

Automatic Classification and Skimming of Articles in a News Video Using Korean Closed-caption

Jung-Won Cho¹, Seung-Do Jeong¹, and Byung-Uk Choi²

¹ Multimedia Laboratory, Department of Electrical and Computer Engineering,
Hanyang University, 17 Haengdang-dong, Sungdong-gu, Seoul, 133-791 Korea
{bigcho, kain}@mlab.hanyang.ac.kr

² Division of Information and Communications,
Hanyang University, 17 Haengdang-dong, Sungdong-gu, Seoul, 133-791 Korea
buchoi@mlab.hanyang.ac.kr

Abstract. We are able to analyze the meaning in the news video by using various data, such as a moving picture, an audio, a caption, and so on. In order to browse the news video effectively, classification and skimming of the news articles are very essential. In this paper, we propose both the automatic classification and skimming methods of the news articles using the closed-caption. The automatic classification method uses tags in the closed-caption for the purpose of distinction of speakers. The skimming method extracts the part of the article introduced by an anchor in the closed-caption as a representative sentence, and also extracts the representative frames consisted of the anchor frame, open-caption frames, and frames synchronized with the frequently appeared terms.

1 Introduction

In order to obtain the precise retrieval results of queries, we have investigated indexing, storing, querying, and showing methods of the retrieval result [1][2]. The first thing we have to consider here is to analyze the user's demand and reflect it on indexing. The indexed news article is then able to be effectively retrieved by using the knowledge-based query. This paper analyzes the meaning of the news article by using useful information extracted from the closed-caption. Also, we present automatic classification and skimming methods of news articles using the closed-caption.

2 Automatic classification of articles

For effective browsing of the news articles, classification according to a category of the article is required. In this paper, for the automatic classification, we extract the reporter's name in the closed-caption and match the name with the records in the database having information of reporter's affiliation [3]. Consequently, the articles can be classified into 7 categories, i.e., politics, economy, society, unification, infor-