizing additional information, a thermal camera which produces for each pixel a gray scale mapping of the infrared radiation at the corresponding location. Traditional video fusion methods fuse the source videos using static image fusion methods frame-by-frame without considering the information in temporal dimension. The temporal information can't be fully utilized infusion procedure. Aiming at this problem, a visible and infrared video fusion method based on Uniform discrete curvelet transform.

## RELATED WORK

Thermal and visible videos show many in efficiency when used separately. These disadvantages have been discussed by several researchers: Congxia Dai, Yunfei Zheng and Xin Li[1] propose a generalized EM algorithm to decompose infrared images into background and foreground layers. It proposes to locate pedestrians from the foreground layer in two steps:1) shape-based classification - does this object contain any pedestrian?2) appearance based localization. In the classification step, compactness and leanness of the object are extracted and trained by support vector machine (SVM).In the localization step, a modified principal component analysis (PCA) technique is developed to statistically infer the most likely position for each pedestrian. Two frames are grouped together if and only if their Hausdorff distance is below a pre-selected threshold. The disadvantages are: The thermal signature of a person is constantly varying due to the walking motion. Moreover when two pedestrians walk closely or pass by each other the overlapped shape of pedestrians experiences severe deformation, which makes tracking even more difficult.

Harsh Nanda, Larry Davis [2] introduce probabilistic templates to capture the variations in human shape .The training data used for developing the probabilistic template consists of 1000 128x48 images all of which are known to contain humans in different poses and orientation but having the same height.128 x 48 window is moved over the entire image and the cp calculated for each pixel (x, y). The probability map is then threshold. The value threshold is calculated based on the training data, which is a set of 1000 rectangular boxes containing pedestrians. We calculate the mean and standard deviation for the pixels belonging to pedestrians and pixels belonging to background. The disadvantages are: pedestrians are not the only hot objects. Lamps, cars and many other objects are also hot and hence are captured by infrared cameras. Lastly, human body does not emit heat uniformly, The amount of heat emitted depends on the body part being captured, dress of the person, orientation of the person, time of day and many other factors. Daw-Tung Lin and Kai-Yung Haung [3] proposed to fuse an object using multi-variance Gaussian function.

This paper inspires a paradigm of human visual perception for multi-camera object tracking, collaboration, and fusion through distributed cameras and computers. The benefit of multiple cameras is that they can expand the monitoring zone. For single-camera tracking, the Kalman filter is utilized to estimate the position of the object in the next frame. To obtain the feature of object movement, the object is skeletonised to extract the gait of the moving object. To obtain color feature, the object is decomposed into upper and lower parts via a K-mean clustering algorithm. Then, each camera conveys the features to the object- fusing system. To fuse an object, the multi-variance Gaussian function is adopted. The main problem in the non-overlapping case is that objects will disappear in the gap between the FOV of two cameras .This method improves the completeness of object extraction, while the results obtained by direct background subtraction in the GMM model were either noisy or too rough. CiarnConaire ,Noel E. OConnor ,Alan Smeaton A [4] suggests a framework that efficiently combines features for robust tracking based on fusing the outputs of multiple spectrogram trackers. the main drawback of using histograms or spectrograms is that their memory requirements (and hence their computational load) increase exponentially as more features are added.

To overcome these difficulties, tracking can be achieved by splitting the feature set over several histogram trackers and combining their outputs. All trackers evaluate a series of potential object position hypotheses and return a similarity score for each one. The combined score for each hypothesis is computed by multiplying the similarity scores from each tracker. Also, the tracking framework we use can incorporate a mean-shift approach to object localisation, allowing rapid object tracking .Mean-shift is an iterative kernel-based procedure to locate the local mode in a distribution. It has been successfully used in many tracking applications to efficiently locate objects in subsequent frames under the assumption that the object overlaps itself in consecutive frames. spatiogram models do not updated during tracking, to account for changes in object appearance . And spatiogram trackers do not scale well to higher dimensions [5-12].

## PROPOSED VİDEO FUSİON FRAMEWORK

This section explains the proposed video fusion algorithm. First, the visible and thermal videos are converted into frames. Next, the visible frames are converted from RGB to LAB form. From the the LAB form of the visible image the luminance alone is extracted. Then, the thermal and visible frames are reduced using curvelet transform by decomposing the frames into co-effecients and then fusing them using the maximum selction rule.

### Video to Frame conversion

The video is read and stored in an object. Then the number of frames in the video is calculated using that object. For the total number of frames in the video a loop is set. Now the number of the frame is concatenated with the image format .Finally each frame is stored with their corresponding names.

### RGB to LAB conversion

The Lab colour space describes mathematically all perceivable colours in the three dimensions L for lightness and a and b for the colour opponents green red and blue yellow. One of the most important attributes of the Lab model

is device independence. This means that the colors are defined independent of their nature of creation or the device they are displayed on. The characteristics of the LAB color space makes it suitable for extracting global color features from a digital image. Therefore we convert the analyzed RGB images into LAB format. The characteristic known as luminance or intensity is represented on the axis named L, that is perpendicular on a pile of ab planes, each one containing all the possible colours for a given luminance.

**Step 1:** The RGB image is converted into LMS cone space. Since l is the transform of LMS cone space. We convert the image into LMS cone space in two steps. The first step is the conversion from RGB to XYZ tristimulus values.

$$\begin{Bmatrix} L \\ A \\ B \end{Bmatrix} = \begin{Bmatrix} 0.5141 & 0.3239 & 0.1604 \\ 0.2651 & 0.6702 & 0.0641 \\ 0.0241 & 0.1228 & 0.8444 \end{Bmatrix} \begin{Bmatrix} R \\ G \\ B \end{Bmatrix} \quad (1)$$

Then, the LMS cone space is constructed from XYZ space as follows:

$$\begin{Bmatrix} L \\ M \\ S \end{Bmatrix} = \begin{Bmatrix} 0.3897 & 0.6890 & -0.0787 \\ -0.2298 & 1.1834 & 0.0464 \\ 0.0000 & 0.0000 & 1.0000 \end{Bmatrix} \begin{Bmatrix} X \\ Y \\ Z \end{Bmatrix} \quad (2)$$

**Step 2:** The data in this color space shows a great deal of skew, which we can largely eliminate by converting the data to logarithmic space.

$$\mathrm{L} = logL$$

$$\mathrm{M} = logM$$

$$\mathrm{S} = logS$$

**Step 3:** The logarithmic LMS cone space is de-correlated using principal component analysis (PCA) to treat the three color channels separately.

$$\begin{Bmatrix} L \\ A \\ B \end{Bmatrix} = \begin{Bmatrix} \frac{1}{\sqrt{3}} & 0 & 0 \\ 0 & \frac{1}{\sqrt{6}} & 0 \\ 0 & 0 & \frac{1}{\sqrt{2}} \end{Bmatrix} \begin{Bmatrix} 1 & 1 & 1 \\ 1 & 1 & -2 \\ 1 & -2 & 0 \end{Bmatrix} \begin{Bmatrix} L \\ M \\ S \end{Bmatrix} \quad (3)$$

**Image decomposition using curvelet transform**

Curvelet transform is a powerful multi-scale multi-orientation image decomposition technique. It was developed to solve the problem of curve singularities. As an image analysis tool, it differs from other directional wavelet transforms in the degree of localization in orientation, which varies with scale. It provides a strong directional characterization in which elements are highly anisotropic at fine scales. With these properties, Curvelet solve the isotropic and limited directional analysis of classic wavelet transform. Unlike the wavelet transform, it has directional parameters. Suppose that we have an object f which exhibits an edge. Upon subband filtering, each resulting fine-scale subband output sf will contain a map of the edge in f, thickened out to a width 22s according to the scale of the subband filter operator. This gives the subband the appearance of a collection of smooth ridges. When we smoothly partition each subband into squares, we see either an empty square if the square does not intersect the edge or a ridge fragment. Moreover, the ridge fragments are nearly straight at fine scales, because the edge is nearly straight at fine scales. Such nearly straight ridge fragments are precisely the desired input for the ridgelet transform.

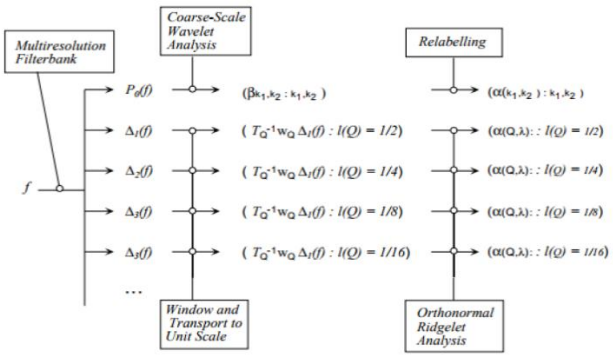



**Figure 2:** Image decomposition using Curvelet Transform

**Fusion using Maximum Selection Rule**

The coefficients obtained by decomposition of source images are fused using absolute maximum fusion rule. Image fusion is the technique to combine relevant information from two or more than two images of the same scene into only one composite image that is more informative and which is mostly suitable for human and machine interaction. The fusion rule used in this work is defined as follows

$$F(x,y) = \max(I(x,y), II(x,y)) \quad (4)$$

Where, *F (x,y)* and *I(x,y)*, *II(x,y)* denotes the fused and input images respectively. In this fusion rule first the two source images, image *I* and image *II* to be fused are read and apply as input for fusion. Then perform independent curvelet decomposition of the two images until level *L* to get approximation and detail coefficients. Finally apply pixel based algorithm for approximations which involves fusion based on taking the maximum valued pixels from approximations of source images *I* and *II*.

**Steps involved in Maximum Selection Rule Algorithm**

**Step 1:** Read the two source images, image *I* and image *II* to be fuse and apply as input for fusion.

**Step 2:** Perform independent curvelet decomposition of the two images until level *L* to get approximation and detail coefficients.

**Step 3:** Apply pixel based algorithm for approximations which involves fusion based on taking the maximum valued pixels from approximations of source images *I* and *II*.

**Step 4:** Thus, the final fused transform corresponding to approximations through maximum selection pixel rule is obtained.

**Step 5 :** Concatenation of fused approximations and details gives the new coefficient matrix.

**RESULTS**

The visible image frame which was obtained from the conversion of visible video to frames is shown in Fig 4.1(a). Fig 4.1(b) shows the thermal frame extracted from a thermal video .The visible frame is converted from RGB to LAB in Fig 4.1(c) and luminance component is extracted from the LAB form which is shown in Fig 4.1(d). Then both the visible and thermal frames are fused using curvelet transform and the resultant fused image is shown in Fig 4.1(e)

.

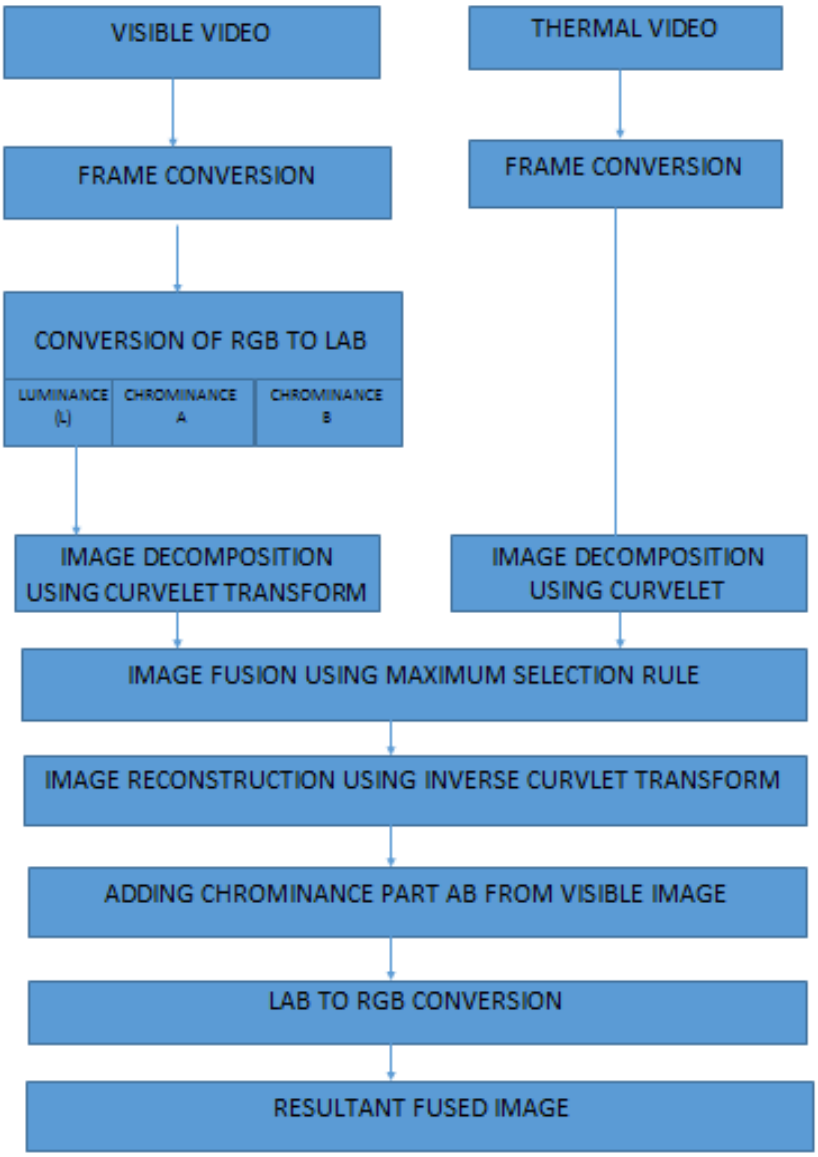



**Figure 3.** Block diagram of Proposed video fusion algorithm

(a) Visible image

(b) Thermal image

(c) Lab conversion

(d) Luminance Component

(f) Fused image

**Figure 4.** Fused image using proposed video fusion algorithm

**Table 1.** Comparison of fusion techniques based on error measures

| **Error measures** | **Curvelet transform** | **Dwt** |
|---|---|---|
| PSNR | 69.7866 | 69.5855 |
| Correlation Co-efficient | 0.4492 | 0.4206 |
| Entropy | 7.1934 | 7.1420 |

The qualtiy of the proposed fusion technique is compared with another fusion technique based on several error measures which is shown in the table below

From the above mentioned table , it can be seen that curvelet transform has higher values in all the measures which shows that it has higher efficiency when compared with other fusion techniques.

## CONCLUSİON

Pedestrian tracking using thermal and visible videos when used separately show certain disadvantages. Hence, a technique of video fusion is proposed which enhances pedestrian tracking by detecting features efficiently. However in order to perform fusion of both thermal and visible videos frames are subject to curvelet transform and then fused using maximum selection rule.

## REFERENCES

[1] Congxia Dai, Yunfei Zheng and Xin Li, Layered Representation for Pedestrian Detection and Tracking in Infrared Imagery, IEEE Computer Society Conference,2005

[2] Harsh Nanda , Larry Davis, Probabilistic Template Based Pedestrian Detection in Infrared Videos, Intelligent vehicle Symposium IEEE,2002.

[3] Daw-Tung Lin and Kai-Yung Haung, Collaborative Pedestrian Tracking and Data Fusion With Multiple Cameras,IEEE,2011.

[4] Ciarn Conaire ,Noel E. OConnor ,Alan Smeaton,Thermo-visual feature fusion for object tracking using multiple spatiogram trackers ,Springer- Verlag Journal , VOL. 6, NO. 4,2011.

[5] M. Isard and J. MacCormik,BraMBLe: A Bayesian Multiple-Blob Tracker,Compaq Systems Research Center,2001

[6] Md Zahangir Alomand Tarek M.Taha,Robust Multi-view Pedestrian Tracking Using Neural Networks,2017.

[7] K.Madheswari and N.Venkateswaran, "An optimal weighted averaging fusion strategy for thermal and visible images using dual tree discrete wavelet transform and self tunning particle swarm optimization", Multimedia Tools and Applications,. PP: 1-22. October-2016.

[8] K Madheswari and N Venkateswaran ,"Swarm intelligence based optimisation inthermal image fusion using dual tree discrete wavelet transform" in Quantitative Infrared Thermography Journal ,Taylor and Francis, First Online, pp: 1-20, september 16, 2016.

[9] K Madheswari and N Venkateswaran , "Swarm Intelligent based Contrast Enhancement Algorithm with Improved Visual Perception for Color Images"Multimedia Tools and Applications,(springer, IF=1.530,Thomson reuters), First online(June 29).

[10] K Madheswari and N Venkateswaran "Particle Swarm Optimization aided Weighted Averaging Fusion Strategy for CT and MRI Medical Images", International journal of biomedical engineering and technology (Scopus indexed), (Accepted)

[11] Madheswari . K , N.Venkateswaran, N. Ganeshkumar, " Entropy optimized contrast enhancement for gray scale images", International journal of Applied Engineering Research (Scopus Indexed) ,Volume 10, Number 55, ISSN 0973-4562, pp. 1590-1595, 2015.

[12] K Madheswari and N Venkateswaran , "Fusion of Visible and Thermal images using Curvelet Transform and Brain Storm Optimization", IEEE TENCON 2016, Singapore 2016.