



International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS)

www.iasir.net

Video inpainting of objects using modified Patch based technique

Mrs.B.A.Ahire,
Student, M.E.(Computer),
MCOERC, Nasik (MS), India

Prof.Neeta A. Deshpande,
Associate Professor, Computer Department,
MCOERC, Nasik (MS), India

Abstract: An efficient moving object removal using modified exemplar based inpainting algorithm is proposed in this paper. In the proposed method, 3D volume of video is converted into 2D slices for maintaining spatial and temporal consistencies. Background subtraction technique is used for object detection and tracking. Then, modified exemplar-based image inpainting algorithm is applied to get the improved results. After combination of completed slices, sequences of virtual contours are formed to retrieve the most similar postures among the number of available postures. Key posture selection and indexing are used to minimize the complexity of posture sequence retrieval. Synthetic posture generation method is used to increase the number of posture generation database.

Keywords: Video inpainting, texture synthesis, Exemplar based inpainting.

Introduction

Inpainting is the technique of reconstructing lost or damaged parts of images and videos. The concept of image inpainting existed a very long years back and from the birth of computer vision, researchers are looking for a way to carry out this process automatically[1]. Video inpainting has a huge number of pixels to be inpainted which makes the searching space much more complex. Application of image inpainting techniques directly into video inpainting without taking into consideration the temporal factors will ultimately lead to failure because it makes the frames inconsistent with each other[1]. Depending upon the way the damaged images are restored, the techniques are classified into three groups: texture synthesis-based methods, partial difference equation-based methods and patch-based methods[1]. The texture synthesis based methods doesn't contain structural information whereas, PDE-based methods results in blurring artifacts[12]. Patch based methods leads to high-quality effects maintaining consistency of local structures[16].

Patwardhan[1] proposed a video inpainting technique for inpainting missing parts of a video sequence when recorded with a moving or stationary camera. In the preprocessing step, three mosaics namely foreground, background & optical flow are constructed. This method can't maintain continuity along a temporal axis. Wexler[3] proposed framework for the completion of missing information based on local structures. Cheung[4] proposed a patch based method for video inpainting which is suitable for synthesizing data and doesn't contain structured information but the inpainting errors are of low resolution with over-smoothing artifacts. In the system proposed by Sun et. al. [5], an user only needs to draw one structure line to perform inpainting. Object based approaches proposed in [7] and [8] employ video inpainting mechanism. In [8], Cheung et. al. proposed an efficient object-based video inpainting technique for dealing with videos recorded by a stationary camera. If the number of postures in the database is insufficient, the inpainting result could not be satisfactory.

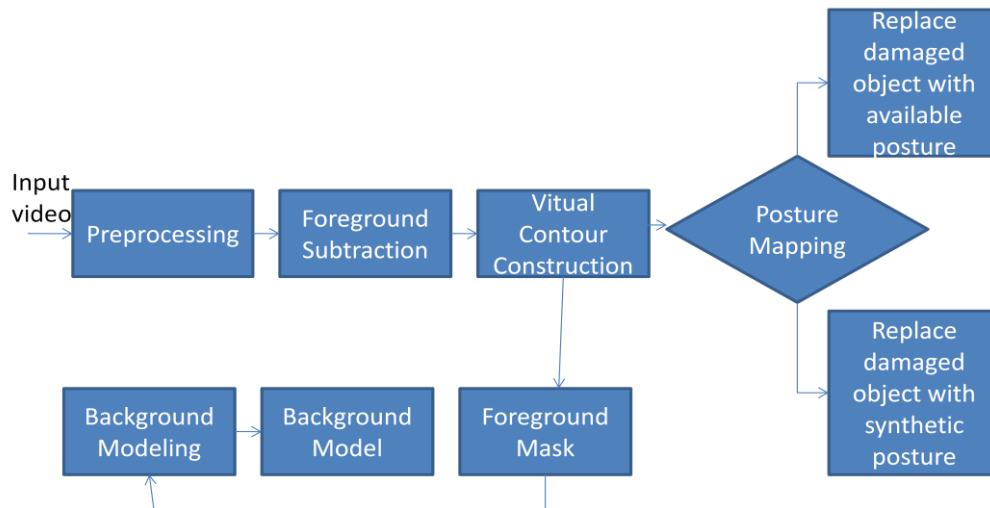
Exemplar-based methods often have difficulty handling spatial consistency and temporal continuity problems [1]. The previous approaches proposed can only maintain spatial consistency or temporal continuity; they cannot solve both problems at the same time. On the other hand, some of the proposed approaches can deal with spatial and temporal information simultaneously, but they lead to the over-smoothing artifacts problem[16].

The rest of the paper is described as follows: Section II provides description of the proposed system. In section III, we present the proposed object extraction method, posture based inpainting scheme for the moving objects and in section IV we evaluate the system's performance. Section V contains some concluding remarks.

Overview of Proposed System

We have proposed an object-based video inpainting scheme that can maintain the spatial consistency and temporal motion continuity of an object simultaneously. The system can also handle the problem of insufficiency of available postures. The system diagram for the proposed work is shown in Figure 1. The proposed system consists of four modules: Object detection and tracking, virtual contour construction, Key posture selection and mapping and synthetic posture generation. First, an object is detected and tracked with the help of background subtraction algorithm. In the next module, a scheme is proposed that is able to derive the virtual contour of an occluded object.

Figure 1 Block Diagram of the Proposed System



Object Completion Using Modified Inpainting Technique

A. Background Subtraction Algorithm

A Video is a function of image over time t $f(x,y,t) = f(x,y)$. The frame rate means the number of images per second. The algorithm for the background subtraction is as follows:

1. Consider the first $M \times N$ colour frame F_i , where $i=1$ and set

$$R(x,y,n(x,y)) = R1(x,y,n(x,y))$$

$$G(x,y,n(x,y)) = G1(x,y,n(x,y))$$

$$B(x,y,n(x,y)) = B1(x,y,n(x,y))$$

$$C(x,y,n(x,y)) = 1$$

2. Capture the next frame F_i , where $i = i + 1$

3. Sequentially calculate the $Q_i(x,y)$ for each pixel from the frame F_i where, $Q_i(x,y) = R_i(x,y), G_i(x,y), B_i(x,y)$, $x = 0$ to $M-1$, $y = 0$ to $N-1$.

If all the pixels of the frame F_i have been calculated go to step 2.

4. Calculate the colour difference $D_i(x,y,n(x,y))$ between the fetched pixel and each classified cluster of the pixel located at (x,y) .

$$D_i(x,y,n(x,y)) = (R_i(x,y) - R(x,y,n(x,y))) + (G_i(x,y) - G(x,y,n(x,y))) + (B_i(x,y) - B(x,y,n(x,y)))$$

5. If the difference between two pixel values is greater than the threshold, the movement of that pixel is considered as significant. Another fixed threshold Th is introduced here. The fixed threshold which is set to be a very small value is used to identify pixels that are very likely to be part of the background.

$$D_{th} = 255, \text{ if } D_i(x,y,n(x,y)) \text{ is greater than or equal to } Th$$

$$D_{th} = 0, \text{ if } D_i(x,y,n(x,y)) \text{ less than } Th$$

After thresholding the difference image, each pixel is classified as a foreground or background image pixel. For the case where $D_{th}(x,y) = 255$, that pixel is classified as a foreground pixel. For $D_{th}(x,y) < 255$, the pixel is classified as a background pixel.

6. If the pixel (x,y) is classified as foreground pixel (i.e. $D_{th}(x,y)=255$), the background model at that location will stay the same, the background model at (x,y) will not be updated. If (x,y) is classified as a background pixel, then the background model updates according to how certain it is to be a background pixel. A smaller fixed threshold, is used to define a pixel with high likelihood of being a background pixel. (The fixed threshold Th is assigned as 10.) That is the case when $D_{th}(x,y)=0$. In this case we update the background model at that location to the corresponding pixel in the current frame (i.e. setting the learning rate, as 1). If the difference is less than the adaptive threshold, and greater than the fixed threshold, i.e. $D_{th}(x,y) = 127$, then we update the background model with a much lower learning rate.

7. At the end, morphological operation is performed.

B. Patch based Image Inpainting Technique

Spatiotemporal slices of a video are used to derive virtual object contours, which are then used as features to infer the occluded object poses. More specifically, after object extraction and removal, we sample a 3-D video volume comprised of several consecutive frames to obtain a set of directional 2-D spatio-temporal slices. To

obtain continuous object trajectories, we use the patch based image inpainting to complete missing regions in the spatio-temporal slices. The method first determines the filling order of the missing regions based on the priority of the missing region. The priority of the pixel of the missing region is calculated as follows :

$$\alpha * C_p + \beta * D_p$$

Because of α and β , the C_p term becomes more prominent as compared to D_p . Hence, we have set the values of α and β as 0.2 and 0.8 respectively. There can be the number of combinations of terms C_p and D_p . But the term which we have used gives very good results as compared to the previous methods.

C. Key-Posture Based Sequence Matching

After composing a sequence of consecutive virtual contours, we use them to match the most similar posture sequence in the set of available postures to complete the occluded objects. To simplify the posture sequence matching process, we use the key posture selection method to select the most representative postures from among the available postures. A posture is deemed a key posture if its degree of dissimilarity to all key postures exceeds a predefined threshold, that is empirically set to be 0.08.

After the key posture selection process, each key posture is labeled with a unique number. The virtual trajectory of each occluded posture is then matched with the key posture that has the most similar context. If a virtual contour cannot be matched in this way, a special label is assigned to it. As a result, a sequence of contiguous available postures and virtual contours can be converted into a string of key-posture labels based on the temporal order. After the encoding process, the problem of retrieving the most similar sequence of postures for a sequence of virtual contours becomes a substring matching problem that, given an input segment of codes, searches for the most similar substring in a long string of codes. The occluded objects are then replaced with the retrieved sequence of available postures. As a result, half of the labels in the initial string are derived from available postures and the remaining labels are obtained from the virtual contours.

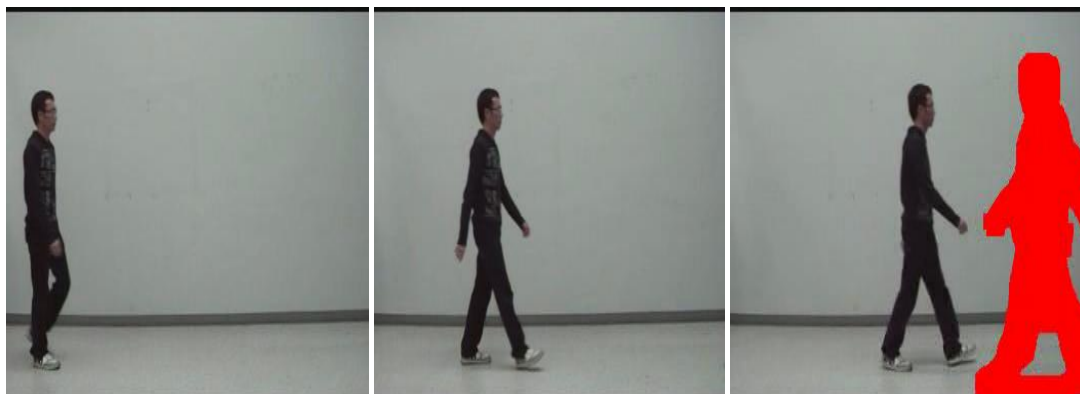
D. Synthetic posture generation

The problem of insufficient postures usually arises when the occlusion period for a to-be-completed object is long, resulting in many reconstructed virtual contours, or when the object's non-occlusion period is too short to collect a sufficiently rich set of non-occluded postures.

First, the skeleton of a virtual contour that cannot find a good match in the posture database is extracted, which is also used to extract the skeletons of all available postures. Then, the constituent components of each selected key-posture are decomposed based on the distribution of the variance in alignment errors between every two aligned key-postures. The component decomposition result of key postures is used to help segment the extracted skeletons into their constituent components. We use the segmented skeleton components of a virtual contour to retrieve similar posture components, which are then used to synthesize new postures.

Experimental Results

We have used test sequences to evaluate the efficiency of our method. In the Fig. 3(a), input test sequence is shown. Fig. 3(b) shows the virtual contours of the occluded object and its corresponding matched postures. From fig. 3(c), the proposed method is still effective in maintaining the spatial consistency and temporal continuity.



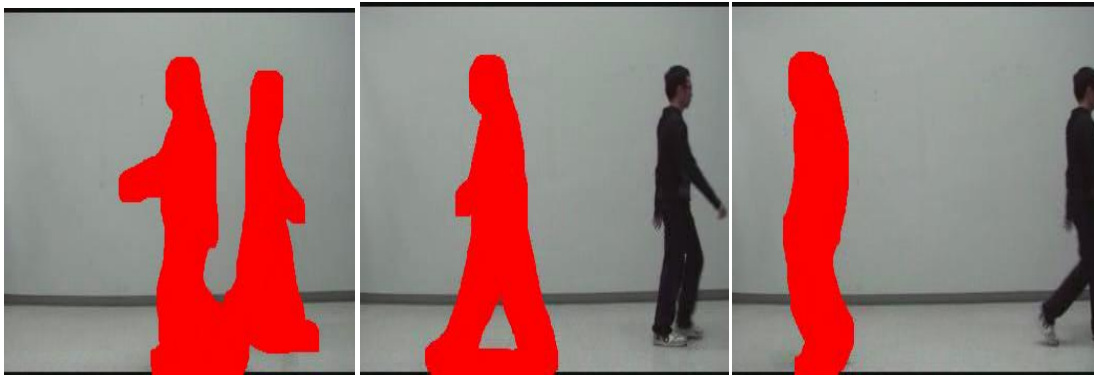


Fig. 3(a). Frames extracted from the original video



Fig. 3(a).Images extracted from the original frames

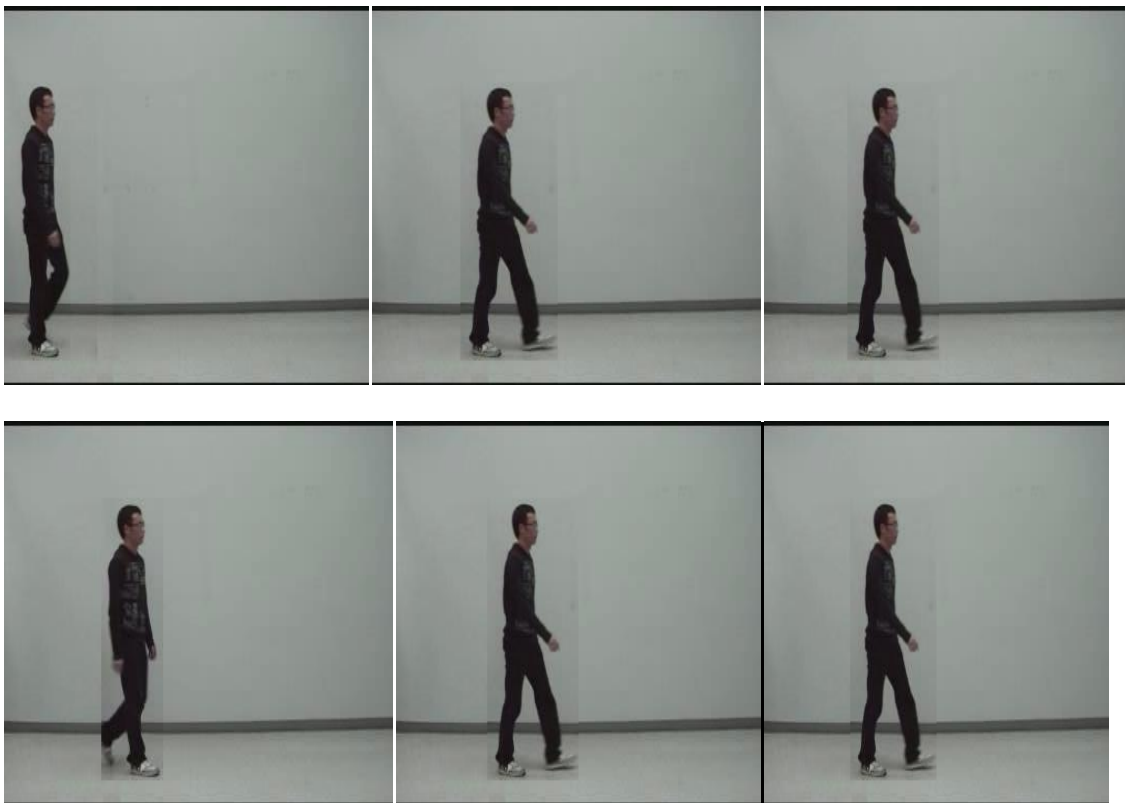


Fig.3(c).Output frames after merging

The proposed system was implemented on a PC equipped with Intel Core2 Duo CPU 2.83 GHz and 3.5 GB system memory. The system has been implemented in Visual C#.

V. Conclusion

This system presents a framework for object removal in a video. The proposed method is comprised of four modules : Object detection and tracking, Virtual trajectories generation, key posture-based sequence retrieval, and synthetic posture generation. We have also modified the patch based image inpainting algorithm for getting the better results as compared to the previous methods. Our experiment results show that the proposed method removes objects with good subjective quality in terms of the object's spatial consistency and temporal motion continuity. It also avoids over-smoothing artifacts and compensates for insufficient available postures. The proposed method also works with the non-linearity of the occluded object. If an object moves nonlinearly during an occlusion period, the virtual trajectories may not compose sufficiently accurate postures.

The proposed system still has a few constraints. Currently the proposed method does not deal with the illumination change problem that occurs if lighting is not uniform across the scene.

VI. References

- [1] K. A. Patwardhan, G. Sapiro, and M. Bertalmio, "Video inpainting under constrained camera motion," *IEEE Trans. Image Process.*, vol. 16, no. 2, pp. 545–553, Feb. 2007.
- [2] Y. T. Jia, S. M. Hu, and R. R. Martin, "Video completion using tracking and fragment merging," *Visual Comput.*, vol. 21, no. 8–10, pp. 601–610, Aug. 2005.
- [3] Y. Wexler, E. Shechtman, and M. Irani, "Space-time completion of video," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 3, pp. 1–14, Mar. 2007.
- [4] V. Cheung, B. J. Frey, and N. Jojic, "Video epitomes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, San Diego, CA, Jun. 2005, pp. 42–49.
- [5] T. K. Shih, N. C. Tang, and J.-N. Hwang, "Exemplar-based video inpainting without ghost shadow artifacts by maintaining temporal continuity," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 3, pp. 347–360, Mar. 2009.
- [6] Y. Shen, F. Lu, X. Cao, and H. Foroosh, "Video completion for perspective camera under constrained motion," in *Proc. IEEE Conf. Pattern Recognit.*, Hong Kong, China, Aug. 2006, pp. 63–66.
- [7] J. Jia, Y.-W. Tai, T.-P. Wu, and C.-K. Tang, "Video repairing under variable illumination using cyclic motions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 5, pp. 832–839, May 2006.
- [8] C. S. Cheung, J. Zhao, and M. V. Venkatesh, "Efficient object-based video inpainting," in *Proc. IEEE Conf. Image Process.*, Atlanta, GA, Oct. 2006, pp. 705–708.
- [9] M. Bertalmio, A. L. Bertozzi, and G. Sapiro, "Navier-stokes, fluid dynamics, and image and video inpainting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Kauai, HI, Dec. 2001, pp. 355–362.
- [10] T. Ding, M. Sznajder, and O. I. Camps, "A rank minimization approach to video inpainting," in *Proc. IEEE Conf. Comput. Vis.*, Rio de Janeiro, Brazil, Oct. 2007, pp. 1–8.
- [11] A. Efros and T. Leung, "Texture synthesis by non-parametric sampling," in *Proc. IEEE Conf. Comput. Vis.*, 1999, vol. 2, pp. 1033–1038.
- [12] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in *Proc. ACM SIGGRAPH*, 2000, pp. 417–424.
- [13] A. Criminisi, P. Perez, and K. Toyama, "Region filling and object removal by exemplar based image inpainting," *IEEE Trans. Image Process.*, vol. 13, no. 9, pp. 1200–1212, Sep. 2004.
- [14] T. Huang, S. Chen, J. Liu, and X. Tang, "Image inpainting by global structure and texture propagation," in *Proc. ACM Conf. Multimedia*, Augsburg, Germany, Sep. 2007, pp. 517–520.
- [15] I. Haritaoglu, D. Harwood, and L. S. Davis, "W4: Who? When? Where? What? A real-time system for detecting and tracking people," in *Proc. IEEE Int. Conf. Automatic Face Gesture Recognit.*, Los Alamitos, CA, 1998, pp. 222–227.
- [16] Y. Li, J. Sun, C.-K. Tang, and H.-Y. Shum, "Lazy snapping," *ACM Trans. Graph.*, vol. 23, no. 3, pp. 303–308, Aug. 2004.
- [17] Mrs.B.A.Ahired, Prof.N.A.Dehpande,"A review on Video inpainting techniques", *International Journal of Computer Engineering & Technology(IJCET)*, Volume 4, Issue 1, January-February (2013), pp. 203-210.