ICASSP 2023 · IEEE

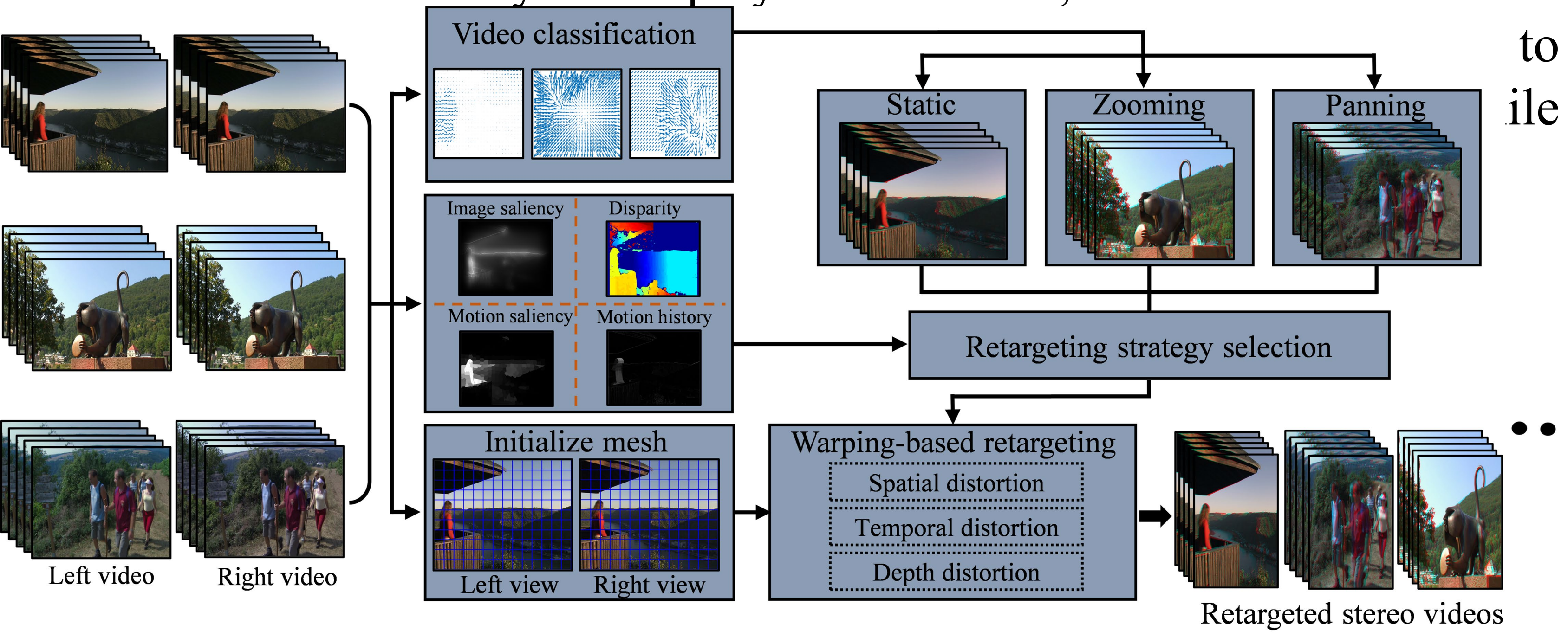
Stereoscopic Video Retargeting Based on Camera Motion Classification

Linghui Cai, Zhenhua Tang

Guangxi University

Introduction

The maturity of stereoscopic display technology and the popularity of media terminals increase the demand for viewing stereo videos. Due to the diversity of display screen sizes, the content of videos to ile

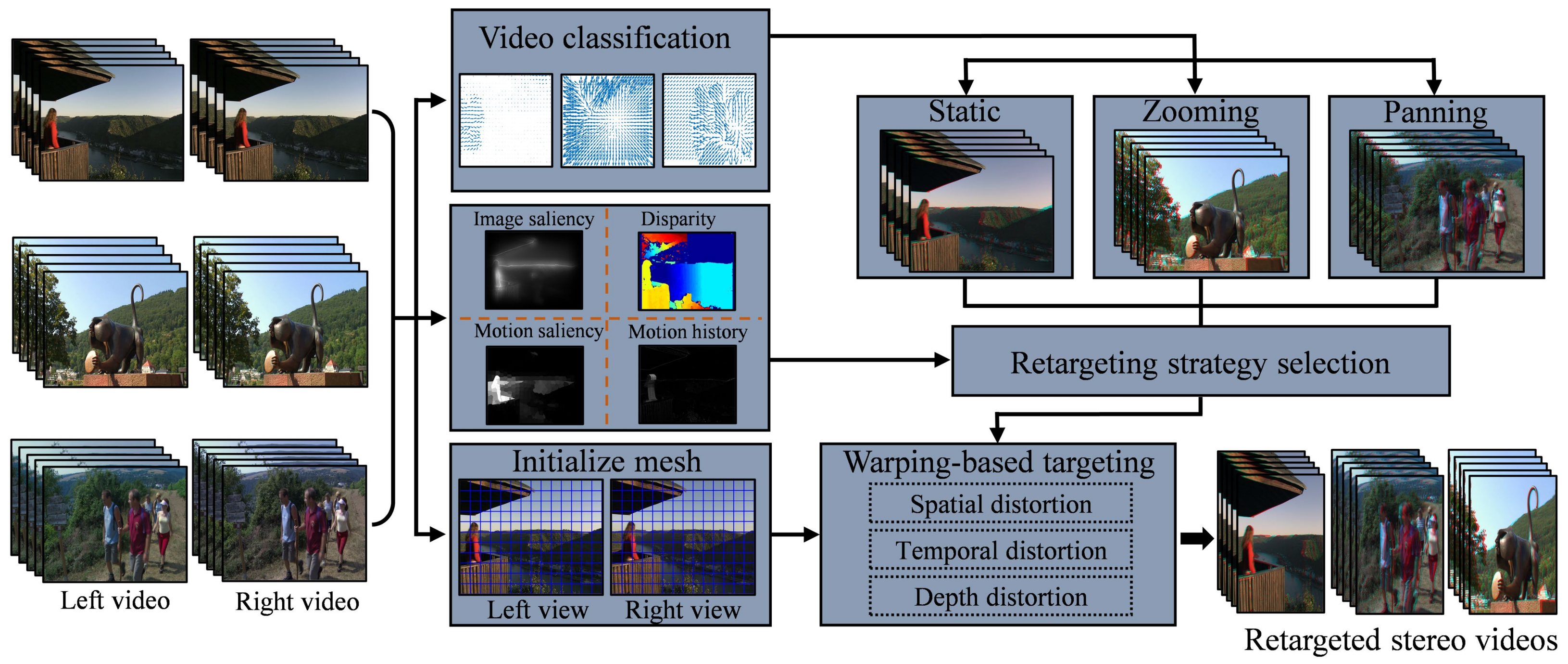


However, existing methods ignore the fact that various features of different stereo videos may have effect on the retargeting quality. They perform resizing with the same methodology, leading to the low quality for retargeted videos.

The Proposed Method

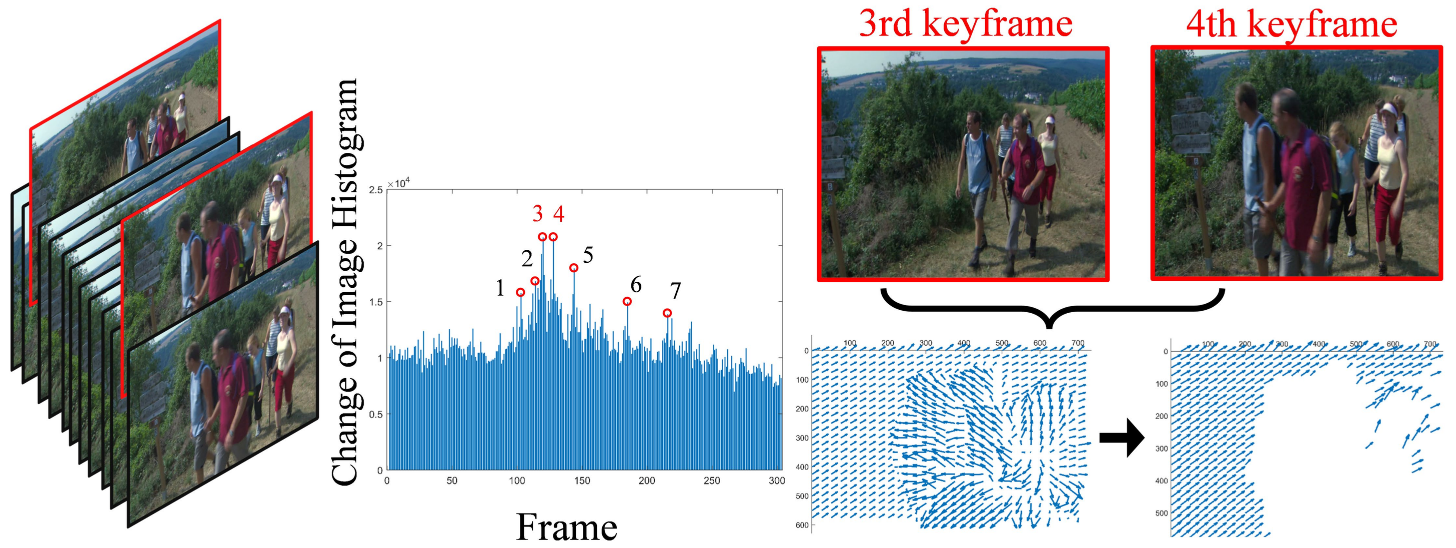
• ***Framework***

The proposed system consists of three modules: **video classification**, **retargeting strategy selection** and **warping-based retargeting**.

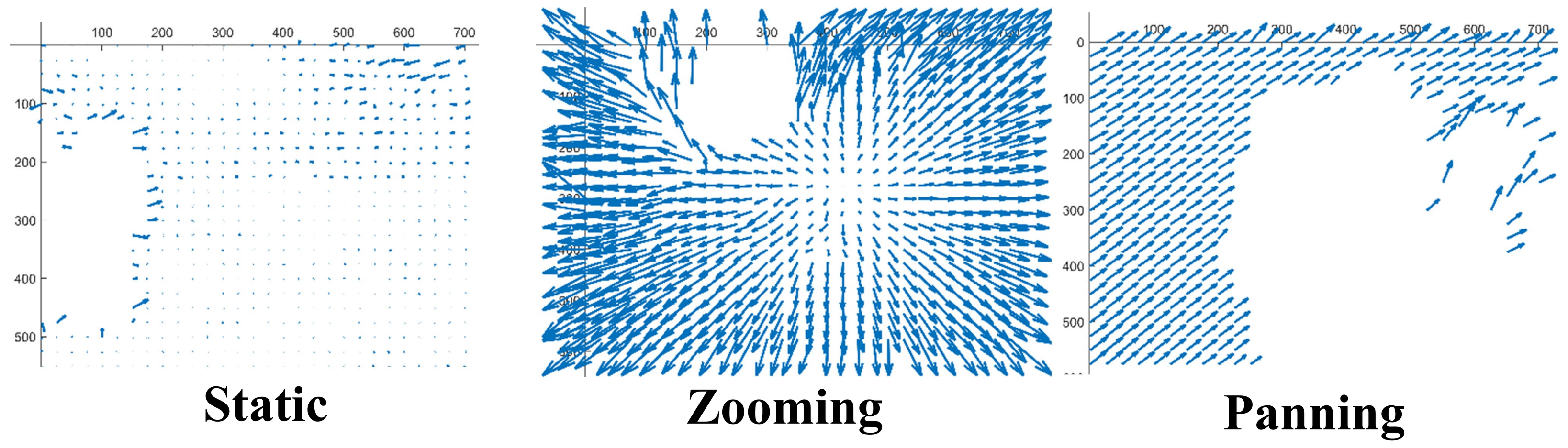


• ***Video Classification***

We extract the global motion vectors of adjacent keyframes by optical flow estimation and remove motion the vectors of foreground.

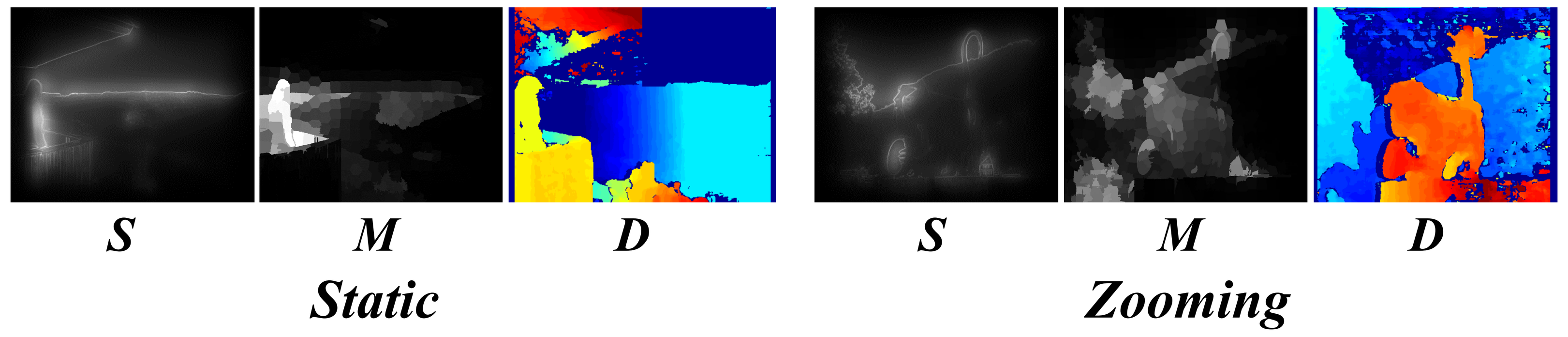


The motion vectors of the background are used to determine the mode of camera motion, including **static**, **zooming** and **panning**.



• ***Retargeting Strategy Selection***

According to the camera motion, we combine different saliency maps to guide optimization in the warping model, including image saliency ***S***, motion saliency ***M***, disparity ***D***, and motion history ***Mh***.



$$I_t^{static}(i,j)=\omega_s S_{avg}(i,j)+\omega_m M_{total}(i,j)+\omega_d D_{avg}(i,j)$$

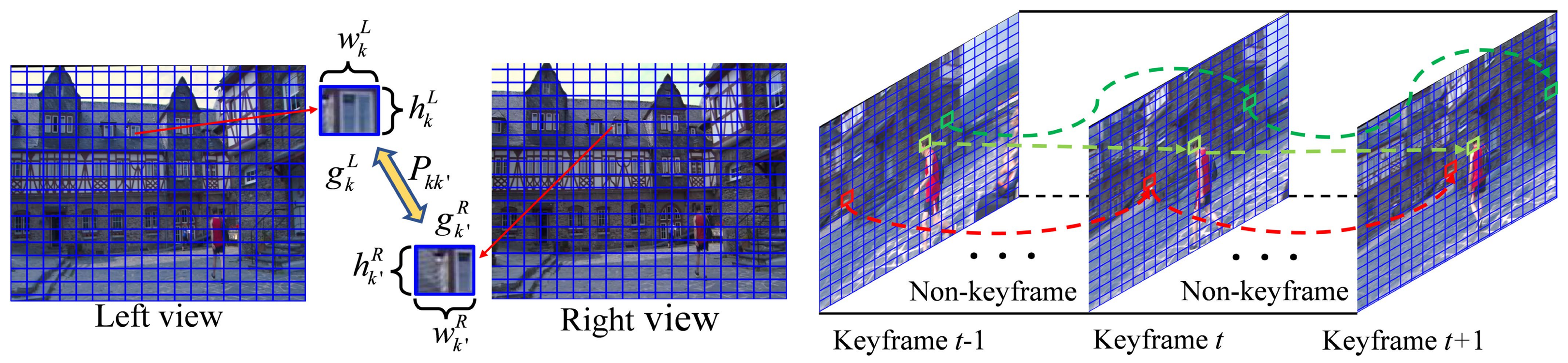
$$I_t^{zoom}(i,j)=\Gamma(i,j)\left(\omega_s S_{avg}(i,j)+\omega_m M_{avg}(i,j)+\omega_d D_{avg}(i,j)\right)$$

$$Mh_t(i,j)=\sum_{l=t-\mu}^{t+\mu}\max\left(R_t(i,j),e^{-\frac{|l-t|}{\mu}}R_l(H_{t\to l})\right)$$

$$I_t^{pan}(i,j)=\omega_s S_t(i,j)+\omega_m M_t(i,j)+\omega_d D_t(i,j)+Mh_t(i,j)$$

• ***Warping-based Retargeting***

We establish a total energy function, concerning on spatial, temporal, and depth distortion measurement. After minimizing the energy loss function and grids interpolation, we can obtain the retargeted grids.



$$\begin{cases} E_S=\sum_{v\in\{L,R\}}\sum_k \left\| w_k^{v,t}\cdot \hat{h}_k^{v,t}-h_k^{v,t}\cdot \hat{w}_k^{v,t}\right\|^2\cdot\sigma_k^{z,t} \\ E_T=\sum_{v\in\{L,R\}}\sum_k\left(\left\|\hat{w}_k^{v,t}-\hat{w}_k^{v,t-1}\right\|^2+\left\|\hat{h}_k^{v,t}-\hat{h}_k^{v,t-1}\right\|^2\right), \quad t\in\mathbb{F} \\ E_D=\sum_k\sum_{k'\in[k-m,k+n]}P_{kk'}^t\left\|\hat{w}_k^{L,t}-\hat{w}_k^{R,t}-(w_k^{L,t}-w_k^{R,t})\right\|^2 \end{cases}$$

$$\arg\min_{\{\hat{w}_k^{L,t},\hat{w}_k^{R,t},\hat{h}_k^{L,t},\hat{h}_k^{R,t}\}}\left(E=\lambda_S E_S+\lambda_T E_T+\lambda_D E_D\right)$$

Experiment Results

• ***Qualitative Evaluation***

We compare the performance with two relative retargeting methods: **WS3D-VidR** (Islam er al. 2019) and **PTIG-svr** (Li et al. 2020). Our method can preserve the spatial shape of salient objects better than the other two methods, and also maintains a good stereoscopic visual experience.

Original video **WS3D-VidR** **PTIG-svr** **Ours**

• ***Quantitative Evaluation***

We conduct objective evaluation for the results using three metrics: saliency similarity measurement (***SSM***), temporal inconsistency distortion (***TID***), and disparity distortion ratio (***DDr***). Our method performs better than other algorithms in terms of ***SSM***. For the ***TID*** and ***DDr*** metrics, we can achieve good performance on most videos.

Dataset	SSM			TID			DDr		
	WS3D-VidR	PTIG-svr	Ours	WS3D-VidR	PTIG-svr	Ours	WS3D-VidR	PTIG-svr	Ours
CVW	0.6050	0.5196	**0.4612**	0.6324	0.6872	**0.6159**	5.82%	4.79%	4.98%
The Eye	0.5246	0.4896	**0.4618**	0.4954	0.7572	**0.4862**	6.95%	4.71%	**4.38%**
Rhine Valley	0.5956	0.5064	**0.4564**	0.5095	0.5591	**0.4896**	6.23%	5.82%	**5.79%**
Treffen	0.4532	04218	**0.3938**	0.4330	0.5192	0.4522	8.59%	7.55%	7.89%
Macroshow	0.3775	0.3294	**0.3092**	0.3874	0.4221	**0.3679**	8.06%	5.23%	5.62%
Skydiving	0.6979	0.5991	**0.5510**	0.4338	0.5593	0.4649	9.59%	6.25%	**6.04%**
Heidelberg	0.4906	0.4308	**0.4009**	0.5795	0.8125	**0.5561**	4.95%	4.26%	**4.19%**

Conclusion

We propose a stereoscopic video retargeting method based on camera motion classification, which adopts different video retargeting strategies for various videos. Quantitative Evaluation show that the performance of the proposed method is superior to the existing methods.