Human Object Video in Painting using Shape Context Descriptor

K. Venkatesh Sharma,PhD
Professor, Dept of CSE
Sri Indu College of Engineering and Technology,Hyderabad

B. Rakesh, PhD
Lecturer, Engg Dept,
Al Musanna College of Technology, Oman

ABSTRACT

Video inpainting is a growing area of research. In this paper we are discussing a video inpainting technique which will be useful in cases where a set of frames are missing or damaged. This technique mainly involves three steps: creating new frames, graphical model construction, estimating the correct frame. Here we limit the distance of the motion trajectory where the motion is not continuous and hence we make sure that the motion continuity is efficient. Shape context descriptor and isomap algorithm is used for finding the matching frame and clustering of the frames in the graphical model.

Keywords

In painting, motion continuity, isomap, shape context.

1. INTRODUCTION

Image inpainting strongly corresponds to image interpolation. It has been widely used in many areas such as restoring old and damaged images, zooming and super-resolution, primal-sketch based perceptual image compression and coding, and the error concealment of image transmission, etc. Video inpainting also has become equally important with the increased. usage of videos in various applications. Most of the primitive video inpainting algorithms incorporated the image inpainting techniques directly. This had many problems since for video inpainting the motion continuity of the object also has to be considered. Also video inpainting involves considering large number of pixels and search spaces.

Video inpainting algorithms can be classified into texture synthesis based methods, patch based methods and object based methods depending upon the way in which the damaged parts are restored.

In this paper an object based video inpainting approach is given, which uses shape context descriptor to find the similar frames of a video. This includes the use of isomap based clustering to develop a graphical model representation of the video

2. RELATED WORK

In painting concept was first introduced by Bertalmio[1]. Though there are many efficient image inpainting algorithms available video inpainting is still a major area of research. There are a number of automatic video inpainting algorithms available. The conventional video inpainting algorithms generally fall under the patch based algorithms and template based algorithms.

Wexler[2] in his approach introduced video inpainting as a global optimization problem. Here frames which do not exist in the dataset are constructed from various space time patches selected from different parts of the video sequence. The global objective function removes local inconsistencies and heuristics of using large patches and it will rank the quality of completion. Fixed size cubes with three dimensions are used as the unit of the similarity measure function which is the Sum of Squared Differences. Though both spatial and temporal in formations are handled simultaneously the multi-scale nature of results will cause blurring and it is very slow.

A user-assisted video layer segmentation technique [3] that decomposes an input video into color and illumination videos was proposed by Jia. It uses a tensor voting technique to address the spatio–temporal issues in background and foreground. Image repairing is used for background inpainting, and occluded objects are reconstructed by synthesizing other available objects. However, a synthesized object created under this approach does not have a real trajectory. This approach is only suitable for periodic motion. In efficient object based video inpainting technique [4] proposed by Zhao video recorded by stationary cameras are considered. Pixels most compatible with the current frame is used for background inpainting and all available objects are used for foreground inpainting. A fixed sliding window includes a set of all continuous objects and defines a similarity function which measures the similarity between two continuous object templates. It can cover large holes and cases where the occluded objects are completely missing from frames. But insufficient number of postures can cause artifacts.

Cheung introduced video epitome [5] which contain the basic structure and motion characteristics of video which is useful for video inpainting. It is a probabilistic patch based model where patches are used for synthesizing images and videos. Patches from a part of the image is stitched together to synthesize new image with similar texture. It fills missing or occluded regions of the video data. Epitomes provide a representation which retains the natural flow of input data and it computationally and statistically advantageous over the patch libraries. It can be trained directly on corrupted or degraded data when the data is repetitive. Here the results are of low resolution and can contain over smoothing artifacts.

Another manifold based approach was proposed by Ding [6].It uses local linear embedding to map the image to a low-dimensional space. Here at first a set of descriptors with the necessary information to reconstruct the frame was found. Secondly, the optimum values for the descriptors were found and at last the frames were reconstructed based on the estimated values. Here Wexler's approach of finding the best matching patch from the adjacent frames was used for filling, difference being in the searching technique used. Rank Minimization Interpolation (RMI) was used to find the descriptors for the missing area thereby reducing the computational complexity. It is non-iterative and computationally attractive but cannot handle scaling information and causes blurring and ghost image artifacts where the object's motion is not periodic.

Lin [7] in his work virtual contour guided video object inpainting using posture mapping and retrieval proposed a technique which includes virtual contour construction, key posture selection and mapping and synthetic posture generation. It is based on the assumption that the trajectory of the occluded objects can be approximated by linear line segmentation during the period of occlusion. Mosaic based schemes and correspondence maps are used for background inpainting. It avoids the problem caused due to the insufficient number of postures. But the synthetic posture generation technique used here is not suitable for generating complex postures and it do not deal with illumination change problems

3. IMPLEMENTATION

Here we describe the proposed a video inpainting scheme where similar frames of missing frames are found inorder to restore any damaged or missing frames. Here we create new frames inorder to deal with the shortage of frames. This mainly involves three steps: 1) creating new frames 2) graphical model representation 3) estimating the correct frame. We discuss the steps in detail in the following sections.

3.1 Creating new frames:

Video inpainting is mainly required in cases where we may lose a set of frames or in cases where a set of frames may be damaged. Creating new frames will help us in cases where there is a shortage for frames. Here we combine together parts of different frames to obtain new frames which can be used to replace the damaged or missing frames [8]. Taking any two frames we align the two frames and then we take the difference between the two frames and project the difference onto the y-axis. Then, from the peaks and valleys of the projected distribution, it is possible to synthesize the new frame. It is not necessary that always the difference should be projected on the y-axis. If the object moves in another direction the difference can be projected along the x-axis also. This method is of low complexity and is more preferred

3.2 Graphical model construction:

Once all the frames are constructed in the next step we will plot the entire frame onto a feature space. This will give the graphical representation object's motion. Next the adjacent frames are linked together. The distribution can be better understood by calculating the distance between the postures in the feature space. The shape context descriptor [9] is used to get the better description about each object. The value of shape context descriptor is obtained for each frame and this will provide us with the details of similarity between two frames. In shape context descriptor we take a set of feature points on each shape known as feature points.

For each point p_i on the first shape we must find the best matching point $q_j$on the second shape. This can be considered as a correspondence problem. a set of vectors originating from a point to all the feature points in the selected shape is considered. This set of vectors is a rich description and as the number of vectors increases the representation of the shape becomes exact. In determining shape correspondences we must find points with similar descriptors. The shape contexts at two points are compared. Since taking all points makes the shape context very descriptive we use n-bin histograms. The costs of two sampled points of different frames are defined as given below

$$D(p_i, q_j) = \frac{1}{2}\sum_{k=1}^{N_{bin}} \frac{\left[hp_i(k) - hq_j(k)\right]^2}{hp_i(k) + hq_j(k)}.$$

where, h_{pi}(k) and h_{qj} (k) denote the kth bin of the two sampled points p_i and $q_j$respectively. Value of Nbin is set to be 60 for all sequences and r is the radius of the circle with N_{bin}. The best match between two different postures can be accomplished by minimizing the following total matching cost:

$$\mathrm{H}(\pi) = \sum_{\mathrm{j}} \mathrm{D}(\mathrm{p_j}, \mathrm{q}_{\pi(\mathrm{j}))}$$

where π is a permutation of 1,2,3…n. Because of the one-to-one matching requirement, shape matching can be considered as an assignment problem that can be solved by a bipartite graph matching method. The shape context distance between the frames can be computed as follows

$$(\mathrm{P}, \mathrm{Q}) = \frac{1}{\mathrm{N_p}} \sum_{\mathrm{i}} \mathrm{D}(\mathrm{p_i}, \mathrm{q}_{\pi(\mathrm{j})}) \frac{1}{\mathrm{N_Q}} \sum_{\mathrm{j}} \mathrm{D}(\mathrm{p_j}, \mathrm{q}_{\pi(\mathrm{j})})$$

Where N_p and N_q are the number of sample points on the shapes P and Q respectively.

After finding the similarity between frames using the shape context descriptor next the similar frames are clustered using a non-linear dimension reduction method known as isometric feature mapping or isomap. Here each frame is taken as the input feature point and the degree of similarity determines the distance between the points.

Step 1: Construct the neighborhood graph. A graph G is defined over all data points by connecting points i and j if they are closer than ϵ(ϵ-isomap) and if i is one of the k nearest neighbors of j (k-isomap).Set edge lengths equal to d_x (I,j)

Step 2: Compute the shortest paths in G. The Floyd's algorithm can be used for this step to find D_G.

Step 3: Construct d-dimensional embedding. Apply classical MDS on $D_{G\text{ to find}}$ the d-dimensional embedding and later find the Eigen vectors.

The advantages of using isomap algorithm include that it is non linear, non iterative polynomial time and it guarantees global optimality.

3.3 Stimating the correct frame:

Once the graphical model for the system is constructed we can estimate or reconstruct the. Correct frame by finding the difference between the graphical models constructed. The best matching frame will be the one with the most similar shape context description. Then the motion can be reconstructed which will be continuous and smooth.

Fig 1 below shows a set of frames from a video. Here a frame is missing in between the frames a and b. The frame which was missing has been reconstructed in Fig 2 as frame b

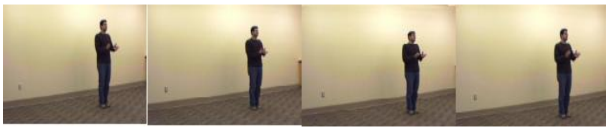

Fig 1. a b c d

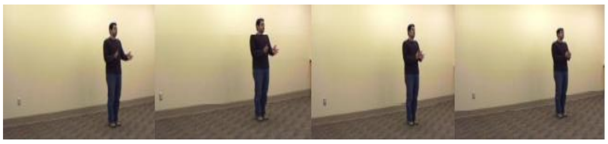

Fig 2.a b c d

4. CONCLUSION

The method discussed above provides better motion continuity for the video reconstructed when compared with the existing patch based video inpainting algorithms since here the new frame is reconstructed from the existing frames. Here since we are using the shape context descriptor the most similar frame is chosen. It has limitations since for isomap algorithm the dimensionality reduction perception is low

5. ACKNOWLEDGMENTS

Our thanks to the friends who have contributed towards development of the research paper.

6. REFERENCES

[1] K. A. Patwardhan, G. Sapiro, and M. Bertalmío, "Video inpainting under constrained camera motion," *IEEE Transactions on Image Processing*, vol.16, no. 2, pp. 545–553, February 2007.

[2] Y. Wexler, E. Shechtman, and M. Irani, "Space-time completion of video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 3, pp. 463– 476, March 2007.

[3] J. Jia, Y. Tai, T. Wu, and C. Tang, "Video repairing under variable illumination using cyclic motions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 5, pp. 832–883, May 2006.

[4] S.-C. S. Cheung, J. Zhao, and M. V. Venkatesh, "Efficient object-based video inpainting," *Proceedings of IEEE Conference on Image Processing*, pp. 705–708,Oct. 2006.

[5] V. Cheung, B. J. Frey, and N. Jojic, "Video epitomes," *of IEEE Conference on Computer Vision and Pattern Recognition*, pp.42–49, June 2005.

[6] T. Ding, M. Sznaier, and O. I. Camps, "A rank minimization approach to video inpainting," Proceedings of IEEE Conference on Computer Vision, pp. 1–8, October 2007.

[7] C.-H. Ling, C.-W. Lin, C.-W. Su, Y.-S. Chen, and H.-Y. M. Liao, "Virtual contour- guided video object inpainting using posture mapping and retrieval," *IEEE Transactions on Multimedia*, vol. 13, no. 2, pp. 292–302, Apr.2011.

[8] Chih-Hung Ling,Yu-Ming Liang,Chia-Wen Lin,Yong-Sheng Chen,Hong-Yuan Mark Liao," Human object inpainting using manifold learning-based posture sequence estimation", *IEEE Transactions on Image Processing*, vol.20,no.11,pp. 3124 - 3135 , 2011.

[9] S. Belongie, J. Malik, and J. Puzicha, Shape matching and objectrecognition using shape contexts," *IEEE Trans. Pattern Anal. Mach.Intell.*, vol. 24, no. 4, pp. 509–522, Apr. 2002.