UNIVERSITE DE GENEVE

CENTRE UNIVERSITAIRE
D'INFORMATIQUE
COMPUTER VISION AND
MULTIMEDIA LABORATORY

Date: 14 November 2006
N° 06.11

TECHNICAL REPORT

Analysis of modality dependence/independence impact on the performance of multimodal fusion systems: Summary of main results

Computer Vision Group
Computing Science Center, University of Geneva
24 rue du Général Dufour, CH - 1211 Geneva 4, Switzerland

e-mail: {Oleksiy.Koval, svolos, Thierry.Pun}@cui.unige.ch

We consider multimodal fusion as a hypothesis testing problem. As an ultimate goal we consider the development of a unified information-theoretic framework for the optimal fusion of multimodal signals. Main tasks to be addressed include:

- information-theoretic investigation of dependent/independent modality fusion impact on performance of fusion systems;
- development of practical low-complexity multimodal fusion systems according to the obtained theoretical results;
- practical validation and demonstration of multimodal fusion systems for the following scenarios: multimodal person identification; multimodal document identification; navigation and document retrieval in multimedia databases using multimodal interaction.

Main setups under the analysis:

- multimodal person identification;
- multimodal interaction and database navigation via printed documents.

Basic application (toy example): for the sake of a fair comparison we have selected multimodal biometric person identification. However, other multimodal fusion systems can be considered using the same theoretical formulations.

Brief summary of main existing results:

- fusion is usually performed on the decision or score levels (see [1] and references therein, [5]);
- known theoretical findings are obtained assuming Gaussian statistics of the considered multimodal data [1,3,5];
- impact of modality vector length on the performance was not studied;
- general theoretical results justifying the influence of statistical dependency between modalities on the performance of multimodal fusion have not been investigated and published yet;
- theoretical justification of fusion system performance improvement due to modality correlation remains an open problem. Some papers even report opposite results. For example, in [3] the authors support the idea that fusion of correlated modalities does not always lead to the fusion performance improvement versus combining independent signals. Contrarily, it is demonstrated in [5] that taking into account correlation between the modalities one can obtain an improvement.

Therefore, the analysis of the existing results in the domain of multimodal fusion of biometric signals in multimodal person identification application poses a fundamental question of justification of the modality dependence influence on the performance accuracy of the multimodal identification protocol. It should be also pointed out that a similar formulation is also valid for other multimodal fusion systems.

Our contribution:

- we perform an information-theoretic analysis of a problem of fusion of dependent/independent modalities using hypothesis testing fundamentals;
- error exponent bounds on probabilities of false alarm and miss for the case of multimodal classification are obtained as theoretical performance limits of systems fusing dependent/independent modalities;
- our analysis is general and does not rely on a particular assumption about the Gaussianity of the observed data.

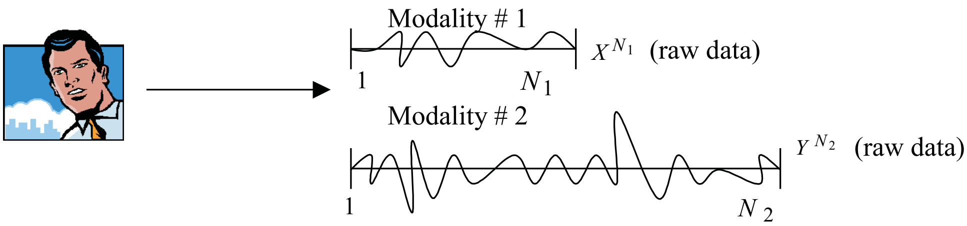

Figure 1: Multimodal observations: the vectors of different lengths might be observed.

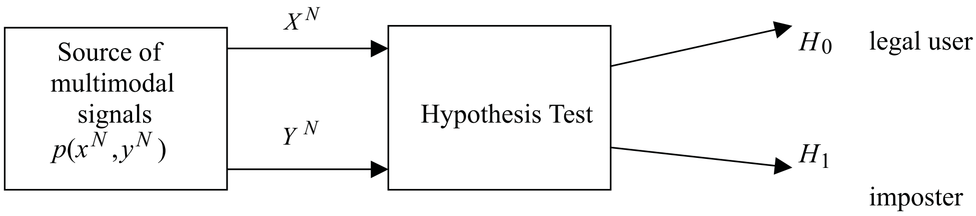

Figure 2: Modality fusion for user identification.

Setup under the analysis: In this setup we assume that a person is characterized by a pair of multimodal biometric signals (Fig. 1).

Given two biometric signals, we would like to solve a binary classification problem, as it is presented in Fig. 2 for the case of dependent and independent modalities.

The problem is formulated in the following way. Given a joint distribution of the observed multimodal vectors, i.e., $(X^N, Y^N) \sim q(x^N, y^N) = \prod_i q(x_i, y_i)$, fixing the a priori statistical model on alternative hypotheses to be $H_0 = p^0(x^N, y^N) = \prod_i p^0(x_i, y_i)$, $H_1 = p^1(x^N, y^N) = \prod_i p^1(x_i, y_i)$, it is necessary to decide, which hypothesis is in forth. We select a log-likelihood ratio decision rule that can be written in terms of relative entropy $D(.\,||\,.)$: $\eta = N\left\{D(q_{XY}(x,y)\,||\,p^1_{XY}(x,y)) - D(q_{XY}(x,y)\,||\,p^0_{XY}(x,y))\right\}$ where designates a length of the considered data vectors that in the scope of this study is supposed to be the same for different multimodal signals.

The measure of performance is derived according to the probability of false alarm P_F and probability of miss P_M.

According to the Stein lemma the performance of the Neyman-Pearson classifier is defined as:

$$P_F \sim 2^{-N\left[D(p^1_{XY}(x,y))||D(p^0_{XY}(x,y))\right]}, \text{for a fixed } P_M, P_M \sim 2^{-N\left[D(p^0_{XY}(x,y))||D(p^1_{XY}(x,y))\right]}, \text{for a fixed } P_F. \quad (1)$$

Thus, the performance is defined by the corresponding relative entropies. In the following, we will consider the cases of independent and dependent modalities.

Independent case: $p(x,y) = p(x)p(y)$.

The bounds on the probabilities of error are:

$$\begin{aligned} P_F &\sim 2^{-N\left[D(p^1_Y(y)||p^0_Y(y))+D(p^1_X(x)||p^0_X(x))\right]}, \text{for a fixed } P_M, \\ P_M &\sim 2^{-N\left[D(p^0_Y(y)||p^1_Y(y))+D(p^0_X(x)||p^1_Y(y))\right]}, \text{for a fixed } P_F. \end{aligned} \quad (2)$$

Dependent case: $p(x,y) \neq p(x)p(y)$.

The bounds on the probabilities of error are given by (1). According to the chain rule for the relative entropy one has:

$$D(p^1_{XY}(x,y)\,||\,p^0_{XY}(x,y)) = D(p^1_Y(y)\,||\,p^0_Y(y)) + D(p^1_{X|Y}(y\,|x)\,||\,p^0_{X|Y}(y\,|x)), \quad (3)$$

$$D(p^0_{XY}(x,y)\,||\,p^1_{XY}(x,y)) = D(p^0_Y(y)\,||\,p^1_Y(y)) + D(p^0_{X|Y}(y\,|x)\,||\,p^1_{X|Y}(x\,|y)). \tag{4}$$

Thus, in order to compare the bounds for dependent (1) and independent (2) cases one should compare two quantities (here we investigate the arguments of P_F since the analysis of P_M is symmetrical):

$$D(p^0_X(x)\,||\,p^1_X(x)) \text{ vs. } D(p^0_{X|Y}(x\,|y)\,||\,p^1_{X|Y}(x\,|y)). \tag{5}$$

In the case,

$$D(p^0_X(x)\,||\,p^1_X(x)) \leq D(p^0_{X|Y}(y\,|x)\,||\,p^1_{X|Y}(x\,|y)). \tag{6}$$

one can conclude that fusion of dependent modalities has a better performance than one obtained by fusion of independent signals.

Proof.

$$\begin{aligned}
&D(p^0_{X|Y}(y\,|x)\,||\,p^1_{X|Y}(y\,|x)) - D(p^0_X(x)\,||\,p^1_X(x)) = \\
&\sum_x \sum_y p^1_{XY}(x,y) \log \frac{p^1_{X|Y}(x|y)}{p^0_{X|Y}(x|y)} - \sum_x p^1_X(x) \log \frac{p^1_X(x)}{p^0_X(x)} = \\
&\sum_x \sum_y p^1_{XY}(x,y) \log \frac{p^1_{X|Y}(x|y)}{p^0_{X|Y}(x|y)} - \sum_x \sum_y p^1_{XY}(x,y) \log \frac{p^1_X(x)}{p^0_X(x)} = \\
&\sum_x \sum_y p^1_{XY}(x,y) \log \frac{p^1_{X|Y}(x|y) p^0_X(x)}{p^0_{X|Y}(x|y) p^1_X(x)} \geq \\
&1 - \sum_x \frac{p^1_X(x)}{p^0_X(x)} \sum_y p^1_Y(y) p^0_{X|Y}(x\,|y) = \\
&1 - \sum_x \frac{p^1_X(x)}{p^0_X(x)} p^0_X(x) = \\
&0,
\end{aligned} \tag{7}$$

where the only inequality in (7) is due to $\log(x) \geq 1 - \frac{1}{x}$.

Thus, based on (6) one can conclude that fusion of independent modalities gives the lower limit of performance enhancement in multimodal fusion classification problem. When modalities are dependent, the gain due to the fusion is higher in terms of reduction of error probabilities.

Particular case. Since in the case of Gaussian data independence is equivalent to the uncorrelation, one can conclude that fusion of correlated modalities leads to a higher accuracy in classification problem.

Further extensions: As an extension of the considered problem, we will analyze the setup presented in Fig. 3. The main task can be formulated as follows. Given a database of M users where each user is represented by a set of J modalities, to answer on the question which person is present observing J multimodal signals (or some of them).

This problem can be considered extending the developed information-theoretic analysis framework from a binary to multiple hypotheses testing formulation. The performance results will be derived.

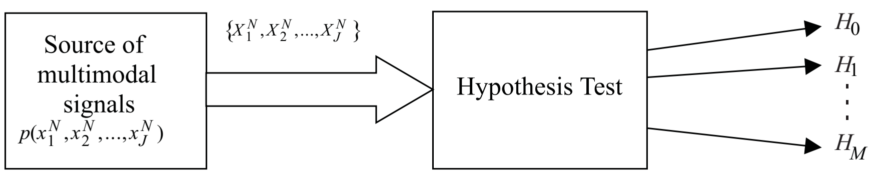

Figure 3: Multimodal user identification setup.

We consider the same formulation for the document retrieval and navigation problem using multiple modality signals. A document under the consideration consists of multiple modalities like text, images, drawings, etc., as well as hidden annotations (Fig. 4).

Given a multimodal document, one obtains an access to various extended multimedia files based on the observed modalities that are distorted due to the imperfectness of the acquisition process as well as extracts the annotations communicated in the document in an invisible manner. Based on a set of $J = J_a + J_m$, where J_a, J_m stay for a number of annotations and multimodal signals, respectively, (Fig. 6), the main task consists in a retrieval in the database of all document related multimedia files such as audio, video, animation, bibliographic references, translations, etc.

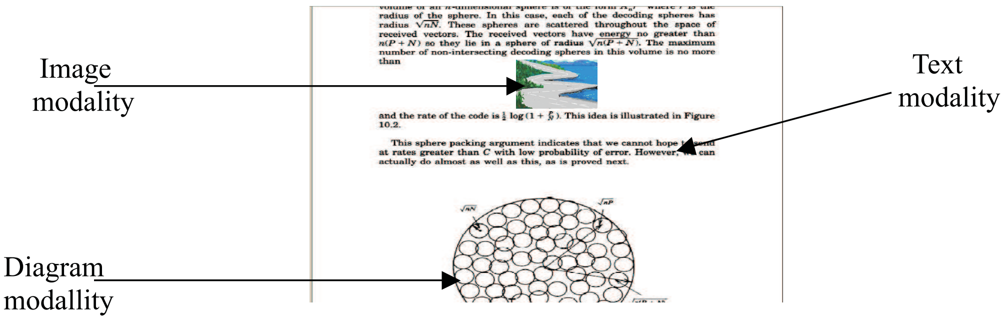

Figure 4: Multimodal user identification setup.

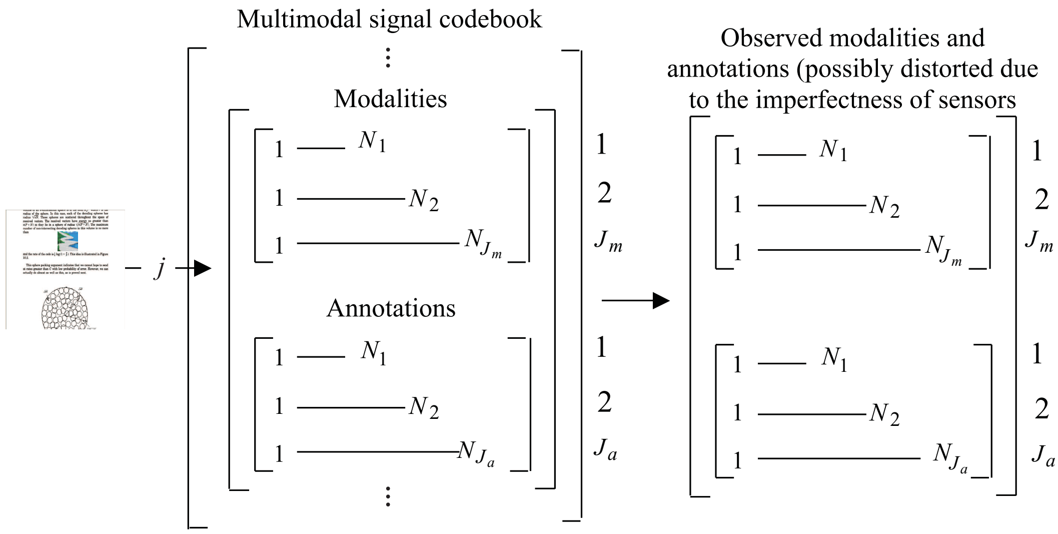

Figure 5: Multimodal document identification setup based on various modalities and hidden annotations.

Practical significance of the above theoretical results:

Given J multimodal signals, practical fusion should prefer dependent modalities if:

- low complexity implementation is required;
- there are limitations on storage or communication rates;
- there is a restriction on the sensor cost.

References:

1. B. Ulery, W. Fellner, P. Hallinan, A. Hicklin, and C. Watson, "Evaluation of Selected Biometric Fusion Techniques", http://www.itl.nist.gov/iad/894.03/pact/ir_7346_C.pdf.

2. J. Kittler, G. Matas, K. Jonsson, and M. Sanchez, "Combining Evidence in Personal Identity Verification Systems," Pattern Recognition Letters, vol. 18, no. 9, pp. 845-852, 1997.

3. N. Poh and S. Bengio, "How do correlation and variance of base classifiers affect fusion in biometric authentication tasks?" IEEE Transactions on Signal Processing, 53(11):4384-4396, 2005.

4. O. Ushmaev, S. Novikov, "Biometric fusion: Robust approach." In Workshop on Multimodal User. Authentication (MMUA 2006), Toulouse, 2006.

5. F. Roli, J. Kittler, G. Fumera, and D. Muntoni (2004) "An experimental comparison of classifier fusion rules for multimodal personal identity verification systems." http://citeseer.ist.psu.edu/cache/papers/cs/26860/http:zSzzSzce.diee.unica.itzSzenzSzpublicationszSzpapers-PattRecAppszSzMCS-Conference-13.pdf/roli02experimental.pdf.

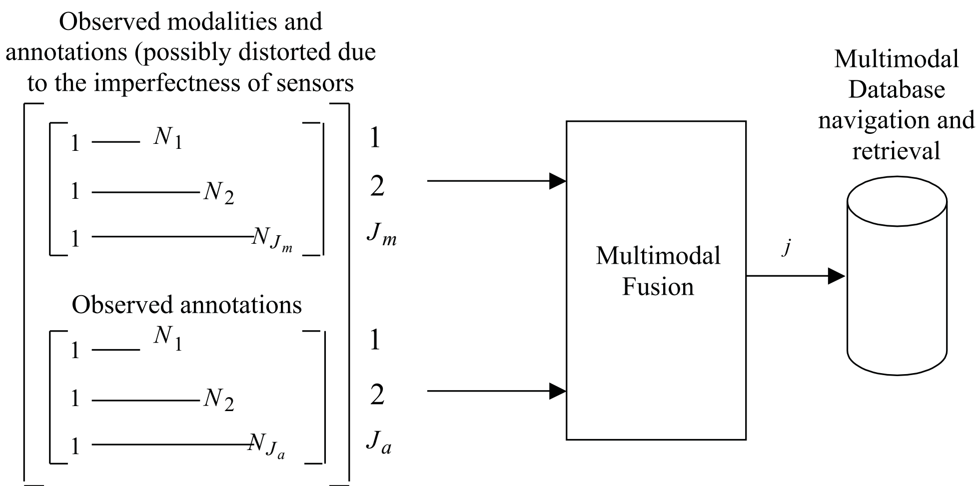

Figure 6: Multimodal document navigation.