ISSN: 2277-9655
(I2OR), Publication Impact Factor: 3.785
(ISRA), Impact Factor: 2.114

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

VIDEO INPAINTING TO REMOVE OBJECTS WITH PIXMIX APPROACH

Tushar A. Pimple*, Prof. Dr. S.K.Shah

* PG student, Smt. Kashibai Navale College of Engineering, Vadgaon (Bk) Pune, India

ABSTRACT

Video inpainting is the process of removing a portion of a video and filling in the missing part by using neighboring frames. Existing approaches for video inpainting do not achieve high quality coherent video stream as they are highly computational expensive. PixMix method is based on a combined pixel based approach and this allows for even faster inpainting. A new object tracking and frame to frame coherence approach for object removal is used, this provides high quality inpainting. Video inpainting can be used to repair damaged footage. It also finds applications in multimedia editing and video modification for privacy protection.

KEYWORDS: Video inpainting, PixMix, Object tracking, Object Removal.

INTRODUCTION

In recent years, transforming photographs and vintage films/videos into digital format has gain an attention in the field of multimedia editing. Video repairing techniques widely used to restore the visual content of vintage films include video de-noising, its stabilization and video inpainting. Video inpainting, is one of the most challenging technique, which helps users to remove undesirable object.Many video inpainting techniques restore the holes in images by propagating linear structures into the target region via diffusion which is inspired by the partial differential equations of physical heat flow. One of the drawback of these techniques is that the diffusion process introduces some blurring, which becomes noticeable when filling larger regions. Exemplar-based image inpainting is introduced to overcome these drawbacks and can produce a reasonably good quality of output for larger regions on still images. Similar techniques can be adopted to remove an object from a video sequence by combining with an object tracking mechanism to fit the need of video inpainting.

This paper introduces pixel based approach for video inpainting. Methods described previously to remove object were patch based. However, in contrast to those purely patch-based approaches, new approach PixMix is based on a combined pixel-based approach and also allows for even faster inpainting while improving the overall image quality. The combined with new tracking approach and frame-to-frame coherence this provides the basis for object removal. By additionally applying a homograph based approach a moving object can be removed from a video which is the future scope of method.

VIDEO INPAINTING

Simultaneous Structure and Texture Video Inpainting

Simultaneous structure and texture video inpainting introduced a digital video inpainting algorithm based on a partial differential equation (PDE) model. If defected region is only small set of pixel, diffusion (solving partial differential equation) works very well to remove damaged or texture part of video. The algorithm treats the input video frame as three separate channels as R, G and B. For each channel, it fills in the areas to be inpainted by propagating information from the outside of the masked region along level lines (isophotes). Isophotes directions are obtained by computing at each pixel along the contour a discretized gradient vector (this gives the direction of largest spatial change) and by rotating the resulting vector by 90 degrees. This intends to propagate information while preserving edges. A 2-D Laplacian is used to locally estimate the variation in smoothness and such variation is propagated along the isophote direction, after every few step of the inpainting process, the algorithm runs a few diffusion iterations to smooth the region which is inpainted. Anisotropic diffusion is used in order to preserve edges across the inpainted region.

ISSN: 2277-9655
(I2OR), Publication Impact Factor: 3.785
(ISRA), Impact Factor: 2.114

Simultaneous Structure and Texture Video Inpainting method is as following

1. Decompose the input video frame into two sub-frames: U, the structure frame, and V , the texture frame
2. Fill in U using frame a video inpainting algorithm
3. Fill in V using a texture synthesis algorithm
4. Recombine the reconstructed U and V frame to form output frame.

Simultaneous structure and texture video inpainting utilized the partial differential equations (PDEs) for video inpainting, but it is only fitted for low-resolution videos with light scratches or little areas. The drawbacks of PDEs method for replacing larger regions or high-resolution videos are lack of consideration for the priorities of the inpainting block sequences and the extension of video frame textures. Thus, the quality of the video inpainting results will decrease visibly.

Video Restoration using Multiresolution wavelet transform and Video Inpainting

In order to effectively retain the image data, various researchers have continually proposed various methods of video inpainting. These video inpainting methods can be divided into two forms of analysis, which can be viewed from two different perspectives: texture analysis and color analysis. In the texture analysis, the video inpainting technique considers spatial texture directly up to the related position used. Conversely, in the color analysis, the color compositions of the original video are first converted into various domains through different color system transformations, and then depending on the diverse color composition trend analysis, the color components of damaged regions are repaired separately. However, the above mentioned methods are unable to combine their respective advantages in the area of video inpainting in different analysis domains. The discrete wavelet transform (DWT) can be used to resolve Y composition (texture) video frame into multiple layers so as to make the spatial frequency analysis possible. The wavelet coefficients of the converted textural video frame include simultaneous spatial-frequency relativity and produce multi-resolution layers with different frequency characteristics. By recognizing the concept of multi-resolution video inpainting a proper video inpainting procedure shall be sequentially started from the lower layer to the higher layer. In Addition, the color components of the image frame (Cb and Cr) serve as a supplementary reference to support the linear interpolation method applied during damaged data prediction.

Wavelet Transform

It provides the time-frequency representation. Wavelet transform is capable of providing the time and frequency information simultaneously; thus giving a time-frequency representation of the signal. Wavelet passes the time-domain signal from various high-pass and low pass filters and filters out either high frequency or low frequency portions of the signal. This procedure is repeated every time, some portion of the signal corresponding to some frequencies being removed from the signal. Suppose a signal which has frequencies up to 1000 Hz. In the first stage one can split up the signal in to two parts by passing the signal from a high-pass and a low-pass filter which results in two different versions of the same signal: portion of the signal corresponding to 0-500 Hz (low pass portion), and 500-1000 Hz (high pass portion). Then, take either portion usually low pass portion or both and do the same thing again. This operation is called decomposition.

Wavelet transform doesn't tell what spectral component exists at any given time instant. But it will investigate what spectral components exist at any given interval of time. High frequencies are better resolved in time and low frequencies are better resolved in frequency. This means that a certain high frequency component can be located better in time (with less relative error) than a low frequency component. On the contrary a low frequency component can be located better in frequency compared to high frequency component.

Rank Minimization Method

In this section basic ideas that allow for recasting the inpainting problem into a rank–minimization one are outlined. As indicated the main idea of the proposed approach is to, rather than directly attempt to interpolate missing pixels, estimate, based on all available spatiotemporal information, the value of a set of descriptors that encapsulate the information necessary to reconstruct missing/corrupted frames. In order to estimate the values of missing descriptors, collect the values of all the descriptors corresponding to the kth frame in a vector $f_k = [f_{k1}, \cdots, f_{kp}]$, where p is the descriptor amount, and assume that these values of each descriptor are generated by a stationary Gauss–Markov random process. This is equivalent to assuming that for *ith* descriptor f_k is related to its values in previous frames by an ARMAX model of the form, where f(k) is instead of f_k for a explicit expression.

ISSN: 2277-9655
(I2OR), Publication Impact Factor: 3.785
(ISRA), Impact Factor: 2.114

Model image sequences of temporal textures using the spatio-temporal autoregressive model (STAR), which expresses each pixel as a linear combination of surrounding pixels lagged both in space and in time. In this method, video data as linear combination of other frames is not represented. Rather, representing the present value of a descriptor (which may be related in a nonlinear way to the values of the pixels) as a linear combination of its past values. This is always possible as stated in the paper. Note that in this context, if the coefficients of descriptor model are known, then the inpainting problem can be trivially solved by using the model, along with the available measurements of f, to reconstruct the missing data. In principle one could try to use a two–tiered approach, where the model that best explains the available data is first extracted from the uncorrupted frames and then used to inpaint the missing values. However, finding an explicit model is unnecessary: missing values of each descriptor in vector f can be directly found by solving a rank–minimization problem, obviating the need for finding an explicit model. All missing values will be calculated through p procedures repeatedly. As illustrated with several examples, this observation leads to simple, computationally tractable inpainting algorithms.

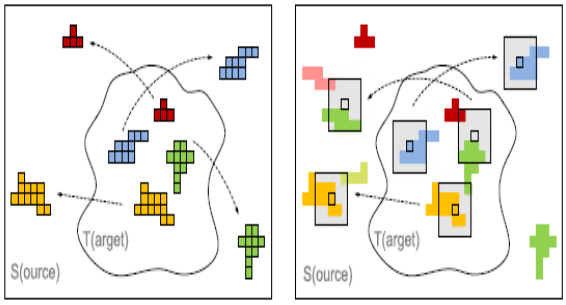
VIDEO INPAINTING TO REMOVE OBJECTS WITH PIXMIX APPROACH

Mapping Function

Image inpainting can be defined as a global minimization problem of finding the transformation function f: T → S producing minimal overall synthesis costs for an arbitrary image I according to a given cost function and the image I is subdivided into the two distinct sets T and S with I = T U S, $T \cap S = \emptyset$ and $S \neq \emptyset$. All pixels from T (target) are to be mapped by pixels defined in S (source). Here, f decides a mapping between target and source pixels inside a frame which is to be repaired or inpainted. Once f has been calculated, the final image can be produced by replacing all target pixels with source information pixels. Usually, the result of manipulated frames may be considered as acceptable, if the replaced (synthesized) frame content blends in seamlessly with the surrounding image information while it remains free of disturbing artifacts and implausible blurring effects.

Further, the new frame information should visually fit to the remaining parts of the video. The transformation function f is based on the following two constraints:
[1] Neighboring pixels defined in T should be mapped to equivalent neighboring pixels in S. This first constraint ensures the structural and spatial preservation of image information (Fig. 1a).

[2] The neighborhood appearance of pixels in T should be similar to the neighborhood appearance of their mapped equivalents in S. Hence a visually coherent result and seamless transitions at the border of the synthesized area are ensured (see Fig. 1b).



(a) Spatial cost ***(b) Appearance cost***

Fig 1 The two cost constraints of the transformation function.

The solution for global minimization problem is to find a transformation function f producing the minimal overall cost for an image I and a target region $T \subset I$ *with* $S = I/T$

$$\min_{f} \sum_{p \in T} cost_{\alpha}(p) \qquad (1)$$

ISSN: 2277-9655
(I2OR), Publication Impact Factor: 3.785
(ISRA), Impact Factor: 2.114

While $p = (p_x, p_y)^T$ is a 2D position and the cost function $cost_\alpha$:T→Q is determined for all elements inside the target region.

3.2 Cost Function

Here in this approach the overall costs are subdivided into a part based on the spatial impact and a part based on the impact of appearance and this can be represented by the following linear combination:

$$cost_\alpha(p) = \alpha . cost_{spatial}(p) + (1-\alpha) . cost_{apperance}(p) \quad (2)$$

while the control parameter allows a balancing between the spatial impact and appearance impact. Here minimization of spatial cost impact is observed, which forces a mapping of neighboring target pixels to neighboring mapping pixels. Also this is represented for an arbitrary neighborhood Ns by costspatial:T→Q:

$$cost_{spatial}(p) = \sum_{\vec{v} \in N_S} d_s[f(p) + \vec{v}, f(p + \vec{v})] . w_s(\vec{v}) \quad (3)$$

Ideally, any neighbor $\vec{v} \in N_s$ of p is mapped to the corresponding neighbor $\vec{v} \in N_s$ of f(p). The spatial cost sums up the spatial distances $d_s(.)$ from this ideal situation for any $\vec{v} \in N_s$ and $p \in T$ (Figs. 1a and 2). Thus, PixMix approach differs from patch-based approach.Spatial cost function allows for a significant faster convergence while reducing image blurring and geometrical artifacts, this novel cost constraint can be seen as an elastic spring optimization automatically minimizing neighboring mapping offsets.

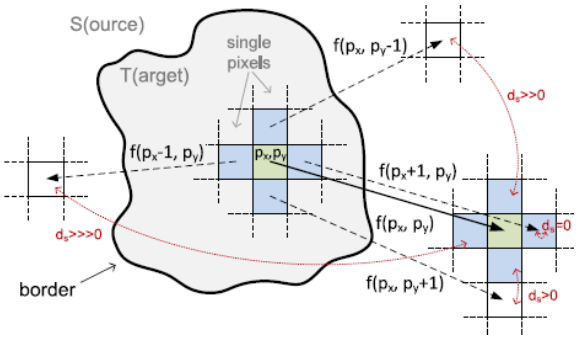


Fig. 2:Spatial cost calculated by neighboring mapping depicted for a four neighborhood.

The application of an appearance cost can be found in several related works. Appearance costs have higher impact to the overall cost to neglect unwanted border effects such as borders or visual discontinuities. Circular neighborhood sets required more processing time while providing only a small improvement in visual. A patch size of 5 × 5 pixels proved to provide enough details regarding the visual content allowing for fast computation. In PixMix approach the sum of squared differences (SSD) is used for the appearance distance da as it provided a good performance and quality when compared to other measures such as the sum of absolute differences (SAD) and the zero-mean SSD.

Iterative Refinement and Propagation

Finding the optimal transformation function f is realized by starting with a rather rough guess of f followed by a series of iterative refinements steps. At each iteration, the mapping for each target pixel is sought to be improved. Randomly individual source positions are tested according to the local cost function and accepted whenever the local cost can be reduced. Here each refinement needs a target information update of an entire image patch, requiring the application of the individual contribution from each patch followed by normalization. PixMix approach directly updates only a single pixel and thus avoids expensive normalizations.

ISSN: 2277-9655
(I2OR), Publication Impact Factor: 3.785
(ISRA), Impact Factor: 2.114

For inpainting purpose multiresolution technique is used. A small change for improvement is applied on an image pyramid starting with a reduced resolution layer and increasing the image size until the original resolution has been reached. Depending on the mask size and frame dimension typically between three and eight layers are used. The coarsest pyramid layer is found by the first layer in that no mask pixel exists that has a larger distance than three pixels to the inpainting border. The algorithm starts with an initial mapping guess $\hat{f}_{n-1}$ in the coarsest pyramid layer Ln-1 and stops if an improved mapping fn-1 with minimal overall cost has been determined. This mapping is then forwarded to the next pyramid layer Ln-2 and is used as the new initialization $\hat{f}_{n-2}$. Again, after a series of iterations within the current layer the optimized transformation fn-2 is forwarded as the initialization of the next layer until the final layer L0 (providing the highest resolution) has been reached and processed (Fig. 3).

The applied image pyramid allows the covering of visual structures with individual frequency speeds up the mapping convergence and thus reduces the chance that the algorithm gets trapped by some local minima. Information propagation improves the overall inpainting significantly. However, instead of propagating the position of entire image patches, PixMix approach forwards single pixel mapping positions only.

Iterative cost refinement is applied on disjoint subsets T_0, T_1……T_{n-1}. Thus each subset T_i can be processed by an individual thread in parallel. In earlier work, static frame subsets have been applied restricting propagation of mapping information to be within individual subsets. This refinement can cause unwanted synthesis blocks in the final video as the mapping exchange between subsets is restricted to the subset border. Further damped propagation may reduce synthesis performance. PixMix approach applies random subsets changing between forward and backward propagation while the subsets' size stays constant. These start rows are optimized explicitly before any refinement iteration processed so because of this, neighboring subsets have access to the same mapping information. The successive refinements toggle between forward and backward propagation, have changing target subsets, and start from common (already refined) rows. A random subsets have a significant impact on the final image quality and convergence performance as information propagation is applied to the entire synthesis mask rather than limited to the sub-blocks.

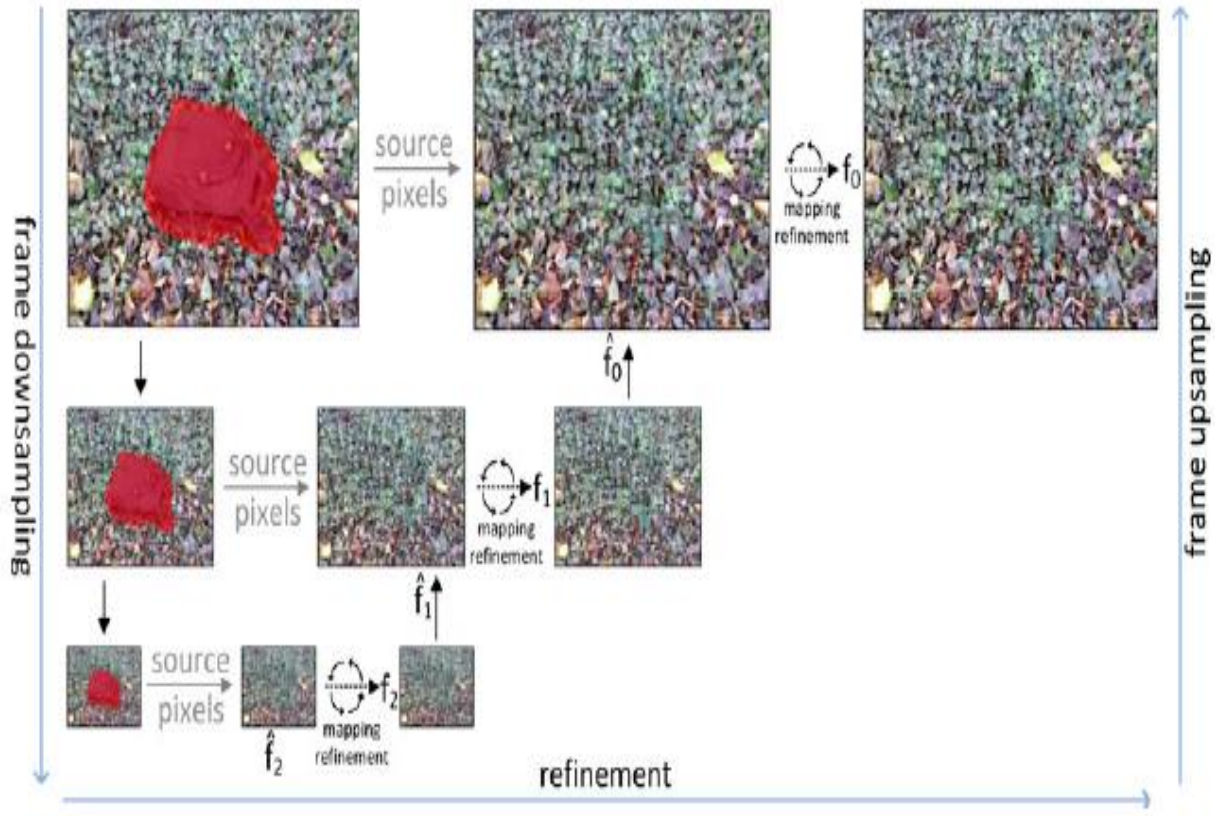


Fig 3: Scheme of pyramid refinement: The original frame is downsampled and iteratively refine and upsampled

ISSN: 2277-9655
(I2OR), Publication Impact Factor: 3.785
(ISRA), Impact Factor: 2.114

Constraints

More advanced constraints may be used to guide the inpainting algorithm providing improved visual results. It depends on the structure of the background or the desired and undesired visual elements; the final inpainting results may be optimized according to the expectations of the users. In a real time video inpainting user defined constraints cannot easily applied, as it requires a certain amount of time. In real time application user is not willing to spend time on defining constraints. As user defined constraint has certain limitations it cannot be used for every video inpainting. Further, when inpainting is applied to video streams not requiring real-time performance or for slightly time shifted live broadcasts, even individual user-defined constraints may be used. PixMix approach allows for the user defined constraints in certain conditions, in the result section video inpainting with and without constraints is shown.

Area Constraints

The most obvious form of inpainting constraints guide the algorithm to explicitly use or avoid image regions from the remaining image content. Algorithm is forced to use image content explicitly selected by the user and discarding content the user does not prefer. An inverse importance map $\bar{m}: S \rightarrow Q$ over all elements (pixels) in S has to be defined to individually rate visual importance of image content. A map with a more detailed granularity clearly allows for more precise algorithm guidance. The final area constraint $constr_A : T \times (T \rightarrow S) \rightarrow Q$ is then directly given by the inverse importance map $\bar{m}$:

$$constr_A(P, f) = \bar{m}[f(p)]$$

RESULTS AND DISCUSSION

Frame Extraction

In this paper, implementation and testing of video inpaintig with PixMix is done. It first separates video frames from a video. These frames afterward help for video repairing purpose. Here MATLAB 10 version and a Computer with core i3 processor is used. Here a video of beach is separated into the frames for the purpose of the video inpainting as shown in fig 4

Fig 4: Extracted frame from a video sequence.

Object selection using user define constraint

As discussed in the section 3.4 here results are obtained by using pre defined constrained. Here object is considered as the umbrella as which is seen in the extracted frame as shown in the fig 4. And umbrella is defined as a constrained as shown in fig 5.Visual quality of the video inpainting with constraint is superior as compared to other inpainting method here object is selected only once and then object is searched in each frame. The constrained video inpainting provides a reconstructed edge with sub pixel accuracy while the remaining undesired image content is synthesized with visual content still matching with the surrounding environment.

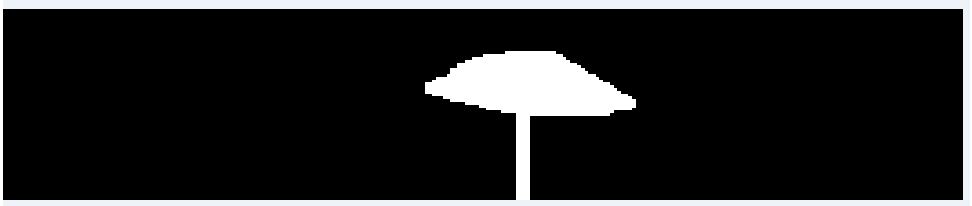


Fig. 5: Selected umbrella for inpainting

ISSN: 2277-9655
(I2OR), Publication Impact Factor: 3.785
(ISRA), Impact Factor: 2.114

Inpainting by PixMix

After objection selection inpaintinng is done with PixMix approach. Here object is selected only in one frame and after that in each frame object is searched for inpainting purpose. As here object is searched in each frame it takes certain time. And after that inpainted frames are obtained which do not contain constrained object,as here umbrella is taken as constrained object frames without umbrella are obtained, in fig 6. And after that by using inpainted frame a final inpainted video is obtained which do not contain umbrella shown in fig 6.

Fig.6: Inpainted video frame without umbrella

CONCLUSION

In this paper, we presented PixMix, this pixel based approach to image and video inpainting. Additionally, capable object selection and tracking algorithm has been introduced. Inpainting approach allows for balancing between the spatial and the appearance term of the cost function in order to provide optimal inpainting results. These overall results showed fewer artifacts than other approaches allowing for high-quality image inpainting. This provided the basis for our self-contained high-quality video inpainting approach. This is achieved by extending the overall cost function by a frame-to-frame coherence term and by applying a homography as a first guess for the mapping in the next frame providing a significantly better initialization. Video inpainting approach is based on the previous and the current frame only and allowing for a high-quality manipulation of live video streams. In future work, we are planning to extend the homography based approach to arbitrary 3D objects. Further intended to publish the results of a user study on the perceived quality and plausabilof our video inpainting approach, the actual user study has already been conducted and requires further analysis of the data obtained. Finally the investigation for smart temporal mapping approaches, providing more sophisticated results for moving objects required, while still allowing for video manipulations.

REFERENCES

1. Criminisi, P. Perez, and K. Toyama, "Region Filling and Object Removal by Exemplar-Based Image Inpainting," *IEEE Trans. Image Processing, vol. 13, no. 9*, pp. 1200-1212, Sept. 2004.
2. I. Drori, D. Cohen-Or, and H. Yeshurun, "Fragment-Based Image Completion," *Proc. ACM SIGGRAPH*, pp. 303-312, 2003.
3. J. Sun, L. Yuan, J. Jia, and H.-Y. Shum, "Image Completion with Structure Propagation," *Proc. ACM SIGGRAPH*, pp. 861-868,2005
4. D. Simakov, Y. Caspi, E. Shechtman, and M. Irani, "Summarizing Visual Data Using Bidirectional Similarity," *Proc. IEEE Conf. Computer Vision and Pattern Recognition* (CVPR '08), 2008.
5. A. Bugeau, M. Bertalm_ıo, V. Caselles, and G. Sapiro, "A Comprehensive Framework for Image Inpainting," *IEEE Trans. Image Processing, vol. 19*, no. 10, pp. 2634-2645, Oct. 2010.
6. J. Jia, T.-P. Wu, Y.-W. Tai, and C.-K. Tang, "Video Repairing: Inference of Foreground and Background under Severe Occlusion," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR '04)*, vol. 1, pp. 364-371, June 2004.
7. J. Jia, Y.-W. Tai, T.-P. Wu, and C.-K. Tang, "Video Repairing under Variable Illumination Using Cyclic Motions," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 28, no. 5, pp. 832-839, May 2006.
8. K.A. Patwardhan, G. Sapiro, and M. Bertalmio, "Video Inpainting under Constrained Camera Motion," *IEEE Trans. Image Processing*, vol. 16, no. 2, pp. 545-553, Feb. 2007.

ISSN: 2277-9655
(I2OR), Publication Impact Factor: 3.785
(ISRA), Impact Factor: 2.114

9. Y. Shen, F. Lu, X. Cao, and H. Foroosh, "Video Completion for Perspective Camera under Constrained Motion," *Proc. 18th Int'l Conf. Pattern Recognition (ICPR '06))*, vol. 3, pp. 63-66, June 2006.
10. T. Shiratori, Y. Matsushita, X. Tang, and S.B. Kang, "Video Completion by Motion Field Transfer," *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR '06)*, vol. 1, pp. 411-418, June 2006.
11. M.V. Venkatesh, S.-C.S. Cheung, and J. Zhao, "Efficient Object- Based Video Inpainting," *Pattern Recognition Letters*, vol. 30, no. 2, pp. 168-179, Jan. 2009.
12. Y. Zhang, J. Xiao, and M. Shah, "Motion Layer Based Object Removal in Videos," *Proc. Seventh IEEE Workshops Application of Computer Vision, (WACV/MOTIONS '05*), vol. 1, pp. 516-521, Jan. 2005.
13. Y. Wexler, E. Shechtman, and M. Irani, "Space-Time Completion of Video," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 29, no. 3, pp. 463-476, Mar. 2007.
14. O. Korkalo, M. Aittala, and S. Siltanen, "Light-Weight Marker Hiding for Augmented Reality," *Proc. IEEE Ninth Int'l Symp. Mixed and Augmented Reality (ISMAR '10)*, pp. 247-248, Oct. 2010.
15. N. Kawai, M. Yamasaki, T. Sato, and N. Yokoya, "AR Marker Hiding Based on Image Inpainting and Reflection of Illumination Changes," Proc. *IEEE Int'l Symp. Mixed and Augmented Reality (ISMAR '12)*, pp. 293-294, Nov. 2012.