Tracking a Screen and Detecting its Rate of Change in 3-D Video Scenes of Multipurpose Halls

N. Charara, I. Jarkass, M. Sokhn, O. Abou Khaled and E. Mugellini

Abstract—**This paper presents an automatic approach to track a wide screen in a multipurpose hall video scene. Once the screen is located, this system also generates the temporal rate of change, using edge detection based method. Our approach adopts a scene segmentation algorithm that explores visual features (texture) and depth information to perform efficient screen localization. The cropped region which refers to the wide screen undergoes a salient visual cues extraction to retrieve the emphasized changes required in rate-of-change computation. In addition to video document indexing and retrieval, this work can improve the machine vision capability in behavior analysis and pattern recognition.**

Index Terms—**Edge histogram, Pattern recognition, Scene segmentation, Slide change detection, Similarity based classifier.**

1. Introduction

In recent years, we have witnessed great advance in electronic imaging and its deployment around the world. The decreasing costs of video record equipment have resulted in enormous volume of video data. This excessive amount of information constitutes a significant challenge for both video management systems and video-based behavior recognition systems.

Due to the complexity of analyzing video footage, several computer vision solutions impose addressing the change detection problem. We mean here by *change* both scene change and slide change that can occur within a wide screen. *Scene* is defined as a sequential collection of shots unified by a common event [1], where *shot* is an unbroken continuous sequence of frames from one camera. Note that the *Slide* keyword refers to the presentation slide which is mainly projected electronically at the wide screen in lectures or conferences. While most of available applications exploit scene/slide change detection in video information management issues as content-based video retrieval and indexing (intelligent lecture recording, efficient browsing …etc.), we benefit from this kind of video analysis, to provide semantic level information for scene analysis. After tracking and locating the screen in video-scene footage of multipurpose hall, the detected rate of change of the screen content is linked to the usage identification of this hall as a partial discriminative feature. We argue this choice since the rate of change varies according to the use; a low rate of change reflects a normal presentation slide swiping therefore a conference purpose, whereas a movie show or cinema manipulation relatively provide a high rate of scene change.

Scene/Slide change detection has been studied in several previous research works [2], [3], [4], [5],[6], [7], [8] and employed in many systems including [9], [10]. The audio and visual cues are adopted separately in most approaches, while combining these two issues [2] provides certainly more efficient results but additional complexity cost. Motion information, layout of text region and SIFT features, are examples of traditional visual cues that are integrated in detection approaches. These cues are combined in other systems with audio and speech features for more advanced recognition levels, trying to match or synchronize the electronic slides to the corresponding presentation videos [2], [7]. By this way, the browsing and indexation of educational and corporate digital video libraries are enhanced. At technical level, some algorithms have been developed on uncompressed video [7] document using the bit rate sequence, but unfortunately these ones appears to be time consuming. Some others have been operated on MPEG [3], [4], [5], [8] compressed domain using the DCT coefficients. In general, there are several technical approaches to compare a couple of consecutive frames in order to handle the shot boundary detection problem. These approaches are based on pixel-level frames comparison, histogram comparison (global block-based) or motion-based similarity measure. In

Manuscript received 19, 04, 2013; revised 8, 05, 2013.

Nour Charara is Member ofUniversity of Applied Sciences of Western Switzerland▪ Perolles 80, CH-1705 Fribourg, Switzerland and the Lebanese University, Beirut , Lebanon (nour.charara@edu.hefr.ch).

Iman Jarkass is with LSI Department, Institute of Technology IUT, Lebanese University, Pb. 813, Saïda, Lebanon. (ijarkass@ul.edu.lb).

Maria Sokhn works on Business Information Systems at the University of Applied Sciences Western Switzerland, Sierre (maria.sokhn@hevs.ch).

Elena Mugellini and Omar Abou Khaled are with the ICT Department, University of Applied Sciences of Western Switzerland, Perolles 80, CH-1705 Fribourg, Switzerland (name.surname@hefr.ch).

Digital Object Identifier 10.3969/j.issn.1674-862X.201.0?.0??

[4]Ngo *et al.* adopted background (conference hall elements including the slide) vs. foreground (presenter) algorithm to remove the effect of motion and occlusion when detecting slide changes. Another approach that is suitable for very large video databases but relatively sensitive to the noisy moving objects is presented in[1],where Oh et al. used the background tracking to take into account both the differences at the pixel and the semantics levels between the video frames.

Our proposed system, shown in Figure.1, is distributed into two separate stages. The first stage takes a still video frame as input and provides the spatial information of screen. The second stage benefits from this information to detect the occurred screen changes by processing only the cropped frames from the original video sequence.

This paper is organized as follows. In the next section, we introduce a segmentation-based screen tracking method using the dynamic rectangular extension strategy. Section 3 describes change detection by exploring the edge detection approach, in both slide projecting and movie scene. Section 4 presents the experimental results illustrating the performance of the proposed method. Finally, the last section concludes the paper and presents the future work.

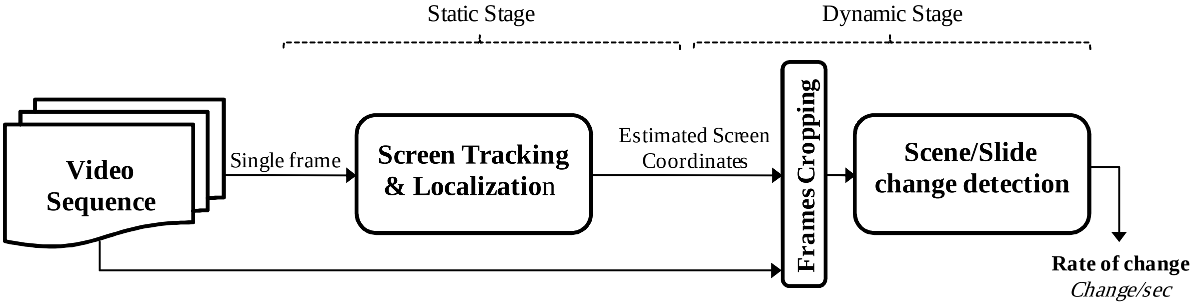


Fig. 1. Overall screen tracking and rate of change detection system.

2. Screen Tracking by Scene Segmentation

For automatic screen location and without intervention from a human operator or a priori information, the scene distribution of the different zones of any multipurpose hall is captured basing on a dedicated segmentation strategy (Fig. 3). By definition, scene segmentation process enables partitioning and labeling of the existing homogeneous regions. What is interesting here is the semantic scene segmentation that aims to classify fixed images into semantic classes. In this part, as well as the visual cues exploration from the video frames, the proposed system is also based on depth information. The depth map is used to increase reliability in the comprehension of the scene elements.

The multipurpose hall, the case-study, is divided normally into three defined zones, first the spectator seat area, then more deeply the sidewalk, and last the theater with the wide screen. Supposing that the image scene covers most of the hall components in a single frame as shown in Fig. 2, we proceed from the deepest point, so-called the vanishing point (VP), toward the camera focal plane. The VP is the common point of intersection of the converging lines in plane image that is captured by camera. At the limits of the homogenous area that surrounds this point, the wide hall screen is determined. Centered by the VP, which is detected with a 2D histogramming based technique [11], the initial rectangular patch is bordered to start the extraction process totransfer the features from fact domain to perception domain. For a given$M$ × N grey framed multipurpose hall scene I, each point $P_{i,j} \in I$, i =1, 2, ···, M and j= 1, 2, ..., N should belong a rectangular patch R_t that is centered by a C_t (X_{Ct}, Y_{Ct}) point with a width W_t and a height H_t. At each iteration, all these variables vary systematically following several criteria that provide dynamic, homogeneous and convergent rectangular patch extension. Here, the extension dynamicity is represented by the width of both horizontal and vertical steps that vary with respect to statistical features. The computation of the median and variance of the difference between the current patch and the next candidate extended patch at each iteration avoids an over-step to the next zone. By this way redundancy is prevented by saving iterations and by respecting the relative homogeneity of the patch. The property of intersection between the rectangular patch corners and the converging lines will guarantee rigorously the convergence property.

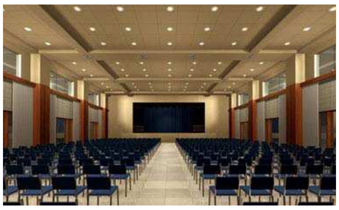


Fig. 2. A typical multipurpose hall model.

In order to characterize the semantic content of the scene, the low-level feature vectors are extracted from each described rectangular patch. The color feature has been neglected due the weak representation capacity in such scenes, which are relatively sensitive to local variation of illumination, and strongly depend on the design (decor) of the whole multipurpose hall. On the other hand, Texture is a spatially repetitive micro-structure of surfaces formed by repeating a particular element or several elements in different relative spatial positions. This repetition involves local variations of scale, orientation, or other geometric and optical features of the elements. The Gray Level Co-occurrence Matrix (GLCM) method extracts second order statistical texture features to be used in our feature vector building. Let G be the number of gray levels which exists in the rectangular patch R_t. The GLCM G×G matrix M is computed where each element $M_{L,\,ɵ}\,(m,\ n)$ is the co-occurrence number of each two pixels having gray values m and n separated by a displacement distance L and at a given direction ɵ. Then, some of Haralick texture features [12]as Inverse Difference Moment, Cluster Shade, Sum average, Dissimilarity, Sum Variance and Max Probability are computed and assigned to the feature vectors. The main motivation behind this segmentation strategy is resumed on constructing a discriminative pattern for the screen to distinguishing it from the other hall elements. This classification problem is ensured by a non-generative similarity based classifier.

Our supervised learning classifier estimates the class label of a test sample based on the pairwise similarities of data samples. The problem of classifying samples based only on their pairwise similarities may be divided into two sub-problems: measuring the similarity between samples and classifying the samples based on their pairwise similarities [13]. While the most similarity based method like as the k-nearest neighbors (k-NN) classifier, the quadratic discriminant analysis (QDA) and the support vector machines (SVMs) introduce the dissimilarity (distance) between their feature vector representations, we use a pure similarity measure, the zero-mean normalized cross-correlation (ZmNCC). The ZmNCC is widely exploited in template matching issues due to its robustness when subtracting the local mean[14]. Moreover, by employing the zero-mean normalized cross-correlation based classifier, we get rid from the description of the samples themselves. So, only a pairwise similarity exploration is required.

Let I be the framed image of size $w{\times}h$, and R_t of size $m_t{\times}n_t$ is the elaborated rectangular patch at each iteration t, where $m_t{<}w$-1 and $n_t{<}(h/2)$-1. The GLCM feature vector $F_t = \{f_{i,t}/\ i{=}0,\ 1\ \ldots N\}$ is extracted from the mentioned rectangular patch with N number of the quantified Haralick features. The problem is to estimate the class label z for a test sample (R_t, F_t) based on its similarities to the previous samples(R_{t-1}, F_{t-1}). The Zero mean Normalized Cross-Correlation between the F_t *and* $F_{t-1}$is shown in equation (1).

$$ZmNCC\,\langle R_{t-1}, R_t\rangle = \frac{\sum_{i=0}^{N}\big(f_{i,t-1}-\mu_{t-1}\big)\,.\,\big(f_{i,t}-\mu_t\big)}{\sqrt{\sum_{i=0}^{N}\big[f_{i,t-1}-\mu_{t-1}\big]^2}\cdot\sqrt{\sum_{i=0}^{N}\big[f_{i,t}-\mu_t\big]^2}} \qquad (1)$$

whereμ_{t-1} and μ_t represent the mean value of F_{t-1} and F_t respectively.

A couple of *Patch* with ZmNCC value over some specific threshold is then classified within the same zone z. A patch R_{tz} is considered at the border if it has a ZmNCC ‹R_{t-1},R_t› value lower than the specific threshold, which is in fact the separating metric that specifies staying in the same homogenous zone (the screen) or moving to the adjacent zone. In order to avoid any estimation error caused by the texture based pattern recognition technique, our proposed system employs depth information as constraint rules at the final stage taking into consideration that the screen is normally located at the deepest area of the hall (from the camera focal plane). To find an approximate depth map, we relied on a passive depth computation method using a stereo vision system [15]. Two calibrated cameras with known physical relation were correlated. The correspondences of common pixel values (from left and right image) were found and by triangulation [16] the distance between the correspondence areas was calculated.

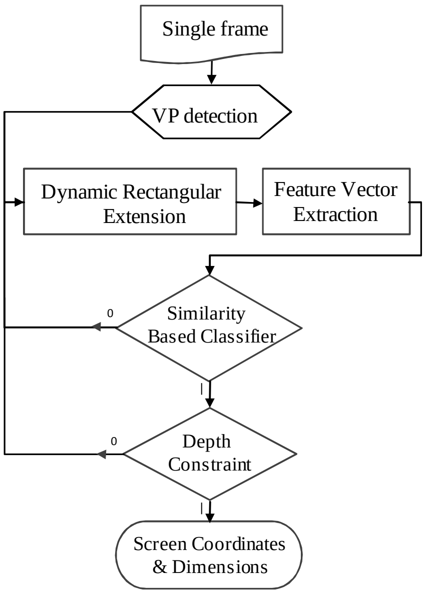

Single frame
VP detection
Dynamic Rectangular Extension
Feature Vector Extraction
0
Similarity Based Classifier
1
0
Depth Constraint
1
Screen Coordinates & Dimensions

Fig. 3.Schematic diagram of the proposed segmentation based screen tracking method.

3. Change Detection

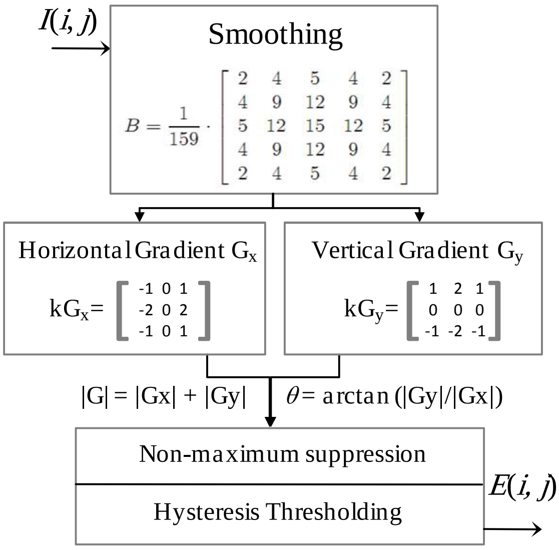
Till this end, the location of the wide screen is extracted from a unique still image. We benefit from the immobility of the equipped camera to crop the screen area from all the video frames. The cropped frame will be analyzed by the screen change detector to obtain the screen rate of change.

Our electronic screen change detection method is generally employed to detect major changes in video streams. We try here to adapt the edge detection based approach to overcome the small changes sensitivity problem. It is important to notice that when we define the different types of possible events in the scene, we don't aim to introduce modeling of all these types as done in [2], but we just try to differentiate them. This means that understanding and describing each of these events would lead to filter the desirable events and neglect the false ones. In general, a wide screen in a multipurpose hall is used by two ways: (1) as a slideshow screen for conferences and seminars or (2) as a movie show screen. In the first use, changes include transitions between the electronic slides in addition to the transition and animation effects. These effects play a negative granularity role and may scatter the essential information by augmenting the total detected changes. In the movie show context, significant changes are resumed by the boundary shots and the key-frames. A key-frame represents the visual content of frames that are the closest in a video sequence according to certain discriminative information, while a boundary shot is the gap between two shots. The detection algorithm must ignore the conflict changes such as camera motion (as tilt, pan and zoom), the appeared human gesture and the gradual transitions between scenes (dissolving, wiping or fading transitions)

In order to encounter this multiplicity challenge that rises from the multipurpose capacity, and to overcome problems concerning time-varying illumination and the small moving object problems, our proposed method first submits each video frame to an edge detector, and then, using a technique inspired from the MPEG-7 Edge Histogram Descriptor, finds features characterizing a scene/slide change detection.

Edge detection principle is frequently used in image processing fields to only preserve the structural representation of an image. A standard Canny edge extractor [17] is exploited in this work by applying the following steps (see Figure. 4): First, the cropped frame is smoothed and blurred to remove the noise. Then the local maxima are selected after computing the image gradients. At last a traditional thresholding step is applied to determine the approved edges by hysteresis.

We aim to select the discriminative features that characterize a scene/slide change, in another term we should define an adaptive dissimilarity measure. In the literature, there are two approaches to solve this problem: histogram-based and pixel-based.



I(i, j)
Smoothing
$B = \frac{1}{159} \cdot \begin{bmatrix} 2 & 4 & 5 & 4 & 2 \\ 4 & 9 & 12 & 9 & 4 \\ 5 & 12 & 15 & 12 & 5 \\ 4 & 9 & 12 & 9 & 4 \\ 2 & 4 & 5 & 4 & 2 \end{bmatrix}$
Horizontal Gradient G_x
$kG_x= \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$
Vertical Gradient G_y
$kG_y= \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$
|G| = |Gx| + |Gy|
θ= arctan (|Gy|/|Gx|)
Non-maximum suppression
Hysteresis Thresholding
E(i, j)

Fig. 4. Canny edge detector steps, where *I* is the original image, *E is* the edge image. kG_x and kG_y are the horizontal and vertical Kernel respectively. |G| and θ are the gradient magnitude and phase respectively.

Despite its low computation cost, the histogram based approach neglects the spatial information and so it may ignore a significant spatial change between two consecutive frames. The pixel based method suffers from the high sensitivity toward the slightest motion. So, a tradeoff method is proposed to optimally adapt our application, e.g. we substitute the basic unit of the second approach (pixel) by sub-image to extract the edge histograms. An edge histogram represents the distribution of the different edge directionalities existed in the image. In particular, the cropped frame is segmented into 4×4 sub-images. Each sub-image is represented by a 5-bins edge distribution histogram where each bin survey one type of the five directional edges types: horizontal, vertical, diagonal-1 (+45 degree), diagonal-2 (-45 degree) and non-directional edge. We exclusively care of the global edge direction composition in sub-images, disregarding the absolute location of edges.

In order to compare the edge histograms of the corresponding sub-images in two successive frames f_i and f_{i+1}, we choose the following normalized dissimilarity measure:

$$d(f_i, f_{i+1}) = \sum_{j=1}^{j=\mathrm{N}} \sum_{k=1}^{k=K} \left[\frac{\left(H_{S^j_{f_i}}(k) - H_{S^j_{f_{i+1}}}(k) \right)^2}{max\left(H_{S^j_{f_i}}(k) , H_{S^j_{f_{i+1}}}(k) \right)} \right] \qquad (2)$$

where $H_{s_{f_i}^j}(k)$ is the bin value of the sub-image $S_{f_i}^j$ histogram. N is the overall number of sub-images, in our case is 16. K is the total number of direction bins; here it is equal to 5. A scene/slide change detection is inferred when the dissimilarity measure exceeds a specific threshold. Therefore, the temporal rate of change is calculated.

4. Experiments

The experiments are conducted on two real world video sets both with the same resolution 160×120; the first is composed of 8 lecture and conference videos from university lectures and international conference oral presentations (AWIC 2011).These conference presentations don't contain videotaping shows on the screen and if there are they are very short. The second dataset consists of several films recorded during a documentary video show. Due to the difference on evaluation method types, the two system stages are tested separately. For the first stage algorithm (Fig. 5), an expert based evaluation method has been adopted, so that the screen is manually segmented at the original fixed scene frame to be compared on pixel-level with our results.

A quantitative evaluation of the screen tracking part is presented by the normalized root-mean-square error (N-RMSE) and the recognition rate (RR) calculation, to measure the differences between the manual segmented screen zone (as a reference) and those obtained automatically by our segmentation based method. N-RMSE is a function dimensions and edge pixels intensity and other statistical features of the detected screen zone(Table 1). Note that the proposed method provides much more accurate results with making consideration of the depth information.

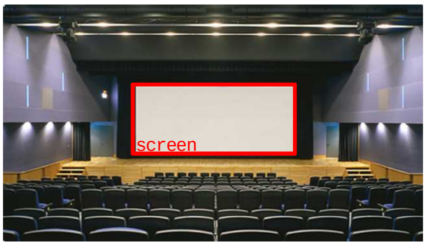


Fig. 5. Resulting tracked screen.

Table 1:N-RMSEand Meanof the Recognition Rate (RR)of the screen tracking stage

Screen tracking method Stages	N-RMSE	Mean of RR
without depth constraints	0.29	88.95 ± 5.56
with depth constraints	0.311	91.76 ± 5.97

To evaluate the second part, a ground truth matching between existed changes (slide or scene change according to dataset) and the automatic detected changes is manually constructed. We use the well-known performance parameters used in the most scene change detection methods; recall (3) and precision (4).

$$Recall = n_{TP}/(n_{TP} + n_{FP}) \quad (3)$$

$$Precision = n_{TP}/(n_{TP} + n_{FN}) \quad (4)$$

wheren_{TP} is the number of true positive (correct detection), n_{FP} is the number of false positive (false detection) andn_{FN} is the number of false negative (miss detection).

A sequence of simultaneous detected slides (and therefore detected changes) in a cropped screen are shown in Fig. 7. The negligence of the small variation such as animation and transition effects (false positives) is shown in the figure. Indeed, the system just marks the slide flipping as a significant change, which is the required result.

Table 2: Averaged Scene/Slide change detection evaluation

	Rate of change (changes/sec)	Recall*100	Precision*100
Conference videos	0.12	79.6%	89.2%
Movie show videos	0.655	78.7%	83%

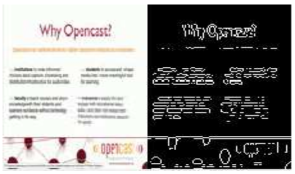


Fig. 6.Example of edge detection, Left: Original cropped frame, Right: the edge cropped frame.

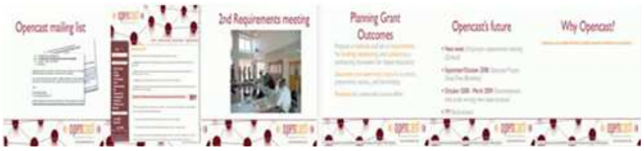


Fig.7. A sequence of simultaneous detected slides in a cropped screen.

5. Conclusion

We have proposed a novel method that locates a screen, in a multipurpose hall, and computes its rate of change. Despite that we distinguish between the definition of Slide change and Scene change, this approach is able to detect both types. The system is composed of two stages, first a segmentation based method tracks the screen borders and then, we benefit from the edge histogram efficiency to map a change detection method. Experiments show that the proposed method succeeds in slide/scene change detection. As future works, additional features have to be integrated on the initial steps to make our change detection system more robust against the suddenly screen occlusions. Moreover, exploiting dynamic thresholding can increase the system accuracy.

6. References

[1] J. Oh, K. A. Hua, and N. Liang. A content-based scene change detection and classification technique using background tracking. In SPIE Conf. on Multimedia Computing and Networking, San Jose, CA, January 2000.

[2]Q. Fan, A. Amir, K. Barnard, R. Swaminathan and A. Efrat, “Temporal modeling of slide change in presentation videos”, *IEEE International Conference on Acoustics, Speech and Signal Processing*(ICASSP 2007), Honolulu, Hawaii, USA, April 15-20, IEEE 2007.

[3] C. Taskiran and E.J. Delp, “video scene change detection using the generalized sequence trace”,in*Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Washington, USA, May 12-15, 1998.

[4] C.W. Ngo, T.C. Pong and T.S. Huang, “Detection of slide transition for topic indexing,”in Proc. of 2002 IEEE International Conference on Multimedia and Expo(ICME ’02), Lausanne, Switzerland, August 26-29, 2002.

[5]A. Cavallaro and T. Ebrahimi. “Change detection based on color edges”. IEEE International Symposium on Circuits and Systems (ISCAS 2001), Australia, May 6-9, 2001.

[6] D. Ma and G. Agam, “Lecture video segmentation and indexing,” in *Proc. SPIE 8297, Document Recognition and Retrieval XIX*, 82970V, January 23, 2012.

[7]G. Schroth, N.-M. Cheung, E. Steinbach and B. Girod. “Synchronization of presentation slides and lecture videos using bit rate sequences”. 18th IEEE International Conference on Image Processing (ICIP 2011), Brussels, Belgium, September 11-14, 2011.

[8] C.-L. Huang and B.-Y.Liao, “A robust scene-change detection method for video segmentation,” IEEE Trans. on Circuits and Systems for Video Technology 11(12), pp. 1281- 1288, 2001.

[9] S. G. Deshpande& J.-N. Hwang, “A Real-time Interactive Virtual Classroom Multimedia Distance Learning System,” *IEEETrans on Multimedia*, vol. 3, no. 4, pp. 432-444, Dec 2001.

[10] Maria Sokhn, Elena Mugellini, Omar Abou Khaled, Ahmed Serhrouchni, "End-to-end adaptive framework for multimedia information retrieval", 9th International Conference on Wired/Wireless Internet Communications (WWIC 2011), 13-17 June, 2011, Vilanovai la Geltru, Catalonia, Spain.

[11] Bo Li, KunPeng, Xianghua Ying, HongbinZha, VanishingPoint Detection Using Cascaded 1D HoughTransform from Single Images, *Pattern RecognitionLetters, Vol 33, Iss 1, pp. 1-102,* 2012.

[12] Fritz Albregtsen, “Statistical Texture Measures Computed from Gray Level Co-occurrence Matrices”, Image Processing Laboratory, Department of Informatics, University of Oslo, November 5, 2008.

[13] L. Cazzanti, “Generative Models for Similarity-based Classification,” Ph.D. dissertation, Dept. Elect. Eng., university ofWashington, 2007.

[14]L.D. Stefano, S. Mattoccia and F. Tombari. “ZNCC-based template matching using bounded partial correlation”. *Pattern Recognition Letters*, Volume 26, Issue 14, 15 October 2005, Pages 2129–2134.

[15]G. Balakrishnan, G. Sainarayanan, R. Nagarajan, and SazaliYaacob, “A Stereo Image Processing System for Visually Impaired,” *International Journal of Information and Communication Engineering* 2:3, 2006.

[16] *S.Mattoccia, “Stereo vision: algorithms and applications, Università di Firenze”*, May 2012.

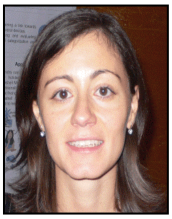
[17] *Canny, J., A Computational Approach To Edge Detection, IEEE Transactions on Pattern Analysis and Machine Intelligence*,8:679-714,1986.

Nour A. Charara received her License and Master1 degrees in Electronics from the Lebanese University Faculty of Sciences in 2008 and 2009 respectively. Also she received Research Master degree in Signal, Telecom, Image, and Speech (STIP) from the Lebanese University, Doctoral School of Sciences and Technology in 2010. Currently, she is member of the Multimedia Information System Group, MISG at the University of Applied Sciences of Western Switzerland, Fribourg, Switzerland. Her research interests include video analysis, image processing and behavior recognition.

Iman Jarkass was born in Tripoli (Lebanon), 23 May 1970. PhD in Control of Systems at the University of Technology of Compiegne (UTC), France, 1998. She is an associate professor at the Lebanese University (Lebanon). She is interested in the data fusion, especially applied to dynamic systems.

Maria Sokhn holds a PhD from the Telecom Paris Tech in 2011. She graduated from the Engineering Faculty of Saint Joseph University, with a Computer and Communication Engineer diploma, and holds a specialized master diploma from Telecom Paris Tech in Multimedia production and creation. Actually she is a professor at the University of Applied Sciences of Sierre, and member of the Multimedia Information System Group MISG. Her researches interest focuses on eGovernement, multimedia information retrieval.

Omar Abou Khaled holds a PhD in Computer Science received from the University of Technology of Compiègne, and a Master in Computer Science from the same university. He is Professor at the University of Applied Sciences of Fribourg EIA-FR. Since 1996 he has been working as research assistant in the MEDIA group of the Theoretical Computer Science Laboratory of EPFL (Ecole Polytechnique Fédérale de Lausanne) in the field of Educational Technologies and Web Based Training research field.He is International Advisor at HES-SO. He is responsible of several projects in the field of Document Engineering, Multimodal Interfaces, Context Awareness, Ambient Intelligence, Blended Learning, and Content-Based Multimedia Retrieval.



Elena Mugellini holds a PhD in Telematics and Information Society received from the University of Florence in 2006, and a Master in Telecommunication Engineering from the same university received in 2002. She is Professor at the Information and Communication Department of the University of Applied Sciences of Western Switzerland, Fribourg (EIA-FR). Elena is the leader of MISG research group. She is also member of the Telematics Technology Laboratory at the University of Florence. Her current research interests are on the areas of Ambient Intelligent, Multimodal Interaction, Tangible User Interface, Personal Information Management, and Document Engineering