한국지능시스템학회 논문지 2008, Vol. 18, No. 1, pp. 133-138

A New Anchor Shot Detection System for News Video Indexing

Hansung Lee*, Younghee Im*, Jooyoung Park**, and Daihee Park*

* Dept. of Computer and Information Science, Korea University
** Dept. of Control and Instrumental Engineering, Korea University

Abstract

In this paper, we propose a novel anchor shot detection system, named to MASD (Multi-phase Anchor Shot Detection), which is a core step of the preprocessing process for the news video analysis. The proposed system is composed of four modules and operates sequentially: 1) skin color detection module for reducing the candidate face regions; 2) face detection module for finding the key-frames with a facial data; 3) vector representation module for the key-frame images using a non-negative matrix factorization; 4) one class SVM module for determining the anchor shots using a support vector data description. Besides the qualitative analysis, our experiments validate that the proposed system shows not only the comparable accuracy to the recently developed methods, but also more faster detection rate than those of others.

Key Words: anchor shot detection, face detection, NMF, SVM, SVDD.

1. Introduction

Nowadays, video materials and video services have been more available than ever before in terms of information repositories as well as commercial perspectives. In particular, with the growing popularity of digital news video as multimedia information repositories, collections of the news video database are recently exploded in numbers. As a result, the news videos appeared to be very valuable information for the data analysts, information providers, and TV consumers because of their information richness [1]. Therefore, news video databases have been a subject of extensive research in the past decade to develop tools for effective and efficient manipulations and analysis of news videos.

In this paper, the scope of our concerns is focused on the anchor shot detection module in particular which is depicted with a highlight in Figure 1. Amongst the preprocessing processes of news video analysis, an anchor shot detection plays a very important role in news video story partitioning especially. Concerning the recent literatures on anchor shot detection, there are in general two main research paradigms [2-3]: 1) template matching method and 2) unsupervised method. The former defines a model in advance for an anchor shot and matches it against all the shots of a news video. Since this method has been proven to be too sensitive to conditions of studio as well as positions of anchor person, it turned out to be not appropriate in the community.

접수일자 : 2007년 11월 13일
완료일자 : 2007년 12월 20일
This paper has been selected as an excellent paper in KFIS Autumn Conference 2007

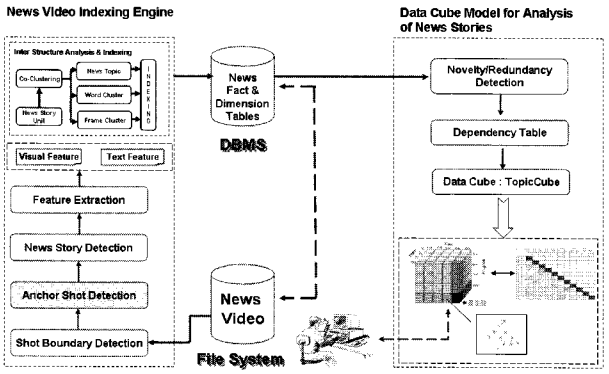


Fig. 1: The Overall Architecture of Our System: focusing on the anchor shot detection

On the other hand, the latter aggregates shots with similar visual contents repeatedly occurred in the whole news video [2-3]. However, by criticizing their work constructively, this method has some serious inevitable drawbacks which states that 1) it consists of two phases (i.e., clustering and pruning) requiring computational cost too much and 2) it uses the assumption that most anchor shots have similar color histograms. However, it is not the case any more according to the recent development of computer graphic technology. The real truth is that background color histograms are often varying. Besides, there are some other ongoing attempts to use audio-visual features for anchor shot detection [4-5]. However these methods still remain time-consuming from the practical point of view.

Since news video produces every day with a greate volume, it is necessary to devise a fast and accurate algorithm for anchor shot detection, needless to say. Essentially, our primary concern in this paper is to construct a novel anchor shot detection system aiming at

the fast detection rate and high precision rate as well. To meet our requirements, in this paper, we propose a novel anchor shot detection system, named to MASD (Multi-phase Anchor Shot Detection), which consists of four modules and operates sequentially: 1) skin color detection module for reducing candidate face regions; 2) face detection module for finding key-frames with a facial data; 3) vector representation module for key-frames using a non-negative matrix factorization (NMF); 4) one class support vector machine (SVM) module for determining the anchor shots using a support vector data description (SVDD).

The rest of this paper is organized as follows. In Section 2, we describe some well-known characteristics of anchor shots. A fast and accurate anchor shot detection system, named to MASD, is proposed in Section 3. In Section 4, we perform some computer experiments to clarify the validity of our approach. Finally, in Section 5, concluding remarks are given.

2. Well-Known Characteristics of Anchor Shots

The temporal and spatial features of anchor shots in a news video are characterized as follows [2]: 1) The news always starts and ends with an anchor shot, and the anchor person appears several times in a news video; 2) The anchor shot is a static shot (i.e., shots containing less changes of camera and object motions); 3) The anchor shot is visually similar with each other and there is no or little changes in anchor's dress and in the background of broadcasting studio as well; 4) The length of anchor shot is longer than normal shots in a news video; 5) The anchor shot consists of at most two persons; 6) The anchor appears with a frontal face; 7) The anchor shot has one of the following spatial types (see Figure 2): a) one anchor on the right, b) one anchor on the left, c) one anchor in the middle, and d) two anchors.

Although the above heuristic rules are not always in accordance with real situations (in fact, it differs slightly from stations to stations), its findings are very valuable in constructing our system in particular. By virtue of these rules, anchor shots can be clustered by the visual similarity and then be classified into one of the spatial types illustrated in Figure 2.

3. New Anchor Shot Detection System

The overall architecture of our newly proposed anchor shot detection (ASD) system is given in Figure 3, named to MASD (Multi-phase Anchor Shot Detection), which consists of four modules and operates sequentially: 1) skin color detection module for reducing candidate face regions; 2) face detection module for finding key-frames with a facial data; 3) vector representation module for key-frame using a non-negative matrix factorization (NMF); 4) one class SVM module for determining the anchor shots using a support vector data description (SVDD). Each module will be described one by one in great details.

Fig. 2. Examples of Four Types of Anchor Shots

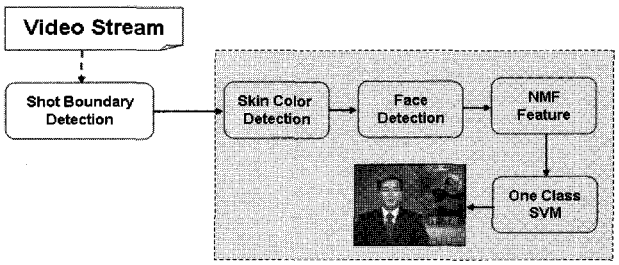


Fig. 3. The Overall Architecture of ASD System

3.1 Skin Color Detection Module

To increase the possibility of anchor shots in a candidate set, and reduce the search space for the face detection module in the next stage, we adopt a fast skin color detection algorithm [6] which eliminates key-frames of each shot with no region of skin color and/or too large region. The simple but cost-effective heuristic rule defined in (1) is highly appropriate here. Consequently, the reduced images with a skin color region are obtained with a fast speed.

$$\begin{aligned}&(R, G, B) \text{ is classified as skin if :}\\ &R > 95 \text{ and } G > 40 \text{ and } B > 20 \text{ and}\\ &R - G > 15 \text{ and } R > B\end{aligned} \tag{1}$$

3.2 Real-Time Face Detection Module

With resulting reduced key-frames produced at the prior step, the main job of a real-time face detection module is to classify them into one with a facial data and the other with non-face data. In case of a

key-frame without a face or with over 3-faces, it is automatically discarded. Accordingly, the search space for the next stage will be considerably reduced. To meet our design requirements (i.e., speed and precision), we choose a real-time face detection model [6] which consists of several weak classifiers in a chain of cascading structure and a support vector machine (SVM) in the last row. A series of weak classifiers incrementally generate the candidate faces with a high speed but low precision. Since the false-negative ratio is high but the false-positive ratio is close to zero, they are conclusively able to generate the candidate faces with a low computational cost. On the other hand, a SVM in the last row detects a face with a high precision finally. The overall architecture of a real-time face detection module is given in Figure 4.

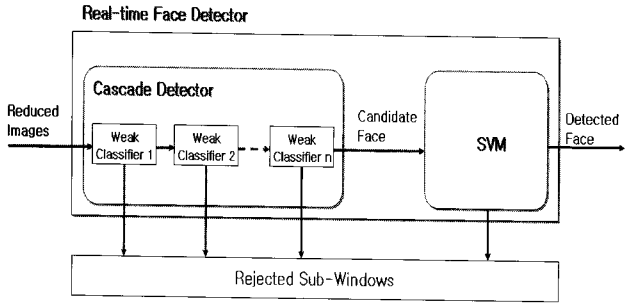


Fig. 4. The Overall Architecture of Real-Time Face Detection System

3.3 Vector Representation Module

As for the vector representation of key-frames, we select a non-negative matrix factorization (NMF) technique, since it has some favourable features comparing to other linear algebra techniques such as a singular value decomposition (SVD) and a principal component analysis (PCA). More precisely, in the induced space derived by a NMF, each axis captures more rich structural features of each shot clusters, and each image vector is conveniently represented with an additive combination of axes. In addition, each axis in the reduced space has more straightforward correspondence with each image cluster than that in the space derived by SVD and PCA [7].

Subject to positive constraints, the NMF factoring of a given positive matrix is derived as follows [8-9]: The image corpus is regarded as an $n \times m$ matrix V, where each column contains n non-negative pixel values of one of the m images. Given matrix V, the NMF is defined as a problem to find non-negative matrix factors W and H such that:

$$V \approx WH \tag{2}$$

where $W=[w_{ij}]$, $H=[h_{ij}]$, $0 \le i \le n$, $0 \le j \le m$.

It is then formalized as the following mathematical optimization problem.

$$\begin{aligned} \min \quad & J=\frac{1}{2} \| V-WH \|^2 \\ s.t. \quad & W,H \ge 0. \end{aligned} \tag{3}$$

Using the Lagrange function and the Kuhn-Tucker condition, the following equations for w_{ij} and h_{ij} are given:

$$(VH)_{ij} w_{ij} - (WH^T H)_{ij} w_{ij} = 0 \tag{4}$$

$$(V^T W)_{ij} h_{ij} - (HW^T W)_{ij} h_{ij} = 0 \tag{5}$$

These equations lead to the following update formulas:

$$w_{ij} \leftarrow w_{ij} \frac{(VH)_{ij}}{(WH^T H)_{ij}} \tag{6}$$

$$h_{ij} \leftarrow h_{ij} \frac{(V^T W)_{ij}}{(HW^T W)_{ij}} \tag{7}$$

For the vector representation of image, NMF is usually defined as follows;

$$V_{i\mu} \approx (WH)_{i\mu} = \sum_{a=1}^{r} W_{ia} H_{a\mu} \tag{8}$$

The r columns of W are called basis. Each column of H is called an encoding and is in one-to-one correspondence with a image in V. The rank r of factorization is generally chosen so that $(n+m)r < nm$, and the product WH can be regarded as a compressed form of the data in V.

3.4 One Class SVM Module

In the last stage of our system, SVM is used to classify the anchor shot with a high precision. In general, the number of data necessary for training varies depending on that of each anchor shot data and non-anchor shot data. Hence, the resulting training may have an influence by other class due to the unbalanced size of training data. Accordingly, it is preferable to select a decision boundary function using one-class SVM (OSVM) (one of the most well-known OSVM is a support vector data description (SVDD)). A SVDD is described as follows [10]:

Given a dataset of N-patterns in d-dimensional input space, $D=\{x_i \in \mathbb{R}^d \mid i=1,...,N\}$, one-class SVM is defined as a problem to obtain a sphere minimizing the volume of it and at the same time including the training data as many as possible It is formalized as the following mathematical optimization problem:

$$\begin{aligned} \min \quad & L_0(R^2,a,\xi) = R^2 + C\sum_{i=1}^{N} \xi_i \\ s.t. \quad & \| x_i - a \|^2 \le R^2 + \xi_i,\ \xi_i \ge 0,\ \forall i. \end{aligned} \tag{9}$$

where a is the center of the sphere that expresses a class, R^2 is the square value of sphere radius, ξ_i is the penalty term that shows how far i th training data x_i is

deviated from the sphere, and C is the trade-off constant. By introducing a Lagrange function and saddle point condition, we obtain the following dual problem:

$$\min_{\alpha} \sum_{i=1}^{N} \sum_{i=1}^{N} \alpha_i \alpha_j < x_i, x_j > - \sum_{i=1}^{N} \alpha_i < x_i, x_i > \quad s.t. \quad \sum_{i=1}^{N} \alpha_i = 1, \alpha_i \in [0, C], \ \forall_i \tag{10}$$

A sphere can express more complex decision boundary in the feature space F and we can map an input space into a feature space using kernel function K. When the Gaussian function is chosen for the kernel, the problem can be further simplified as follows [11]:

$$\min_{\alpha} \sum_{i=1}^{N} \sum_{i=1}^{N} \alpha_i \alpha_j K(x_i, x_j) \quad s.t. \quad \sum_{i=1}^{N} \alpha_i = 1, \alpha_i \in [0, C], \forall_i \tag{11}$$

Note that in this case, the decision function of each class can be summarized as follows

$$f(x) = R^2 - (1 - 2\sum_{i=1}^{N} \alpha_i K(x_i, x) + \sum_{i=1}^{N} \sum_{j=1}^{N} \alpha_i \alpha_j K(x_i, x_j)) \geq 0 \tag{12}$$

4. Discussions and Experiments

Now, by putting everything together stated so far, we have implemented our new anchor shot detection system, named to MASD by means of NMF toolbox [12] and SVDD toolbox [13] for MATLAB. Figure 5 illustrates the processing snapshot of our system phase by phase with a typical example.

To evaluate the effectiveness of a proposed anchor shot detection system overall, we collect a dataset from two main Korean broadcasting stations, namely, KBS and MBC. The test dataset consists of 1,226 shots including 55 real anchor shots, 43 reporter shots, and 70 interview shots. A dataset is described in more details in Table 1. The ground truth is manually labeled in advance.

The standard *recall* and *precision* criteria for the evaluation measure is adopted here, where *recall* is the ratio of the number of relevant items retrieved to the total number of relevant items in the relevant set, while *precision* is the ratio of the number of relevant items retrieved to the total number of irrelevant and relevant items retrieved. These are expressed as a percentage as follows:

$$Recall = \frac{A}{A+B} \times 100(\%) \tag{13}$$

$$Precision = \frac{A}{A+C} \times 100(\%) \tag{14}$$

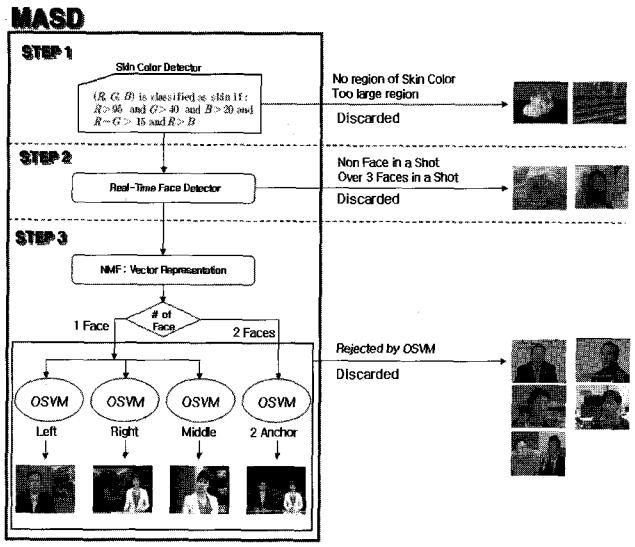


Fig. 5. The MASD With a Typical Example

Table 1. Dataset for Evaluation

	# of Shots
Total Shot	1,226
Anchor Shot	55
Reporter	43
Interview	70

where A is the number of relevant items which are retrieved by the system, B is the number of relevant items which are not retrieved by the system, and C is the number of irrelevant items which are retrieved by the system. As a single figure of merit for comparing different algorithms, the so-called F-measure [3] combining *recall* and *precision* is also used:

$$F = \frac{2 \times recall \times precision}{(recall + precision)} \tag{15}$$

Our computer experiment reports that 399 (32.5 %) key-frames as the candidate set is obtained without losing any anchor shots after a skin color detection phase. Furthermore, it also reports that 148 (12.1 %) key-frames as the candidate set is determined after a skin color detection phase with including all anchor shots (see Figure 6). After all, it is confirmed from the empirical observation that a huge amount of non-candidate anchor shots is eliminated successfully by means of the successive processes of a skin color detection and a face detection in a row as we expected.

With the resulting reduced search space produced at the prior steps, the anchor shot detection module classify the facial data into one with an anchor shot and the other with non anchor shot. Corresponding to experiment, our system shows 98.14 % of *precision*, 96.8 % of *recall*, and 97.2 % of F-measure, respectively. Roughly comparing ours with the results of other previous methods

[1-5] (In fact, it is not advisable. Since each method uses its own dataset, the comparing empirical result is not reliable in a sense), our system shows the comparable accuracy to the recently developed methods. The summary is given in Table 2.

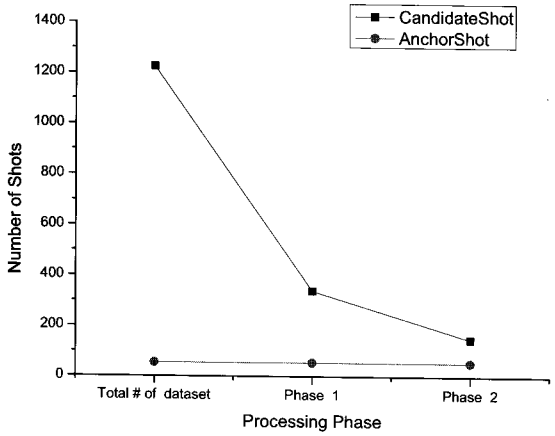


Fig. 6. The # of candidate sets at each phase

As stated before, the methodology behind our system considers not only the high speed by means of reducing the search space extensively, but also the high accuracy by means of a SVDD with NMF. Now, we have seen that the computer experiments support our system's designing goals from the empirical point of views. Therefore, associated with a qualitative analysis, we are entitled to claim that our system is quite effective and efficient.

Table 2. Comparison with Other ASD Methods

	[1]	[2]	[3]	[4]	[5]	Proposed method
Precision	100	98.0	99.4	92.6	96.8	98.1
Recall	97.7	93.0	92.3	90.5	97.9	96.4
F-measure	98.8	95.4	95.7	91.5	97.3	97.2

5. Conclusions

In this paper, we proposed a new anchor shot detection system, named to *MASD* (*Multi-phase Anchor Shot Detection*), which is a core step of the preprocessing process for the news video analysis. A computer experiments was presented to illustrate the proposed system, and associated with a qualitative analysis, the resulting system validated by showing not only the comparable accuracy to the recently developed methods, but also more faster detection rate than those of others.

References

[1] G. Xinbo, L. Jie, and Y. Bing, "A Graph-Theorectical Clustering based Anchorperson Shot Detection for News Video Indexing", *Proc. of ICCIMA'03*, pp. 108-113, 2003.

[2] X. Luan, Y. Xie, L. Wu, J. Wen, and S. Lao, "AnchorClu: An Anchorperson Shot Detection Method Based on Clustering", *Proc. of PDCAT'05*, pp. 840-844, 2005.

[3] M. Santo, P. Foggia, C. Sansone, G. Percannella, and M. Vento, "An Unsupervised Algorithm for Anchor Shot Detection", *Proc. of ICPR'06*, Vol. 2, pp. 1238-1241, 2006.

[4] D. Lan, Y. Ma, and H. Zhang, "Multi-level Anchorperson Detection Using Multimodal Association", *Proc. of ICPR'04*, Vol. 3, pp. 890-893, 2004.

[5] L. D'Anna, G. Marrazzo, G. Percannella, C. Sansone, and M. Vento, "A Multi-stage Approach for Anchor Shot Detection", *LNCS*, Vol. 4109, pp. 773-782, 2006.

[6] J. Song, H. Lee, and D. Park, "Real-Time Face Detection System Using Cascade Structure and SVDD", *Proc. of KCC05*, Vol. 32, No. 1(B), pp. 763-765, 2005.

[7] W. Xu, X. Liu, and Y. Gong, "Document Clustering Based on Non-negative Matrix Factorization", *Proc. of ACM SIGIR03*, pp. 267-273, 2003.

[8] Daniel D. Lee and H. Sebastian Seung, "Learning the parts of objects by non-negative marix factorization", *Nature*, Vol. 401, pp. 788-791, 1999.

[9] Daniel D. Lee and H. Sebastian Seung, "Algorithms for Non-negative Matrix Factorization", *In Advances in Neural Information Processing Systems*, Vol. 13, pp. 556-562, 2001.

[10] D. Tax and R. Duin, "Uniform Object Generation for Optimizing One-class Classifiers", *Journal of Machine Learning Research*, Vol. 2, Issue 2, pp. 155-173, 2001.

[11] Jooyoung Park, Jinsung Kim, Hansung Lee, and Daihee Park, "One-Class Support Vector Learning and Linear Matrix Inequality", *International Journal of Fuzzy Logic and Intelligent Systems*, Vol.3, No.1, 2003.

[12] NMF Toolbox for MATLAB
Available at: http://mole.imm.dtu.dk/toolbox/nmf/

[13] SVDD Toolbox for MATLAB
Available at: http://ict.ewi.tudelft.nl/~davidt/dd_tools.html/

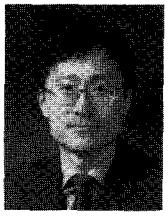
저 자 소 개

Hansung Lee
received his BS and MS degree in computer science from Korea University, Korea, in 1996 and 2002, respectively. He worked for DAEWOO engineering company from 1996 to 1999. He is now pursuing his PhD degree in the Department of Computer Science, Korea University, Korea. His research interests include machine learning and data mining.

E-mail : mohan@korea.ac.kr

Younghee Im
received her BS degree in computer science from Korea University, Korea, in 1994, and her PhD degree in computer science from Korea University, Korea, in 2001. She joined Korea University in 2005, where she is currently an Invitational Professor in the Department of Computer and Information Science. Her research interests include machine learning, context awareness, and intelligent database.

E-mail : yheeim@korea.ac.kr



Jooyoung Park
received his BS degree in electrical engineering from Seoul National University, Korea, in 1983, and his PhD degree in electrical and computer engineering from the University of Texas at Austin, USA, in 1992. He joined Korea University in 1993, where he is currently a Professor in the Department of Control and Instrumental Engineering. His research interests include neural networks and nonlinear systems.

E-mail : parkj@korea.ac.kr

Daihee Park
received his BS degree in mathematics from Korea University, Korea, in 1982, and his PhD degree in computer science from the Florida State University, USA, in 1992. He joined Korea University in 1993, where he is currently a Professor in the Department of Computer and Information Science. His research interests include data mining and intelligent database.

E-mail: dhprk@korea.ac.kr