

Incorporating Audio Signals into Constructing a Visual Saliency Map

Jiro Nakajima¹, Akihiro Sugimoto², and Kazuhiko Kawamoto¹

¹ Chiba University, Chiba, Japan

`nakajima13@chiba-u.jp`, `kawa@faculty.chiba-u.jp`

² National Institute of Informatics, Tokyo, Japan
`sugimoto@nii.ac.jp`

Abstract. The saliency map has been proposed to identify regions that draw human visual attention. Differences of features from the surroundings are hierarchially computed for an image or an image sequence in multiple resolutions and they are fused in a fully bottom-up manner to obtain a saliency map. A video usually contains sounds, and not only visual stimuli but also auditory stimuli attract human attention. Nevertheless, most conventional methods discard auditory information and image information alone is used in computing a saliency map. This paper presents a method for constructing a visual saliency map by integrating image features with auditory features. We assume a single moving sound source in a video and introduce a sound source feature. Our method detects the sound source feature using the correlation between audio signals and sound source motion, and computes its importance in each frame in a video using an auditory saliency map. The importance is used to fuse the sound source feature with image features to construct a visual saliency map. Experiments using subjects demonstrate that a saliency map by our proposed method reflects human's visual attention more accurately than that by a conventional method.

Keywords: gaze, visual attention, visual saliency, auditory saliency, audio signal, video, sound source feature.

1 Introduction

Visual focus of attention can be an important cue for understanding human behaviors or supporting human activities [17]. To estimate human visual focus of attention, a computational model of visual saliency map, which identifies image regions that draw more human attention, was proposed by Itti *et al.* [9]. The saliency map is computed based on center-surround contrast of image features such as color, intensity or orientation in an image in a fully bottom-up manner.

Following a method computing a saliency map of a still image [9], many types of saliency maps have been intensively studied. Itti *et al.* [8] employed a feature-integration theory [16] and incorporated low-level dynamic features such as motions or flickers to extend the model to be applicable to a video. Harel *et al.* [6] introduced a graph structure over image maps that arise in

saliency computation to improve the accuracy of a saliency map. Bruce *et al.* [2] used Shannon's self-information in measuring saliency to construct a model of bottom-up overt attention. Cerf *et al.* [3] proposed combining face detection with a saliency map computed from contrasts of image features. Taking into account the difference between central (fovea) and peripheral areas in computing a saliency map was proposed by Kubota *et al.* [12]. Saliency computation from an ego-centric video was proposed by Yamada *et al.* [17].

Eye movements are indeed affected by visual stimuli and visual attention is drawn accordingly. Human attention, however, is attracted by auditory stimuli as well. When we hear an extraordinary sound, we tend to look at the direction of the sound source even if it is not visually salient. A video usually contains sounds and auditory stimuli are available. Nevertheless, most existing methods for computing saliency from a video discard such sounds and use image features alone. Combining image features with sounds in saliency computation is promising to significantly improve the accuracy of a saliency map.

In contrast to visual saliency, few models to determine salient audio signals have been proposed [11,14,10]. Kayer *et al.* [11] proposed an auditory saliency map model to detect salient sounds embedded in noisy backgrounds. In that work, the sound wave is converted to a time-frequency representation to have an intensity image and then important auditory features (intensity, frequency contrast and temporal contrast) are extracted on different scales. An auditory saliency map is computed using the center-surround contrast of the features in the same way of computing a visual saliency map.

In the context of human-machine interaction, there are a few attempts in which audio signals are taken into account to construct a visual saliency map [7,14]. Since they focus on real-time performance for the interaction, the accuracy of constructed saliency maps is not deeply investigated; how to control the system in real time is more important. In [7], for example, a microphone array is used as a device to estimate high salient auditory stimuli in real time. Visual saliency map and auditory saliency map are fused by a weighted linear combination into a single audio-visual saliency map, which is used for robot head control. This method, however, cannot be applicable without any microphone array. Moreover, it cannot deal with recorded videos.

This paper presents a method for constructing a visual saliency map for a video¹ by incorporating sounds in its computation. We assume a single moving sound source. To treat sounds like image features, we introduce a sound source feature which represents the location of a sound source in each frame. The location of a sound source is computed using the correlation with audio signals and sound source motion by [18]. Our method then computes the importance of the sound source feature in each frame using an auditory saliency map by [11]. The importance is then used to fuse the sound source feature with image features. By this way, we obtain a visual saliency map that reflects sounds contained in

¹ Hereafter, we discriminate an image sequence from a video to highlight the existence/non-existence of sounds. In this paper, a video contains audio signals while an image sequence does not.

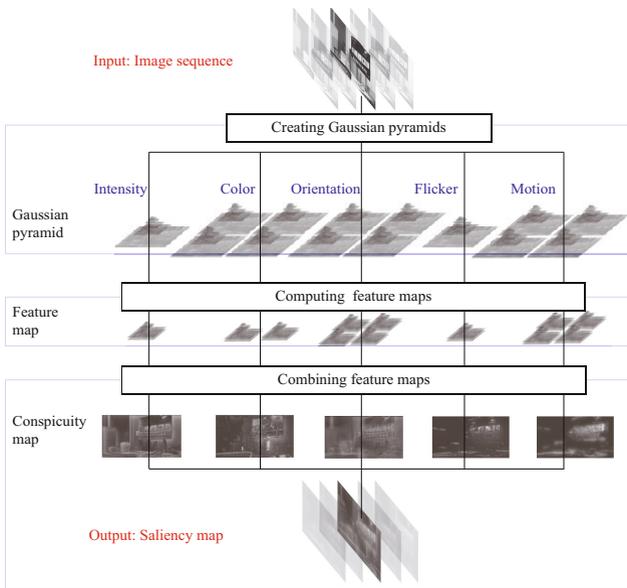


Fig. 1. Computation flow of a saliency map

a video. Experiments using subjects confirm that our proposed saliency map outperforms a saliency map computed using image features alone.

2 Visual Saliency Map

We here present a brief overview on how to compute a bottom-up visual saliency map [8]. The core concept in computational visual saliency is extracting regions with vastly different image features than their surrounding regions. Figure 1 depicts the computation flow. An input frame is decomposed into a set of multi-scale feature maps which extract local spatial discontinuities in the modalities of color, intensity, orientation, motion and flicker. All feature maps are then combined into a unique scalar saliency map which encodes for the salience of a location in the scene irrespectively of the particular feature which detected this location as conspicuous.

Gaussian pyramids [5], hierarchical images in multi-scales, are first created from an input frame. For color and intensity, five base images are first generated using an input RGB channels, denoted by R, G and B , of a frame: $i = (R + G + B)/3$, $r = R - (G + B)/2$, $g = G - (R + B)/2$, $b = B - (R + G)/2$ and $y = (R+G)/2 - |R-G|/2 - B$. Four base images r, g, b and y are then individually normalized by i and small values (for example, less than $1/10$) are set to zero. Next, Gaussian pyramids of each image are created by iteratively applying the pair of smoothing with a Gaussian filter and down-sampling. We denote the pyramids by $i(\sigma), r(\sigma), g(\sigma), b(\sigma)$ and $y(\sigma)$ where $\sigma = 0, 1, \dots, 8$ represents the

Algorithm 1. The max-lacal normalization $N(\cdot)$

1. Normalizing the values in the map to a fixed range $[0, M]$.
 2. Set \bar{M} as the average of the local maximal values over the map, each of which is strictly less than M .
 3. Multiply $(M - \bar{M})^2$ to all the values in the map.
-

scale. Gaussian pyramid $i(\sigma)$ is used as the intensity feature while Gaussian pyramids $r(\sigma), g(\sigma), b(\sigma)$ and $y(\sigma)$ are as the color features. For orientation, on the other hand, four Gabor pyramids $o(\sigma, \theta)$ ($\theta = 1, 2, 3, 4$) are created from base image i [5]. For the modalities of dynamic features (motion and flicker), orientation and intensity are used for creating pyramids. Four motion pyramids $m(\sigma, \theta)$ are created from $o(\sigma, \theta)$ and its shift while flicker pyramid $f(\sigma)$ is created by taking the absolute difference of $i(\sigma)$'s between two successive frames.

Feature maps are computed in a center-surround structure akin to visual receptive fields from the Gaussian pyramids of image features. Center-surround operations are implemented by the absolute across-scale subtraction, i.e., absolute difference between a fine and a coarse scales in a pyramid in concern. When two images at different scales in a pyramid are given, the bilinear interpolation [4] is first applied to a smaller image so that the two images have the same size. The corresponding pixel-wise subtraction between the two images is then computed. We note that color feature maps follow the opponent color theory (we thus have two color-opponent feature maps). We first compute the subtraction in each pair of opponent color and then apply the absolute across-scale subtraction.

For each feature, its corresponding feature maps are individually normalized and then combined to have a single conspicuity map. The max-local normalization²[9], denoted by $N(\cdot)$, highlights the most discriminative features in each feature map. Algorithm 1 shows the detail of this normalization. The across-scale addition is applied to normalized feature maps, which results in a conspicuity map. We remark that two color-opponent feature maps are combined into a color conspicuity map. We denote the conspicuity map for each feature (intensity, color, orientation, flicker, and motion) by $\bar{i}(\mathbf{x}, t)$, $\bar{c}(\mathbf{x}, t)$, $\bar{o}(\mathbf{x}, t)$, $\bar{f}(\mathbf{x}, t)$, and $\bar{m}(\mathbf{x}, t) \in \mathbb{R}$, where \mathbf{x} is the coordinates of a point in a frame and $t \in \{0, 1, \dots, T - 1\}$ is the frame number.

Finally, the normalized conspicuity maps are all linearly combined to have a single saliency map for the frame:

$$s_v(\mathbf{x}, t) = \frac{1}{5} \left\{ N(\bar{i}(\mathbf{x}, t)) + N(\bar{c}(\mathbf{x}, t)) + N(\bar{o}(\mathbf{x}, t)) \right. \\ \left. + N(\bar{f}(\mathbf{x}, t)) + N(\bar{m}(\mathbf{x}, t)) \right\}. \quad (1)$$

If a point has a large value, the point highly attracts the visual focus of attention. A point with a low value, on the other hand, tends not to attract attention.

² The range becomes $[0, M]$ after this normalization. M is a given constant value. We set M to 1.0 in our experiments.

3 Saliency Map with Audio Signals

We design a visual saliency map model that reflects saliency of the spatial location of a sound source. To treat sounds like image features, we introduce a sound source feature which represents the location of the sound source in each frame. We also compute an importance of the sound source feature in each frame.

3.1 Detecting a Sound Source Feature

As we assume a single moving sound source, we have to identify the location of the sound source to use the sound source feature. Motion of a sound source can be detected using image features in a video. Therefore, the correlation between sound source motion and audio signals can be used to identify the location of the sound source.

We employ a method [18] to detect the sound source location where the location of a sound source is detected by using a correlation with sound source motion and audio signals. This method computes a visual inconsistency feature representing motion changes at each point in a frame, and an audio inconsistency feature representing audio changes at each frame in a video. From these features, an audiovisual correlation at each point at each frame is computed by tracking points that maximizing the correlation between the visual and audio features using a beam search. This audiovisual correlation represents synchronization of motion changes and audio changes in a video. Assuming that human visual attention is attracted by motions that synchronize with audio signals, it is reasonable to use this audiovisual correlation as our sound source feature. We denote the sound source feature by $h(\mathbf{x}, t)$ where \mathbf{x} is the coordinates of a point in a frame and t is the frame number.

3.2 Computing Weights for Frames

The sound source feature does not account for changes in audio signal because the correlation between sound source motion and audio signals cannot capture changes of audio signals. When we hear a loud sound, our attention is attracted by the sound at the beginning. Such attention, however, is not preserved if we keep listening to the same loud sound. Therefore, in order to reflect changes of audio signals into a saliency map, we have to assign a weight to each frame which represents discriminative changes of audio signals over frames. We use auditory saliency to measure this weight according to the insight that human attention is attracted by salient sounds.

The auditory saliency map is a model to analyze saliency of audio signals. We employ an auditory saliency map [11] to compute auditory saliency. The auditory saliency map is computed by converting audio signals to a two-dimensional image, i.e., a sound spectrogram, which is a time-frequency representation of a sound. Sound intensity, frequency contrast, and temporal contrast are extracted as image features from the sound spectrogram. The auditory saliency map is

obtained by hierarchially computing differences of these features from the surroundings in multiple resolutions and fusing them in a fully bottom-up manner like a visual saliency map. We denote the auditory saliency map by $s_a(t, f)$ where f is frequency.

For each frame, we compute its weight $w(t)$ of the sound source feature from the auditory saliency map $s_a(t, f)$. Since the dimension of $s_a(t, f)$ is two and that of $w(t)$ is one, dimension reduction is required. Assuming that human attention is attracted by a sound derived from a single sound source and that human turns attention to the most salient sound in each frame, we extract the most high salient frequency in each frame as the weight of the frame. This selection of frequency allows us to reduce the effect of audio noises as well. The most salient sound at frame t is defined as

$$w(t) = \max_f s_a(t, f). \quad (2)$$

$w(t)$ is normalized the range of $[0, W]$. As we see in the experiments, W depends on the video. We experimentally determine W .

3.3 Constructing a Saliency Map with Audio Signals

We incorporate the sound source feature $h(\mathbf{x}, t)$ and its weight $w(t)$ into constructing a visual saliency map. Since $h(\mathbf{x}, t)$ is not a conspicuity map, $h(\mathbf{x}, t)$ cannot be directly combined with conspicuity maps $\bar{i}(\mathbf{x}, t)$, $\bar{c}(\mathbf{x}, t)$, $\bar{o}(\mathbf{x}, t)$, $\bar{f}(\mathbf{x}, t)$ and $\bar{m}(\mathbf{x}, t)$, each of which is derived from an image feature. According to the feature integration theory by Treisman [16], each feature is independently processed in parallel and fused in the end. In our case, independently of the image features, $h(\mathbf{x}, t)$ is computed as the sound source location that is synchronized with audio signals in a video. We can thus deal with $h(\mathbf{x}, t)$ like other image features and construct a conspicuity map for $h(\mathbf{x}, t)$ in the same way as those for the image features.

The process to construct the conspicuity map for $h(\mathbf{x}, t)$ is as follows:

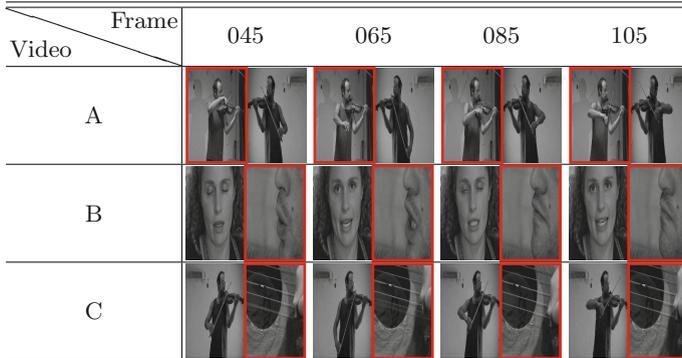
1. create a sound source Gaussian pyramid $h(\sigma)$ from the audiovisual correlation in each scale of the intensity Gaussian pyramid $i(\sigma)$,
2. compute sound source feature maps based on the mechanisms of center-surround,
3. compute a sound source conspicuity map $\bar{h}(\mathbf{x}, t)$ by normalizing and combing the sound source feature maps.

We construct our proposed saliency map for a video by linearly combining $\bar{h}(\mathbf{x}, t)$ with the conspicuity maps derived from static image features (intensity, color, orientation) and dynamic image features (flicker and motion) using weight $w(t)$.

$$s(\mathbf{x}, t) = \frac{1}{5 + W} \left\{ N(\bar{i}(\mathbf{x}, t)) + N(\bar{c}(\mathbf{x}, t)) + N(\bar{o}(\mathbf{x}, t)) \right. \\ \left. + N(\bar{f}(\mathbf{x}, t)) + N(\bar{m}(\mathbf{x}, t)) + w(t)N(\bar{h}(\mathbf{x}, t)) \right\}. \quad (3)$$

Table 1. Details of videos

Video	#frames	fps	sound source location
A	121	25	left
B	115	25	right
C	112	25	right

**Fig. 2.** Example of the input videos. Image sequences surrounded in red indicate the sound source locations.

4 Experiments

The proposed saliency map was evaluated in terms of how much scan-path data fit human visual attention. We compute the conventional saliency map s_v (Eq. (1)) and our proposed saliency map s (Eq. (3)) from input videos, and compared their performances and found that our proposed saliency map outperformed the conventional saliency map.

4.1 Experiment Set-Up

We used three pairs of an image sequence and its corresponding audio signals: videos A, B, and C [1]. Table 1 shows the details of the videos A, B, and C, and some example frames are illustrated in Fig. 2. Each video used in our experiment consists of combined two image sequences aligned in the right and left, and the audio signal sound is attached to the sequence in only one side. We say right if the sound is attached to the image sequence in right-hand side and left if the sound is attached in left-hand side (cf. Table 1). The color feature c was not used in our experiments since all the frames of the videos are monochrome.

We used an eye tracker Tobii TX300³ to detect scan-pathes of 10 subjects. The scan-path data were used as the ground truth for evaluation. We used a chin

³ Tobii TX300 eye tracker. <http://www.tobii.com/en/eye-tracking-research/global/products/hardware/tobii-tx300-eye-tracker/>

support to restrict subjects' head movements. With this restriction, the accuracy of eye tracking is increased and stable scan-path data become available. We used speakers that are built in the eye tracker to output audio signals. We obtained 10 scan-path data by displaying the videos to the subjects. We identified as gaze points the points among the scan-path data whose movements are not greater than 2.12 pixels per 0.1 seconds.

4.2 Evaluation Criteria

We used a Receiver Operating Characteristic (ROC) curve [15] and a Normalized Scanpath Saliency (NSS) [13] as evaluation criteria of the saliency map. This is because these criteria are known to be better suited to scan-path evaluation.

The ROC is a curve that plots a false detection rate (labelling a non-gazed location as fixed) as a function of a detection rate (labelling gazed locations as fixed). The more the ROC curve is located at an upper left, the higher the accuracy of the saliency map becomes. Moreover, we calculated an Area Under the Curve (AUC) value of the ROC curve. The closer the AUC value is 1.0, the higher the accuracy of the saliency map becomes.

The NSS measures significant difference from a random scan-path movement. The NSS value is calculated as the average of saliency over the entire saliency map that is normalized into mean 0.0 and variance 1.0. If the NSS value is greater than 0.0, the saliency map is significant compared with a random case.

4.3 Results and Discussion

As mentioned above, W depends on the video. This is because W affects the balance between the sound source feature and the image features. To find an appropriate W for each video, we changed W by 1.0 from 0.0 to 10.0 and evaluated the saliency map of each video using our evaluation criteria⁴ (we also tested the cases of $W = 15.0$ and 20.0). Figures 3 and 4 show AUC values and NSS values under different W s.

From Figs. 3 and 4, as the overall tendency, we observe that as W increases, both AUC and NSS increase at the beginning, then decrease and finally become almost constant. This observation suggests that the accuracy of our proposed saliency map depends on W and that an appropriate W exists that achieves the most accurate saliency map: as W increases, the accuracy is improved at the beginning and then degraded to converge at a constant. The reason for this property comes from the following facts. (1) Saliency of gaze points is increased by incorporating the sound source feature at the beginning. (2) Such tendency continues until the image features and the sound source feature are balanced. (3) Once the sound source feature is overweighted, saliency of gaze points starts decreasing. (4) The sound source feature gradually dominates saliency values (and image features are almost neglected finally). We remark that if the accuracy

⁴ Note that $W = 0.0$ indicates $s = s_v$. The case of $W = 0.0$ corresponds to the method without using auditory feature.

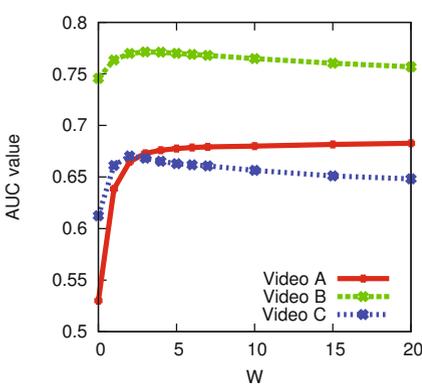


Fig. 3. AUC under different W

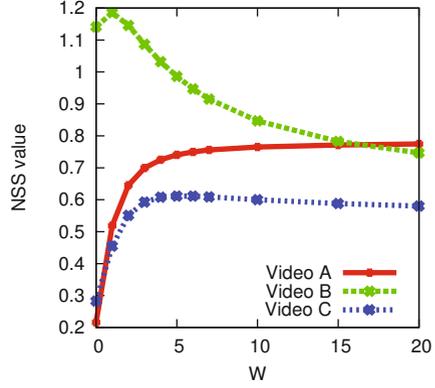
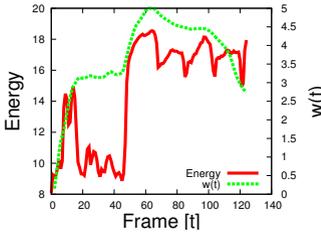
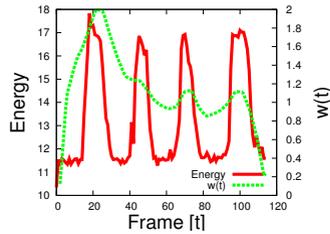


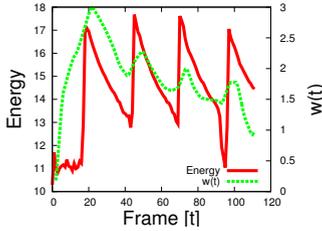
Fig. 4. NSS under different W



(a) Video A



(b) Video B



(c) Video C

Fig. 5. Audio energy and $w(t)$

of sound source feature detection is poor, saliency of a wrong region, i.e., a region incorrectly identified as the sound source location, is increased by incorporating the sound source feature. The accuracy of the saliency map is degraded in such a case. Indeed, in Figs. 3 and 4, there are cases where the accuracy of the saliency map greatly decreases. We selected and fixed the most effective value of W for each video: $W = 5.0$ (video A), $W = 2.0$ (video B) and $W = 3.0$ (video C). These W s were used to obtain the results below.

Figure 5 shows an audio energy calculated by audio signals and $w(t)$. We observe that $w(t)$ becomes large at frames when the audio energy greatly changes.

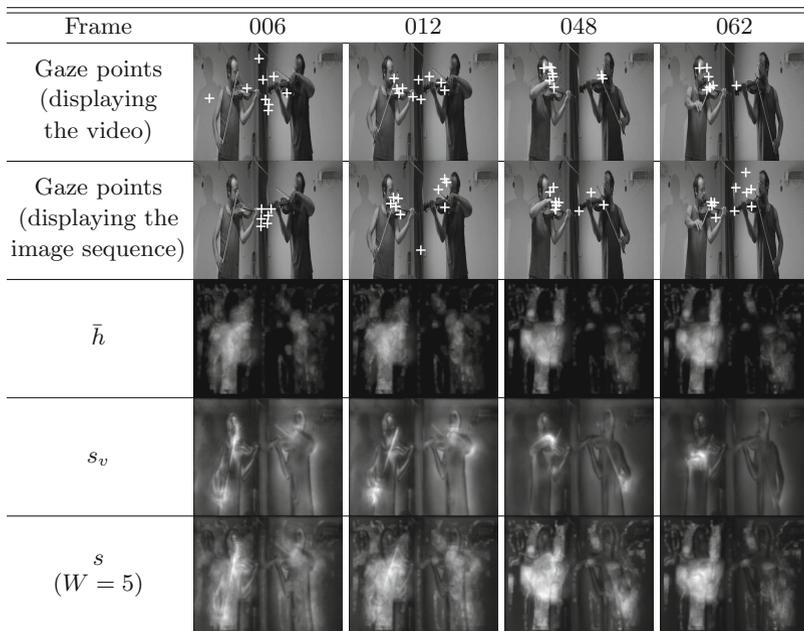


Fig. 6. Examples of images and saliency maps of Video A

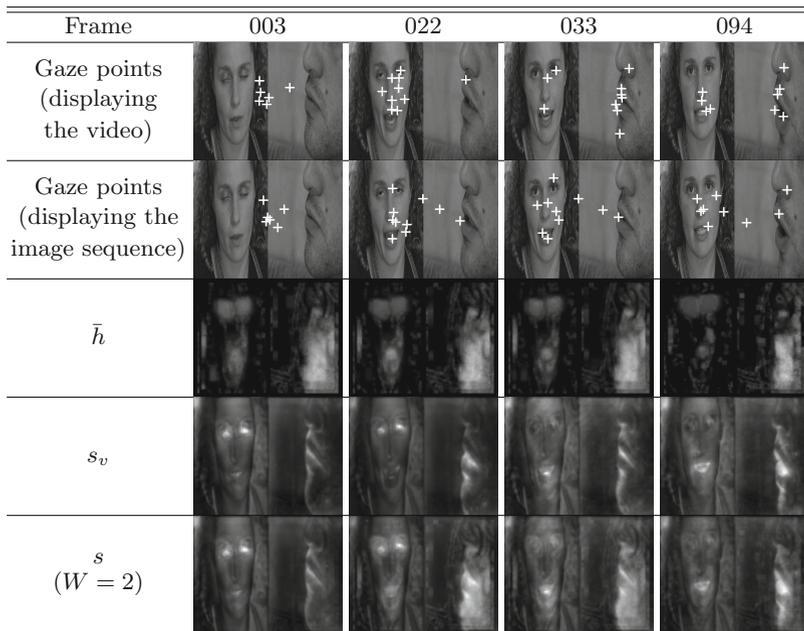


Fig. 7. Examples of images and saliency maps of Video B

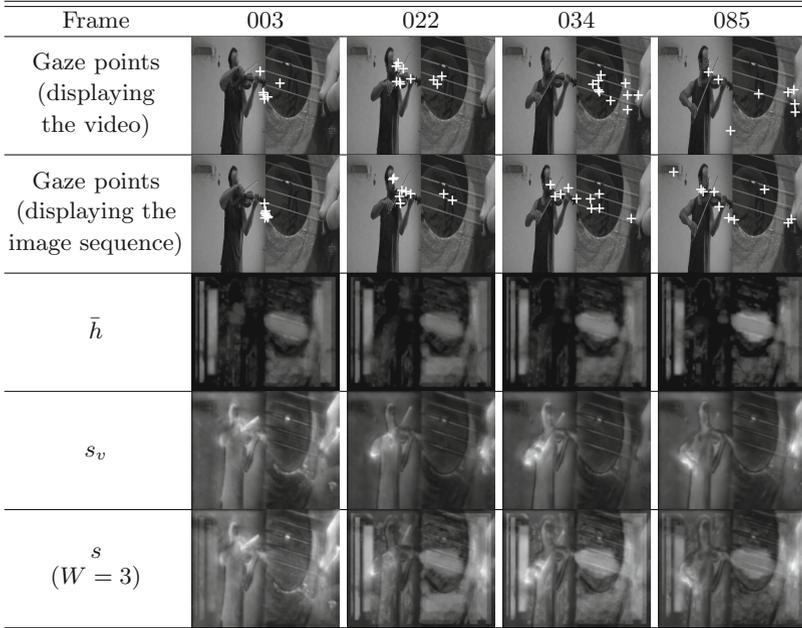


Fig. 8. Examples of images and saliency maps of Video C

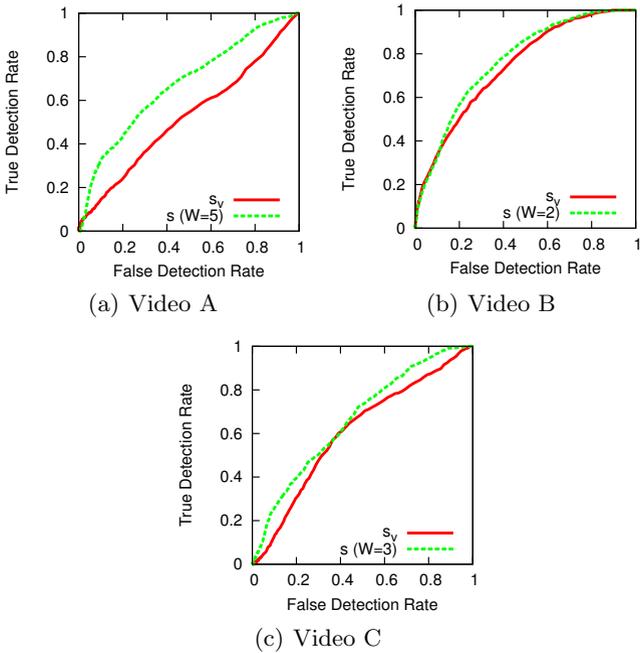


Fig. 9. ROC curve

Thus, it is possible to reflect discriminative changes of audio signals over frames by using $w(t)$.

Figures 6, 7, and 8 show examples of images and saliency maps of videos A, B, and C, respectively. In each figure, gaze points obtained by displaying the video, gaze points obtained by displaying only the image sequence, the conspicuity map of the sound source feature \bar{h} , the conventional visual saliency map s_v , and our proposed saliency map s are illustrated in this order. We remark that gaze points are represented as white crosses. In the saliency maps, high salient regions are displayed in white. If the white crosses fall into the white regions, the saliency map is highly accurate. We observe that gaze points are affected by the existence/non-existence of audio signals and the gaze points are drawn to the sound source (left or right) depending on the frame. This indicates that human gazes are attracted by not only visual stimuli but also auditory stimuli. We also observe that (1) \bar{h} is indeed reflected in s_v , and (2) how much \bar{h} is reflected in s_v depends on the frame. Therefore, we can conclude that the proposed saliency map appropriately reflects changes of audio signals in terms of $w(t)$.

Figure 9 shows the ROC curves of videos A, B, and C. The ROC curve of s is located in a more upper left than s_v for all the videos. Therefore, the accuracy of the saliency map increases by incorporating sound source feature \bar{h} . This confirms that incorporating audio signals significantly improves a saliency map.

5 Conclusions

This paper presented a method for constructing a visual saliency map for a video by incorporating sounds in its computation. To deal with sounds like image features, we introduced a sound source feature which represents the location of a sound source in each frame. The location of a sound source was computed using the correlation with audio signals and sound source motion. Our method then computed the importance of the feature in each frame using an auditory saliency map. The importance was then used to fuse the sound source feature with image features. By this way, we obtained a visual saliency map that reflects sounds contained in a video. Experiments using subjects confirmed that our proposed saliency map outperforms a saliency map computed using image features alone.

Acknowledgments. This work was in part supported by CREST.

References

1. Barzelay, Z., Schechner, Y.Y.: Harmony in motion. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2007, pp. 1–8 (2007)
2. Bruce, N.D.B., Tsotsos, J.K.: Saliency based on information maximization. In: NIPS (2005)
3. Cerf, M., Harel, J., Einhäuser, W., Koch, C.: Predicting human gaze using low-level saliency combined with face detection. In: NIPS (2007)

4. Gangnet, M., Perny, D., Coueignoux, P.: Perspective mapping of planar textures. *Computers & Graphics* 8(2), 115–123 (1984)
5. Greenspan, H., Belongie, S., Goodman, R., Perona, P., Rakshit, S., Anderson, C.H.: Overcomplete steerable pyramid filters and rotation invariance. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 222–228 (1994)
6. Harel, J., Koch, C., Perona, P.: Graph-based visual saliency. In: *Advances in Neural Information Processing Systems 19, Proceedings of the Twentieth Annual Conference on Neural Information Processing Systems*, Vancouver, British Columbia, Canada, December 4–7, pp. 545–552 (2006)
7. Huber, D.J., Khosla, D., Dow, P.A.: Fusion of multi-sensory saliency maps for automated perception and control. In: *Proceedings of the SPIE* (2009)
8. Itti, L., Dhavale, N., Pighin, F.: Realistic avatar eye and head animation using a neurobiological model of visual attention. In: Bosacchi, B., Fogel, D.B., Bezdek, J.C. (eds.) *Proceedings of the SPIE 48th Annual International Symposium on Optical Science and Technology*, vol. 5200, pp. 64–78. SPIE Press (2003)
9. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* 20(11), 1254–1259 (1998)
10. Kalinli, O., Narayanan, S.S.: Prominence detection using auditory attention cues and task-dependent high level information. *IEEE Transactions on Audio, Speech & Language Processing* 17, 1009–1024 (2009)
11. Kayser, C., Petkov, C.I., Lippert, M., Logothetis, N.K.: Mechanisms for Allocating Auditory Attention: An Auditory Saliency Map. *Current Biology* 15(21), 1943–1947 (2005)
12. Kubota, H., Sugano, Y., Okabe, T., Sato, Y., Sugimoto, A., Hiraki, K.: Incorporating visual field characteristics into a saliency map. In: *Proceedings of the Symposium on Eye Tracking Research and Applications, ETRA 2012*, pp. 333–336. ACM (2012)
13. Peters, R.J., Iyer, A., Itti, L., Koch, C.: Components of bottom-up gaze allocation in natural images. *Vision Research* 45(8), 2397–2416 (2005)
14. Schauerte, B., Kühn, B., Kroschel, K., Stiefelhagen, R.: Multimodal saliency-based attention for object-based scene analysis. In: *Proceedings of the 24th International Conference on Intelligent Robots and Systems (IROS)*. IEEE/RSJ (2011)
15. Tatler, B.W., Baddeley, R.J., Gilchrist, I.D.: Visual correlates of fixation selection: effects of scale and time. *Vision Research* 45(5), 643–659 (2005)
16. Treisman, A.M., Gelade, G.: A feature-integration theory of attention. *Cognitive Psychology* 12(1), 97–136 (1980)
17. Yamada, K., Sugano, Y., Okabe, T., Sato, Y., Sugimoto, A., Hiraki, K.: Attention prediction in egocentric video using motion and visual saliency. In: Ho, Y.-S. (ed.) *PSIVT 2011, Part I. LNCS*, vol. 7087, pp. 277–288. Springer, Heidelberg (2011)
18. Yuyu, L., Sato, Y.: Visual localization of non-stationary sound sources. In: *Proceedings of the 17th ACM International Conference on Multimedia*, pp. 513–516. ACM (2009)