VRMixer: Mixing Video and Real World with Video Segmentation

Tatsunori Hirai
Waseda University
Shinjuku, Tokyo, Japan
tatsunori_hirai@asagi.waseda.jp

Satoshi Nakamura
Meiji University / CREST, JST
Nakano, Tokyo, Japan
satoshi@snakamura.org

Tsubasa Yumura
Meiji University / CREST, JST
Nakano, Tokyo, Japan
yumu.stp@gmail.com

Shigeo Morishima
Waseda University/ CREST,
JST
Shinjuku, Tokyo, Japan
shigeo@waseda.jp

ABSTRACT

This paper presents *VRMixer*, a system that mixes real world and a video clip letting a user enter the video clip and realize a virtual co-starring role with people appearing in the clip. Our system constructs a simple virtual space by allocating video frames and the people appearing in the clip within the user's 3D space. By measuring the user's 3D depth in real time, the time space of the video clip and the user's 3D space become mixed. *VRMixer* automatically extracts human images from a video clip by using a video segmentation technique based on 3D graph cut segmentation that employs face detection to detach the human area from the background. A virtual 3D space (i.e., 2.5D space) is constructed by positioning the background in the back and the people in the front. In the video clip, the user can stand in front of or behind the people by using a depth camera. Real objects that are closer than the distance of the clip's background will become part of the constructed virtual 3D space. This synthesis creates a new image in which the user appears to be a part of the video clip, or in which people in the clip appear to enter the real world. We aim to realize "video reality," i.e., a mixture of reality and video clips using *VRMixer*.

Author Keywords

2.5D; Mixed media; Human extraction; 3D graph cut segmentation; Visual interaction.

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

INTRODUCTION

In the present human life, people desire extraordinary experiences. For many people, going out to theme parks and having

an exciting time is a great relaxation method and one of the best ways to disentangle from day-to-day life. People also watch movies and enjoy fantasy worlds that we never expect to reach.

Science and technology can provide such experiences. The representative research area is virtual reality (VR), which allows people to experience other worlds as if they were actually there. However, in return of an immersive feeling, many VR technologies require equipment or environments that are not readily available.

On the other hand, extraordinary experiences through media, such as movies, are within reach. However, people can only experience the extraordinary as an observer and cannot actually experience the world inside the movie. Nevertheless, people's longing for the silver screen is so strong that some of them imagine being part of the action. There are some simple and well-known methods for synthesizing humans into an image, such as a chroma key system. In most chroma key systems, the background is static, while the user's time actively continues. This gap between the user and the synthesized space is a barrier to user immersion. Although a chroma key system that uses a video clip as a background might be better, time in a video clip passes regardless of the user's time; thus, the gap remains.

Hence, we propose *VRMixer*, a system that allows the user to be immersed into the world of a video clip with greater reality by sharing both space and time. Figure 1 shows a mixed image generated by *VRMixer*. This system assumes that space-sharing assists time-sharing which affect to immersion. By acting as a part of the video clip, such as dancing with dancers in a music video, the user can feel immersed. Although there is an obvious gap between the real world and the video clip, when people take part in the clip or when people in the clip come out from its world, this gap may become narrowed with its ambiguity.

To realize immersion, our system constructs a simple virtual 3D space (i.e., 2.5D space) from a video clip through video processing. In particular, we have implemented a video segmentation method that uses an automatic 3D graph cut and face detection to extract the human area from a video clip.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.
ACE '14, November 11 - 14 2014, Funchal, Portugal Copyright is held by the owner/author(s). Publication rights licensed to ACM. ACM 978-1-4503-2945-3/14/11...\$15.00.
<http://dx.doi.org/10.1145/2663806.2663834>

Figure 1. Image generated by VRMixer.

This method enables the system to automatically extract the human body area from the clip and detach it from the background. This realizes feelings of immersion in the virtual world and makes it seem like people in the clip are emerging from the video world. As a result, the worlds (real and video) can be mixed.

The proposed system constructs 2.5D space by positioning the background in the back and people in the front. A user can enter the 2.5D space of the video clip and stand between the people appearing in the clip and the background by measuring 3D depth using a depth camera. By sharing the real world and video clip spaces, the user’s time consciousness can be mixed with the time flow of the video clip. We aim to mix the real world and the world of video clips by sharing both space and time to realize “video reality” with VRMixer.

APPLICATIONS OF VRMIXER

In this section, we present some potential applications of VRMixer. Applications that require depth information cannot be realized with chroma key systems which do not share space. Consideration of depth in VRMixer broadens the range of application of the system.

As a straightforward application, users can co-star with people inside a video clip. In the case of chroma key content, users are always in front of people in a video clip; therefore, the users will always take the main part. However, VRMixer enables the users to stand behind the main actor in the video clip. Thus, the users will not always take the leading part, and hence, VRMixer realizes co-starring more equitably. In addition, a user can act as a stand-in for someone in a movie. In this case, the voice of the person who the user is replacing should be erased beforehand. This stand-in application can be used to practice acting skills. Although acting skills

are difficult to practice alone, VRMixer helps facilitate acting practice with virtual actors from existing movie content.

Depth information is important when considering application to dance video content, which requires various formations of people. On video sharing services, such as YouTube, a video genre called “Me Dancing” is popular among amateur dancers. Users dance to music and upload a video of their performance to the video sharing service. Using VRMixer, users can collaborate with other “Me Dancing” videos or with their favorite dancers or singers. Moreover, most “Me Dancing” videos are recorded in the user’s room, which is not always an attractive background environment. VRMixer can solve this problem by using video clips with attractive backgrounds. In addition, dance practice will be more practical if users can dance as part of the actual dance video.

RELATED WORK

To get into the world of movies, Morishima *et al.* proposed an instant movie casting project called “Dive into the Movie”[12]. In this project, Maejima *et al.* proposed the *Future Cast System*, which allows users to participate in a pre-created film as instant CG characters[11]. The project presented a novel movie appreciation method. However, the available content for “Dive into the Movie” is limited and is difficult to experience because it requires special equipment.

To bring the extraordinary world closer, using readily available equipment is required. Hunter *et al.* presented *WaaZam!*, a system that allows users to interact in a virtual environment composed of digital assets layered in 3D space, by using depth cameras[6]. Fortunately, depth cameras, such as the Kinect, are becoming readily available. *WaaZam* is a telepresence system that uses a virtual environment; is not a system for users to enter and immerse themselves in the video content. Although *WaaZam* presents a virtual 3D space instantly, it focuses on interaction between users in different spaces rather than that between the user and the video content. Therefore, it is not suitable for the applications such as dancing with people in the video content which we aim to realize with VRMixer. In addition, *WaaZam* requires manual positioning of digital assets. Focusing on human extraction, VRMixer made it possible to automatically position the people in the video content and the background in different depth.

In VRMixer, we create a richer virtual 3D environment from a video clip rather than simply allocating video frames to 3D space. This is important in realizing previously mentioned applications. To achieve this, we construct 2.5D space, which is a simple 3D image created from a 2D color image with an added depth channel. Using a 2.5D video, Kurihara *et al.* proposed *DDMixer2.5D*, which enables drag-and-drop copy-and-paste of a video object into another video clip[9]. In 2.5D videos, depth information makes it easier to segment an object in the video than it is to segment the same object using only color channels. With 2.5D video content, construction of simple virtual space from video content is significantly easier. However, a 2.5D video is not a common format, and hence, cannot be accessed easily. Thus, we use regular video clips, which are abundantly available.

Li *et al.* proposed a method to cut-and-paste a 2D video object using 3D graph cut segmentation[10]. This method can be applied to regular video clips. However, our goal is to realize a mix of video and reality mixing rather than video-to-video editing.

Creating 2.5D (or 3D) space from a 2D video clip automatically is a difficult problem to solve with the current video processing methods. Chen *et al.* proposed a 2D-to-3D video conversion method using 2.5D paper cutouts[2]. Although the resulting video has high quality, this method requires significant user input to specify the boundaries and depth of the object. Iizuka *et al.* presented an interactive technique to create 3D space from a single image; however, it also requires user input[7]. Therefore, to create 2.5D space automatically, we simplify the problem by constructing 2.5D space from the main object and the background of a video clip. We define the main object as the people appearing in a video clip and all other objects as part of the background. Although this constraint reduces the selection of available video clips, we can still use clips with people appearing in them, which is advantageous because the availability of such video clips is vast. For the optimum utilization of this constraint, we use video clips in which people play important roles, such as dance videos, as the main target.

SYSTEM IMPLEMENTATION

VRMixer preprocesses a video clip to construct its 2.5D space. This preprocess is automatically achieved with video processing, which does not require human interaction. While there are some time-consuming tools to extract regions from a video clip, the proposed system does not require any human input. After preprocessing, the user can enter the video clip by moving around his/her room as the clip plays. The synthesized result is shown in real time on a display or projector. Note that the only delay in the synthesis process is the minimum delay of the depth camera which is unnoticeable.

Figure 2 shows an overview of the VRMixer system, including preprocessing. The following section describes the methods involved in VRMixer.

Preprocess (Human extraction)

Figure 2. System overview of VRMixer.

CONSTRUCTION OF 2.5D SPACE

To construct 2.5D space, we propose a video segmentation method and a method to allocate the segmented layer within 3D space.

Video Segmentation

This section describes our video segmentation method, i.e., the part of preprocessing that is related to 2.5D space construction from a video clip. In VRMixer, each video frame is segmented into human and nonhuman (background) areas.

Automatic Human Extraction

The automatic extraction of the human area from a video clip is difficult without training. There are methods for extracting human areas from video sequences, such as histogram of oriented gradient (HOG) and local binary pattern (LBP) based methods, which require specific training data[17]. However, these methods narrow the target and are not suitable for our purposes. Therefore, to realize human extraction for a wide range of available content, we use a face detection method to identify the human area and a human detection method as support. Face detection is comparatively more accurate than human area detection even if the target person’s filming conditions are unknown. We use hierarchical fitting based on the active structure appearance model proposed by Irie *et al.*[8] as our face detection method. Face detection method is performed for each video frame. The method simultaneously detects multiple facial feature points; therefore, an approximate face size W can be estimated. Based on this face size, the area from the bottom of the face to $n \times W$ down can be roughly predicted as the probable body area of the person ($n = 8$ in the current implementation). Simultaneously, an area sufficiently far from the face can be estimated as the probable background area. To detect this background area reliably, a human extraction method based on an HOG descriptor trained by a support vector machine[4] is applied to all frames as a support to face detection. In this way, the human area in a frame that is unsuitable for face detection, e.g., a face with occlusion, can be extracted. These rough estimation results are used as a seed in the subsequent segmentation technique.

From the roughly estimated human area, we estimate the color distribution of both human and background areas by using a variational Bayesian Gaussian mixture model (VB-GMM)[3]. The estimated color distribution corresponds to the RGB pixel values of the video frames. VB-GMM construction corresponds to the inference of the representative colors required to draw each human and background areas. This is equivalent to learning the human and background areas for each video clip. In the segmentation method, we use only the mean pixel value of each Gaussian distribution in the GMM as representative colors. We do not use variance or co-variance because the roughly estimated area does not include the entire human body area; thus, the inferred distribution may be biased. In fact, the roughly estimated body area rarely includes arms; therefore, the distribution rarely reflects the color of arms. The face detection method is imperfect, and there are many omissions. Thus, the roughly estimated body and background areas represent a limited subset of the desired area.

Based on the representative colors of the human and background areas, a 3D graph cut using a min-cut/max-flow algorithm is performed[1]. The 3D graph cut creates a natural segmentation boundary. To generate better segmentation results, we introduce weighting based on the motions in a frame to the graph cut's data term. This weight is based on the assumption that the human in a video clip moves, while the background tends to be static. The motion is quantitatively calculated from an optical flow estimated by the Gunnar Farneback algorithm[5]. Then, a dilation and erosion process is performed to eliminate small noises.

Video Segmentation Results

Figure 3 (e) shows a segmentation result using the above mentioned method. For comparison, results from a simple frame-by-frame graph cut segmentation (Figure 3(b)); GrabCut[13], which is an interactive segmentation method (Figure 3(c)); and adaptive binarization (Figure 3(d)) are shown. Adaptive binarization method is a segmentation method based on the color of the human area and the background learned by VB-GMM. The proposed segmentation method and the adaptive binarization method do not require any manual input; however the graph cut and GrabCut methods require considerable manual input for each video frame. The leftmost results of graph cut and GrabCut include an example of manual input. In GrabCut, the user specifies an approximate human area with a rectangle so that miss detection does not occur outside of the rectangle.

To compare segmentation results quantitatively, we perform a human area extraction experiment. We randomly extract 30 video frames that include human beings from three dance videos (10 frames each). Using the obtained 30 frames, we manually segment the human areas and calculate the accuracy of each segmentation method. Figure 4 and Table 1 show the extraction accuracy for each segmentation method.

Figure 4. F score for each method (30 frames).

Table 1. Accuracy comparison of segmentation methods

method	precision	recall	F score
Graph cut	0.417	0.798	0.535
GrabCut	0.719	0.856	0.768
Adpt. Binarization	0.534	0.754	0.537
Proposed	0.751	0.809	0.734

Figure 4 shows the F score for each method for 30 video frames. The results indicate that both GrabCut and the proposed method perform with high accuracy. Average F scores as well as precision and recall are presented in Table 1. GrabCut shows the highest accuracy. However, it limits the prediction area to the inside of the manually input rectangle; consequently, its precision and recall values are higher than those of the other methods.

If the manual input required by GrabCut is considered, the results for the proposed method, which does not require manual input, are reasonable. However, some artifacts remain in the segmentation results, and the accuracy improvement will be addressed in future work.

In addition, the present implementation requires parameter tuning. For example, adding greater weight to the stable area is effective for a video in which a human tends to move; however, it has the opposite effect in a video where a human does not move. To achieve better segmentation results, allowing the user to provide some manual input during the initial segmentation might be an option, rather than allowing unsatisfactory results from the fully automated method. However, allowing user input is not a simple solution to video segmentation in which objects move in every frame. For the condition, numerous user inputs are required to achieve a satisfactory result. To allow efficient user input, Wang *et al.* proposed an interactive video cutout method[16] and Li *et al.* proposed an efficient boundary correction method[10]. Such effective interaction might be an option in future implementations.

2.5D Space Construction

This section describes 2.5D space construction from a pre-processed video clip. As a starting point for our research, we place the entire video frame in a user's 3D space. The user stands behind or in front of the frame and a depth camera(Kinect) is used to measure the depth. Figure 5 shows the video frame layered in the user's 3D space. The user can position a part of his/her body in front of the frame.

Figure 5. Simply layering a video frame in 3D space.

This virtual environment is similar to that constructed with WaaZam[6] in aspect of using content as it is. Here, the video

Figure 3. Comparison of human extraction methods: (a) original video frame, (b) frame-by-frame graph cut segmentation, (c) interactive human area segmentation with GrabCut [Rother et al. 2004], (d) adaptive binarization, (e) proposed automatic human area segmentation.

frame in 3D space is a flat wall, and the video sequence proceeds regardless of the user's time flow. Therefore, to share the space time of the real world and a video clip more effectively, VRMixer constructs 2.5D space.

Using the segmentation result of the main object (human area) in the video clip, the system positions the human area and the background at different depths. 2.5D space is constructed by positioning the human area at the front and the background area at the back. The user can adjust the depth of an extracted human from a video clip and that of the background manually, as well as their size in the virtual environment. VRMixer mixes objects in a real world and a video clip as shown in Figure 6. Figure 6 is a top view diagram; the areas underlined in red are shown in the output images.

Figure 6. Top view of mixed real world and video content using VRMixer.

Figure 7 shows a VRMixer output image. The user can stand between a person in the video clip and the background.

Figure 7. The user in constructed 2.5D space, standing between the person and the background.

In the current implementation, the user must set the depth and size of each layer manually using slide bar. These parameters will be adjusted automatically in a future implementation that offers normalization of the real world and the video clip.

MIXING A VIDEO CLIP AND THE REAL WORLD

With a video segmentation method and a depth camera, the actual 3D space occupied by a user can enter the world of a video clip. This would be an extraordinary experience because all video clips containing people can be used, which means the user can be immersed into any type of video content. Beyond entering the clip, the user's 3D space becomes a part of the mixed 2.5D space. For example, the user's favorite chair can be part of the video scene. Figure 7 shows a user's locker entering the video clip.

Note that the floor in the user's 3D space near the depth camera will be a floor in the mixed space. As the real-world floor recedes from the camera, it transitions to the video clip's floor. This will be the boundary where video and reality meet. As people in the video clip approach the camera, their feet will reach the floor in the user's 3D space. It appears as if the people are coming out from the 2D space into the user's 3D space. Just as people from the video can come out of their space, the user can enter the 2D video space. By merging the background of a video clip, the user can create new 2.5D worlds. For example, the user can set the back wall as the wall appearing in a video clip and a side wall as a real wall in the actual 3D space. VRMixer can exchange real world elements for elements in the video content through this mixed 2.5D space with an ambiguous boundary.

Summon a 2D person to a 3D space

Using the video segmentation method, VRMixer can position people in the video in the user's 3D space. This phenomenon was previously achieved with augmented reality (AR) technology using 3DCG characters. VRMixer realizes this with actual photographed 2D people. For example, Figure 8 shows a 2D person emerging from her video world and dancing in the user's room.

Figure 8. A person in a video clip entering the 3D world.

Application to Dance Practice

A practical application of VRMixer is dance practice. People often learn dance choreography by watching a video clip. In such situations, they tend to confuse direction. If the user is in the same space as the dancer, discerning right and left and learning dance formations become much easier. To adapt the system for dance lessons, we implement a horizontal inversion mode that presents the VRMixer display as a virtual mirror.

Tsuchida *et al.* proposed a system for practicing dance formations with a self-propelled screen[15]. VRMixer can also provide dance formation lessons using the virtual mirror with a virtual dancer inside. Note that VRMixer reflects user motion with nearly no delay, which results in the output appearing more like a mirror. A dance practice where a user is substituted for a specific member of a dance group can be realized with VRMixer.

EARLY REACTIONS

In this section, we discuss early reactions from VRMixer users. One user felt as if the worlds were mixing as the feet of a person in a video clip stepped into the real world.

Another user felt that the accuracy of both video segmentation and depth measurement was noticeable. Improving the accuracy of video segmentation will be a focus of our future work, and the accuracy of depth measurement can be improved as depth cameras improve.

Another piece of feedback is that the system would be more interesting if users could dance in place of a member of a dance group. To realize this function more adequately, we must erase a specific object and inpaint the background. Kurihara *et al.* proposed an inpainting method using the Patch-Match algorithm[9], which can be applied to the proposed system in a future implementation.

One test subject reported a pleasurable dance experienced because she could dance with her favorite dance group and join hands with them.

The quantitative evaluation of VRMixer is part of our future work.

Limitations and Future Work

The accuracy of the video segmentation method is important. As mentioned earlier, our segmentation method relies on a face detection method; thus if no face can be detected in a video clip, the entire frame will be considered the background and 2.5D space will not be constructed effectively. An HOG-based human area detection method can support this; however, the results are not perfect. The present scope of a VRMixer target video is limited to videos in which the human area can be segmented with our segmentation algorithm. This limits the target video content to videos for which the face detection method and 3D graph cut segmentation algorithm work, i.e., video clips with relatively clear faces and simple objects. To improve accuracy and address failure cases, allowing user input prior to segmentation may be an option in future implementations. If we accept user input, our video segmentation method can be extended to any video object while improving segmentation accuracy. An interactive segmentation method, such as those proposed by Wang *et al.*[16], and Li *et al.*[10] is a possible approach.

In addition, due to technical limitations, video clips with many scene changes are not suitable as a target video of VRMixer. While scenes in the video change, the user's environment does not change; thus, immersion will be interrupted during scene changes. Therefore, video content with minimal scene changes, such as dance video, is suitable for the target video.

In the future, we aim to mix more than two video clips so that a user can recast videos with desired co-stars. Furthermore, by introducing telepresence function, such as WaaZam[6] presented, mixing of multiple real environments and video clips may be possible.

We also intend to consider the interaction between the user and the people in a video clip. For example, if the person in the clip moves as the user pushes, the boundary between the real world and a video clip might become more ambiguous. It may be possible to physically summon a person by using a robot or other device such as the self-propelled screen proposed by Tsuchida *et al.*[15].

CONCLUSION

In this paper, we have presented *VRMixer*, a system that mixes the real world and a video clip using video segmentation. *VRMixer* allows the user to become immersed in the video clip. An individual appearing in the clip can appear in the user's actual 3D space. This mixture of real world and video content provides a new way to appreciate video content. However, improving the accuracy of video segmentation is a problem that must be solved. This will be addressed in a future implementation of the system, e.g., by using an interactive approach.

As interactive devices continue to develop, the immersion provided by *VRMixer* will also improve. For example, by using a free air tactile-feedback device, such as *AIREAL*, it may be possible for the user to feel a virtual object[14].

We aim to make it possible for any existing video clip to be used with the proposed system. By using films from the last century, it might be possible to duplicate a historical event and observe the scene with you as a participant. This can potentially be used to simulate time travel.

Video clips, such as movies or music videos, are some of the best sources of entertainment. The number of videos continues to increase; however, the way such content is consumed have not changed significantly. We will continue to improve this system to realize new interactions between humans and digital content.

ACKNOWLEDGMENTS

This work was supported by OngaCREST, CREST, JST and partially supported by JSPS Grant-in-Aid for JSPS Fellows, JSPS.

REFERENCES

1. Boykov, Y., and Kolmogorov, V. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Trans. on PAMI* Vol. 26, 9 (2004), 1124–1137.
2. Chen, J., Paris, S., Wang, J., Matusik, W., Cohen, M., and Durand, F. The video mesh: A data structure for image-based three-dimensional video editing. In *Proc. of ICCP2011* (2011), 1–8.
3. Corduneanu, A., and Bishop, C. Variational bayesian model selection for mixture distributions. In *Proc. of AI Statics2001* (2001), 27–34.
4. Dalal, N., and Triggs, B. Histograms of oriented gradients for human detection. In *Proc. of CVPR2005* (2005), 886–8893.
5. Farneback, G. Two-frame motion estimation based on polynomial expansion. *Image Analysis* (2003), 363–370.
6. Hunter, S., Maes, P., Tang, A., Inkpen, K., and Hessey, S. Waazam!: supporting creative play at a distance in customized video environments. In *Proc. of CHI2014* (2014), 1197–1206.
7. Iizuka, S., Kanamori, Y., Mitani, J., and Furui, Y. Efficiently modeling 3d scenes from a single image. *IEEE Computer Graphics and Applications* Vol. 32, 6 (2012), 18–25.
8. Irie, A., Takagiwa, M., Moriyama, K., and Yamashita, T. Improvements to facial contour detection by hierarchical fitting and regression. In *Proc. of ACPR2011* (2011), 273–277.
9. Kurihara, T., Okabe, M., and Onai, R. Ddmixer2.5d: drag and drop to mix 2.5d video objects. In *Proc. of UIST2013 Adjunct* (2013), 69–70.
10. Li, Y., Sun, J., and Shum, H. Video object cut and paste. In *ACM Trans. on Graphics*, vol. 24 (2005), 595–600.
11. Maejima, A., Wemler, S., Machida, T., Takebayashi, M., and Morishima, S. Instant casting movie theater: The future cast system. *IEICE Trans. on Information and Systems* (2008), 1135–1148.
12. Morishima, S., Yagi, Y., and Nakamura, S. Instant movie casting with personality dive into the movie system. *Virtual and Mixed Reality 2011* (2011), 187–196.
13. Rother, C., Kolmogorov, V., and Blake, A. "grabcut": interactive foreground extraction using iterated graph cuts. *ACM Trans. on Graphics* Vol. 23, 3 (2004), 309–314.
14. Sodhi, R., Poupyrev, I., Glisson, M., and Israr, A. Areal: Interactive tactile experiences in free air. *ACM Trans. on Graphics* Vol. 32, 4 (2013).
15. Tsuchida, S., Terada, T., and Tsukamoto, M. A system for practicing formations in dance performance supported by self-propelled screen. In *Proc. of AH2013* (2013), 178–185.
16. Wang, J., Bhat, P., Colburn, R., Agrawala, M., and Cohen, M. Interactive video cutout. In *ACM Trans. on Graphics*, vol. 24 (2005), 585–594.
17. Wang, X., Han, T., and Yan, S. An hog-lbp human detector with partial occlusion handling. In *Proc. of ICCV2009* (2009), 32–39.